

**Assessing the Development Effectiveness of
Multilateral Organizations:
Guidance on the Methodological Approach**

OECD DAC Network on Development Evaluation

Revised June 2012

Table of Contents

Acronyms.....	ii
1.0 Introduction	1
1.1 The Challenge: Gap in Information on Development Effectiveness.....	1
1.2 Background on Developing the New Approach.....	2
1.3 Common Criteria for Assessing Development Effectiveness	4
1.4 Preliminary Review	7
1.5 Implementing Option Two: Meta-Synthesis	13
2.0 Selecting Multilateral Organizations for Assessment	14
3.0 Selecting Evaluation Studies.....	15
3.1 Objective of the Sampling Process.....	15
3.2 Sampling Process	16
4.0 Screening the Quality of Evaluations Chosen for Review	17
4.1 Criteria for Screening Quality of Evaluations	17
4.2 Analysis of the Quality Screening Criteria.....	20
5.0 Scaling and Classifying Evaluation Findings.....	21
5.1 Scale for Evaluation Findings	21
5.2 Operational Guidelines for Classifying Evaluation Findings.....	23
6.0 Analysis of Effectiveness Results.....	34
6.1 Analysis of Quantitative Results	34
6.2 Factors Contributing to or Inhibiting Effectiveness	36
7.0 Organizing and Carrying Out the Development Effectiveness Assessment	37
7.1 Lead and Supporting Agencies and Their Roles	37
7.2 Carrying out a Meta-Synthesis of Evaluations	38
8.0 Development Effectiveness Assessment Report	39
Annex 1: Evaluation Quality Screening Scoring Guide for Multilateral Organizations	42
Annex 2: Operational Guidelines for Classifying Evaluation Findings	45

Acronyms

ADB	Asian Development Bank
CIDA	Canadian International Development Agency
DAC	Development Assistance Committee
MAR	Multilateral Aid Review
MDB	Multilateral Development Bank
MDG	Millennium Development Goals
MfDR	Managing for Development Results
MO	Multilateral organization
MOPAN	Multilateral Organization Performance Assessment Network
NGO	Non-governmental organization
RBM	Results-based management
SADEV	Swedish Agency for Development Evaluation
UK – DFID	United Kingdom Department for International Development
UN	United Nations
UNDP	United Nations Development Programme
UNEG	United Nations Evaluation Group
USAID	United States Agency for International Development
WHO	World Health Organization
WFP	World Food Programme

1.0 Introduction

1.1 The Challenge: Gap in Information on Development Effectiveness

Decision makers in bilateral development agencies which provide funding to Multilateral Organizations (MOs) have for some time identified a persistent gap in the information available to them regarding the development effectiveness of MOs. While they have access to reasonably current and reliable information on Organizational Effectiveness (how MOs are managing their efforts to produce development results), they often lack similar information regarding development effectiveness. This creates a challenge for decision-making about where and how to allocate resources in an evidence-based manner.

This persistent gap arises from a number of factors:

- Considerable variability from one MO to another in the coverage, quality and reliability of their reporting on development effectiveness in terms of the attainment of objectives and the achievement of expected results of their activities and the programs they support;
- The relative infrequency, and long duration (usually over two years from concept to completion) and costs normally exceeding \$1.5 million USD of independent joint evaluations of MOs. While these studies do often result in important, actionable findings on development effectiveness they are not able to provide coverage of most MOs. For these reasons, they do not represent a feasible or cost effective approach to cover the large number of MOs; and
- The focus of most existing, systematic efforts to assess the effectiveness of MOs is on systems and procedures for managing towards development results rather than on the results themselves. While these systems are in evolution, at least for the present the Multilateral Organization Performance Assessment Network (MOPAN), COMPAS, and the Development Assistance Committee (DAC)/United Nations Evaluation Group (UNEG) Peer Reviews of the Evaluation Function of multilateral organizations, they do not directly address development effectiveness.

In response to an initiative proposed by the Canadian International Development Agency (CIDA), the Network on Development Evaluation of the DAC has been investigating an approach to addressing the challenge of improving the body of available information on

the development effectiveness of MOs. The approach which has been developed and tested over the past two years has two key features:

1. It seeks to generate a body of credible information on a common set of criteria that would provide a picture of the development effectiveness of each MO; and
2. It builds on information, particularly evaluation reports, which is already available from the MOs themselves using methods which are modest in time and cost requirements and with limited burden on each organization.

1.2 Background on Developing the New Approach

Following a proposal by CIDA in June 2009 to the DAC Evaluation Network, the Network established a Task Team to work with CIDA to further discuss and refine the approach. After meeting in Ottawa in October 2009 to further discuss the initiative the sixteen-member Task Team agreed to establish a smaller Management Group, composed of representatives from the Canadian International Development Agency (CIDA), the Department for International Development (UK – DFID), the Swedish Agency for Development Evaluation (SADEV), the United States Agency for International Development (USAID), the World Bank and the UNEG, to guide the work. The Management Group was responsible for day-to-day interaction with the CIDA team that carried out pilot studies of the approach in 2010 and was active in the further refinement of the guidance document.

A revised proposal, prepared by CIDA under the guidance of the Management Group, was submitted to the February 2010 meeting of the DAC Evaluation Network for discussion. The meeting agreed to a pilot test phase to be carried out in 2010 and asked CIDA to coordinate the work, working in close cooperation with MOPAN to avoid duplication. The Management Group met in Ottawa in April 2010 to finalize the details and agreed to pilot test the approach on the Asian Development Bank (ADB) and the World Health Organization (WHO), two of the organizations being surveyed by MOPAN in 2010.

The pilot test was carried out from June to October 2010 and the results were reported to the Network meeting in November. The pilot test concluded that the proposed approach was workable and that it provided a reasonable picture on the development effectiveness of the MO that can be obtained within a reasonable time-frame of 6-8 months and a resource envelope of approximately \$125,000, with little burden for the MO under review.

The pilot test also concluded that the proposed approach works well when the evaluation function of the MO produces a volume of evaluation reports of reasonable quality over a three to four-year period covering a significant portion of its investments. Where the MO has not produced a sufficient number of evaluation reports over the 3 - 4 year time-frame of the study and where data on both evaluation and program coverage is lacking, the approach is still workable and useful in providing valuable information on development effectiveness; however, the findings cannot be easily generalized to the performance of the MO with a reasonable level of confidence. But it still complements the data available from other exercises and offers insights than can be used for dialogue between the MO and its stakeholders through their regular channels of communication.

The pilot test noted that there were opportunities to further refine and improve the methodology and to reflect these improvements in updated guidelines. The pilot test team met in February 2011 to discuss the detailed improvements proposed in the pilot test report along with other suggestions from team members. The results of the February workshop, as discussed in the May 2011 meeting of the Management Group, were incorporated into an update of this guidance document. The approach was then implemented in 2011/12 to assess the effectiveness of two other UN agencies, namely the United Nations Development Programme (UNDP) and the World Food Programme (WFP). The lessons learned from the process of conducting these two reviews have been integrated into this guidance document.¹ Advice for future reviews is included throughout this document in text boxes titled “Tips.”

Throughout the pilot test, the technical team continued to monitor ongoing developments among other systems and processes for assessing MO performance such as the Multilateral Aid Review (MAR) completed by the UK’s Department for International Development as well as ongoing efforts to strengthen the COMPAS and MOPAN systems. The recent initiative which comes closest to this one in its focus on development effectiveness is clearly the MAR, carried out in 2010/11 by DFID. It seems clear that, had the initiative as pilot tested been in operation for three to four years before the completion of the MAR, it could have made a strong contribution to that exercise.

Ongoing efforts to improve systems such as MOPAN and COMPAS have not yet entered directly into the area of development effectiveness and do not yet address the key information gap which this initiative aims to fill. MOPAN has recently begun

¹ This included lessons learned with respect to humanitarian, as well as development, programming effectiveness.

internal discussion on the possibility of extending its remit to encompass some elements of development effectiveness but these are only exploratory at this time.

1.3 Common Criteria for Assessing Development Effectiveness

1.3.1 Defining Development Effectiveness

From the beginning of this initiative in 2009, the technical team and members of the Management Group and the Task Team have considered the question of whether an explicit definition of development effectiveness was needed and what it should include. The currently available definitions seemed either too broad, in that they blended elements of development and organizational effectiveness, or too specific. In the latter category, goals and targets such as the Millennium Development Goals (MDG) may fall outside the mandate of some MOs and, in any case, are not always the specific focus of either monitoring systems or evaluation reports.

A more workable approach, in the absence of an acceptable definition of development effectiveness, may be to build on criteria which best describe the characteristics of MO programs and activities which are clearly linked to supporting development (and humanitarian) results. The characteristics of recognizably effective development programs would include:

1. Programming activities and outputs would be relevant to the needs of the target group and its members;
2. The programming would contribute to the achievement of development objectives and expected development results at the national and local level in developing countries (including positive impacts for target group members);
3. The benefits experienced by target group members and the development (and humanitarian) results achieved would be sustainable in the future;
4. The programming would be delivered in a cost efficient manner; and
5. The programming would be inclusive in that it would support gender equality and would be environmentally sustainable (thereby not compromising the development prospects in the future).

A sixth relevant criterion, which is not directly related to assessing development effectiveness, was also identified. However, it is useful as an explanatory variable, but will be analyzed separately:

6. Programming would enable effective development by allowing participating and supporting organizations to learn from experience and use of performance management and accountability tools, such as evaluation and monitoring, to improve effectiveness over time.

Thus, the methodology focuses on a description of developmentally effective MO programming rather than a definition as such. It has been developed for its workability and its clear relationship to the DAC Evaluation Criteria. This description is provided only for the purpose of grounding the work of the proposed approach to assessing the development effectiveness of an MO. It is not meant to provide a comprehensive and universal definition of the phenomenon of development effectiveness.

1.3.2 Common Criteria

There are six main criteria and 19 sub-criteria. The six common development effectiveness assessment criteria are:

- Achievement of development objectives and expected results;
- Cross-cutting themes (environmental sustainability and gender equality);
- Sustainability of results/benefits;
- Relevance of interventions;
- Efficiency; and,
- Using evaluation and monitoring to improve development effectiveness.

Table 1 below presents the Common development effectiveness Criteria and detailed sub-criteria to be used in the review.

Table 1: Common Development Effectiveness Assessment Criteria

One: Criteria for Assessing Development Effectiveness	
Achieving Development Objectives and Expected Results	
1.1	Programs and projects achieve their stated development objectives and attain expected results.
1.2	Programs and projects have resulted in positive benefits for target group members.
1.3	Programs and projects made differences for a substantial number of beneficiaries and where appropriate contributed to national development goals.
1.4	Programs contributed to significant changes in national development policies and programs (including for disaster preparedness, emergency response and rehabilitation) (policy impacts) and/or to needed system reforms.

One: Criteria for Assessing Development Effectiveness	
Cross-Cutting Themes – Inclusive Development which is Sustainable	
2.1	Extent to which multilateral organization supported activities effectively address the cross-cutting issue of gender equality.
2.2	Extent to which changes are environmentally sustainable.
Sustainability of Results/Benefits	
3.1	Benefits continuing or likely to continue after project or program completion or there are effective measures to link the humanitarian relief operations, to rehabilitation, reconstructions and, eventually, to longer term development results.
3.2	Projects and programs are reported as sustainable in terms of institutional and/or community capacity.
3.3	Programming contributes to strengthening the enabling environment for development.
Relevance of Interventions	
4.1	Programs and projects are suited to the needs and/or priorities of the target group.
4.2	Projects and programs align with national development goals.
4.3	Effective partnerships with governments, bilateral and multilateral development and humanitarian organizations and NGOs for planning, coordination and implementation of support to development and/or emergency preparedness, humanitarian relief and rehabilitation efforts.
Efficiency	
5.1	Program activities are evaluated as cost/resource efficient.
5.2	Implementation and objectives achieved on time (given the context, in the case of humanitarian programming).
5.3	Systems and procedures for project/program implementation and follow up are efficient (including systems for engaging staff, procuring project inputs, disbursing payment, logistical arrangements etc.).
Using Evaluation and Monitoring to Improve Development Effectiveness	
6.1	Systems and process for evaluation are effective.
6.2	Systems and processes for monitoring and reporting on program results are effective.
6.3	Results-based management (RBM) systems are effective.
6.4	Evaluation is used to improve development effectiveness.
Notes on Using the Criteria	
<p>In applying the above criteria to the review of evaluations concerning a given MO it is important to note the following guidelines.</p> <ul style="list-style-type: none"> • The findings of each evaluation as they relate to each of the 15 sub-criteria listed above (plus 4 relating to the use of evaluation) are to be classified using a four-point scale which is detailed below. • Analysis and reporting will be by sub-criteria and criteria only. There is no intention to combine criteria into a single index of development effectiveness. • Specification of micro-indicators is not recommended because experience with similar evaluation reviews suggests that over-specification of the review criteria soon leads to very few evaluations 	

One: Criteria for Assessing Development Effectiveness

being specific enough in their findings to address a significant number of micro indicators. In essence, this would involve over retro-fitting of external criteria to the evaluations in a way which distorts their original design. The experience of the pilot tests suggests that the level of specification illustrated is precise enough to allow for reliable classification of evaluation findings and still broad enough that most reasonably well designed evaluations will have findings which contribute to a credible analysis.

1.4 Preliminary Review

1.4.1 Information Sources for the Preliminary Review

The process of assessing the development effectiveness of a multilateral organization begins with a Preliminary Review of its regularly published reports which focus on development effectiveness as well as a brief assessment of the coverage and quality of the program and project evaluations it produces. The most common documents which may include significant elements of development effectiveness reporting from a given MO include:

- **An Annual Report on Development Results** at the organizational level. For Multilateral Development Banks (MDBs), this is often based on a review of project completion reports which may or may not have been audited for accuracy by the central evaluation group. For United Nations (UN) organizations this report may track MDG results across partner countries and may be supplemented (as in the case of UNDP) by highlights from evaluations. For UN agencies it can usually be found in the documents submitted to the governing body at its main annual meeting.
- **An Annual Summary of Evaluation Results.** This is a common document among MDBs and some UN organizations and typically presents both extracted highlights and some statistical data on the coverage and results of evaluations published in a given year.
- **A Report on Progress Towards the Objectives of the Strategic Plan.** This report is not necessarily issued annually. It can either relate to the biennial budget of a UN organization or to a three to five year strategic plan of a MDB.
- **The COMPAS Entry for an MDB.** While COMPAS reports focus mainly on a given MDB's progress in Managing for Development Results (MfDR) it does include some information on the status of completed projects as compiled in project completion reports and, sometimes, evaluations.

As well as reviewing the published development effectiveness reports listed above, the Preliminary Review should include a brief survey of program evaluation reports available from a given MO and of any surveys of the evaluation function carried out either internally or externally. These would of course include available reports on the evaluation function produced by the DAC/UNEG Peer Reviews of the Evaluation Function. It would also include, for example, reports by the internal auditor of a given MO which describe the evaluation function as submitted to the governing body and, if available, the results of any recent MOPAN survey as they relate to the evaluation function.

1.4.2 Scenarios and Options

If the Preliminary Review illustrates both that the MO produces credible, regular reports on the development effectiveness and has a credible evaluation function which provides adequate coverage of the range of MO's investments, the MO's information can be used to as the basis for assessing development effectiveness. If not, the other options indicated below should be implemented.

The Preliminary Review will determine whether a given MO can be classified in one of three possible scenarios, each leading to a course of action or option. The content of the options changes depending on the reliability of the MO's systems for reporting on effectiveness as follows:

- Scenario A: The MO's reporting on the core development effectiveness criteria is adequate and based on an adequate body of reliable and credible information, including information from evaluations covering a large proportion of its investments, which provides a picture of the organization's performance. If this is the case, (the MO's reporting on the development effectiveness criteria is rigorous and evidence-based, covering a large proportion of its investments) then its reporting may be relied upon. This would correspond to Option 1 in the framework (see Annex A).
- Option 1 of the framework is simply reliance on the MO's own reporting of development effectiveness. It is to be used whenever MO reports on development effectiveness are rigorous, evidence-based and of sufficient coverage and frequency to satisfy the information needs of the bilateral shareholders. This is the medium to long-term goal of all activities undertaken under the proposed approach.
- Scenario B: The MO's reporting on the core development effectiveness criteria is not adequate but the MO has an evaluation function which produces an adequate body of

reliable and credible evaluation information from which a picture of the organization's performance on the development effectiveness criteria can be synthesized to satisfy the information needs bilateral shareholders.

- Option 2 of the framework is intended to be used when there is a perceived need to improve information on the development effectiveness of the chosen MO provided its own evaluation function produces a reasonable body of quality evaluations. The core of this option is the use of meta-synthesis and analysis methods to “conduct a systematic synthesis of available evaluations as they relate to MO development effectiveness.”²
- Scenario C: The MO's reporting on the core development effectiveness criteria is not adequate, and the body of evaluation information available from its evaluation function is either not adequate or is not credible or reliable for a synthesis to provide a picture of the organization's performance on the development effectiveness criteria for the bilateral shareholders.
- Option 3: For those MOs without a reasonable body of quality evaluations for review, Option 3 of the framework suggests participating agencies would need to consider actions aimed at strengthening the MO's development effectiveness reporting. These could include (among others): joint evaluation by a group of bilateral donors; direct support of MO results reporting and evaluation systems (e.g. through DAC/UNEG Peer Review of the evaluation function); and provision of support to MO efforts to strengthen decentralized evaluation systems.

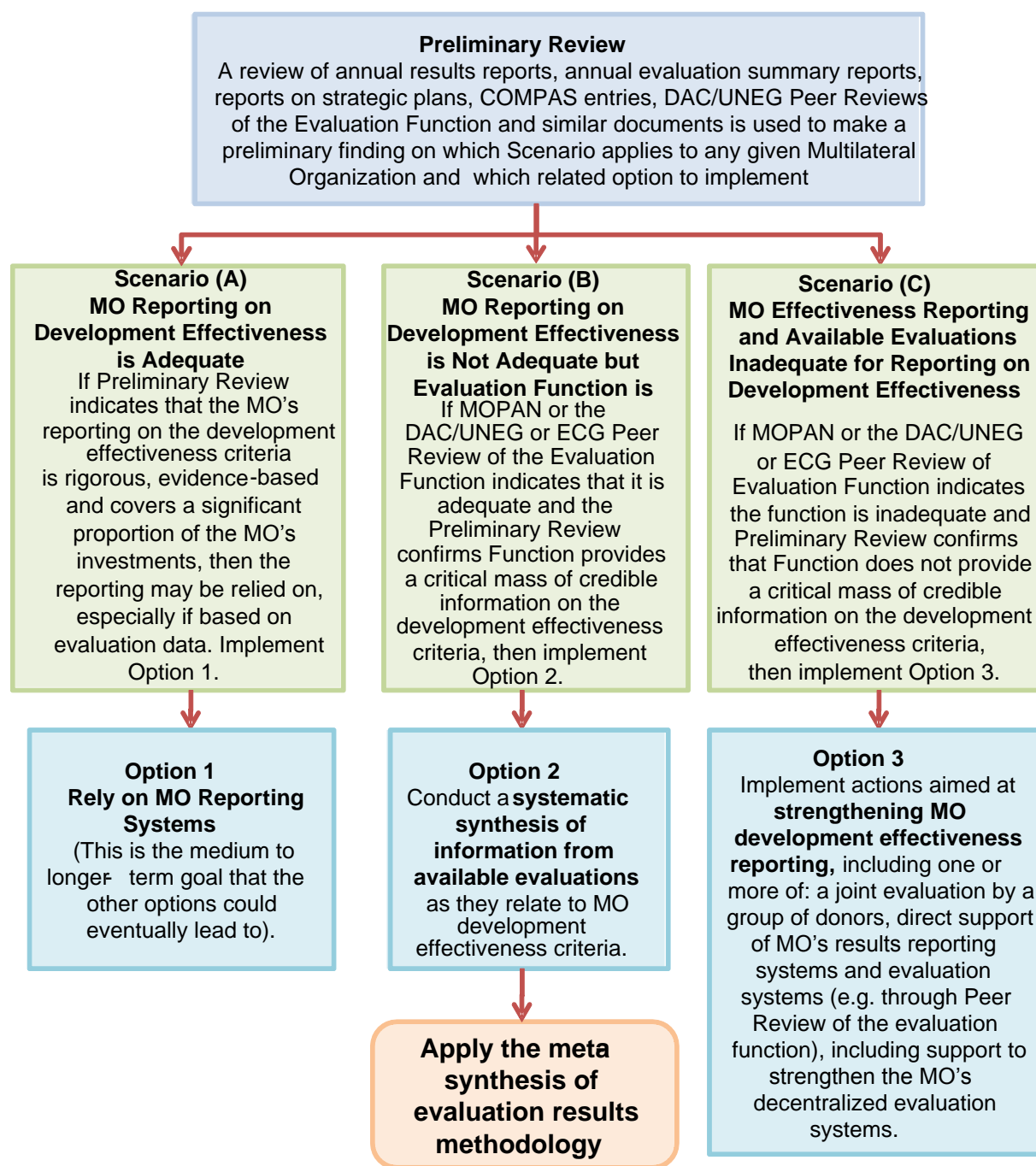
The findings would enable member agencies of the DAC Network on Development Evaluation to raise issues relating to the adequacy of results reporting, evaluation and RBM systems through their participation in the governing bodies of the relevant MO.

1.4.3 Guidance for Selecting Scenarios and Options

The decision to classify a given MO into the scenarios (and thus to select the appropriate option) as described in Table 2, necessarily involves a considerable element of judgment on the part of the lead agency for the assessment.

² Proposed Approach to Strengthening Information on the Development Effectiveness of Multilateral Organizations. DAC Network on Development Evaluation, Task Team on Multilateral Effectiveness. P8.

Table 2: Steps in the Development Effectiveness Assessment Process: Scenarios and Options



The classification is much easier to carry out after at least one round of meta-synthesis since that exercise will provide critical information on the quality and coverage of evaluations carried out by the MO concerned; evaluations which in turn may form an

important element in the evidence based supporting MO reporting on development effectiveness. In fact, the pilot test of the methodology illustrated clearly that one of the MOs reviewed would have been best classified under Scenario 3, after it was included in the test because all readily available evidence (prior to the meta-synthesis) suggested it belonged in Scenario 2.

This provides an important clue to the key element to be used in arriving at a classification of an MO under one of the scenarios in Table 2: the quality of the evidence (and its coverage) used to support claims of development effectiveness made in the key reporting documents of the MO.

In practical terms this means that classification under each scenario requires the following:

Scenario A: Reporting on Development Effectiveness is Adequate (No Further Action, Option One)

1. MO reports on effectiveness cover all six key criteria of development effectiveness, although the reporting may be spread across more than one instrument;
2. MO reports clearly describe the evidence base used to support claims of development effectiveness. These may include end-of-project reports, references to monitoring results of quantitative targets at country level such as the MDGs, and project and program evaluations, including impact evaluations. Where project completion reports are relied on the MO, the reports should include the results of an independent review of the accuracy of those reports carried out by an independent evaluation group within the MO;
3. The evidence used to support assertions of development effectiveness should provide coverage of most MO programming and activities including country and regional programs and key sectors of development. This may be on a rolling basis rather than a single annual report but the report should indicate how coverage is being achieved and provide an estimate of coverage; and
4. There should be some indication that the evidence used has been independently checked for quality, this check may be done by the independent evaluation office or by an external group engaged in a process such as the Peer Reviews of the Evaluation Function carried out under the auspices of the Development Evaluation Network of the DAC.

Scenario B: MO reporting on development effectiveness is not adequate but the MOs evaluation function provides an adequate information base on development effectiveness (Conduct Meta Synthesis, Option Two)

1. MO reporting on development effectiveness does not cover all six key criteria, even when sought across more than one report type, or;
2. MO claims of development effectiveness are not substantiated with quantitative data gathered with the source and relation to MO programming clearly identified, or;
3. Where the MO reports of development effectiveness do refer to quantitative data from more than one source (including end of project reports and evaluations for example) the sources do not provide coverage of a significant portion of MO programming (this coverage can be attained over a 3-5 year period if the MO reports it that way) or it is impossible to estimate the level of coverage from the information provided.

If conditions one, two, or three apply, the next step is to determine if a meta-synthesis is likely to produce useful results. If so, the MO should meet the following criteria:

4. MO has published or can make available a sufficient number of evaluation reports with enough scope to allow for selection of a sample which covers a representative selection of MO programming over a three to four year time frame; and,
5. The evaluations published or made available by the MO are of reasonable quality and include findings relating to most of the assessment criteria in the model. If a Peer Review of the Evaluation Function for this MO has been carried out within the past three to four years it can provide important corroborating information for determining that the evaluation function is strong enough to support a meta-synthesis. If not, it will be necessary to review the quality of a sub-set of the evaluation sample to determine if this condition is met.

Scenario C: MO reporting on development effectiveness is not adequate but the MO evaluation function does not provide an adequate information base on development effectiveness (Support Development Effectiveness Reporting, Including Monitoring and Evaluation function, Option Three)

Scenario C will be in place and Option Three will be appropriate when the first three conditions under Scenario B above apply (lack of coverage of development effectiveness criteria in reporting, or reporting not substantiated by clearly identified, independently verified quantitative data, or inadequate coverage) but the last two conditions (a

sufficient supply of reasonably quality evaluations covering the bulk of MO programming over a three to four year time frame) do not apply.

At least in the first cycle of development effectiveness assessments, it seems likely that most MOs will be classified under Scenario Two and a meta-synthesis will be the preferred option. For MOs with more developed systems of results reporting a single cycle may result in their reclassification under Scenario One. The pilot test indicated, for example, that this would be the case for the Asian Development Bank.

1.5 Implementing Option Two: Meta-Synthesis

If a meta-synthesis of evaluation results is required (Scenario B, Option 2), the remaining sections of the guideline can be applied to that task. As already noted, where scenarios A and C are found (sufficient, credible development effectiveness reporting is already in place (A) or there is no credible set of program evaluations (C), Network members have access to the tools described in Options 1 and 3.

In order to provide a strong, common basis for implementing Option 2, these guidelines are organized around the process of designing and implementing the meta-synthesis of the development effectiveness of an MO in a given year. Thus the guidelines focus on the overall design and implementation process of a development effectiveness Assessment which is prepared mainly based on a meta-synthesis of available evaluations. Information is provided on:

- The process and criteria used for selecting MOs and for drawing a sample of evaluations to be reviewed;
- The process and criteria to be used in assessing the quality of evaluations so that they can be included in the review of evaluation findings and contribute to the assessment of development effectiveness;
- The approach to be followed in scaling and classifying evaluation findings in the reports under review;
- The method for conducting a content analysis of open ended information in evaluation reports, especially relating to factors contributing to or impeding development effectiveness;
- The relationship among participating agencies and the organization and management of the assessment team; and
- The reporting format to be followed.

2.0 Selecting Multilateral Organizations for Assessment

Since their early meetings in 2009, the Task Team and the Management Group have emphasized that the selection of MOs for assessment should begin with the list of organizations subject to a MOPAN survey and report each year. In order to harmonize with MOPAN and to allow the two initiatives to be complementary rather than overlapping and duplicative (especially in terms of reporting) the development effectiveness Review MOs are to be drawn from the list of MOPAN survey organizations each year.

The second step in the process of identifying MOs for assessment involves member agencies of the DAC Network on Development Evaluation indicating which organizations they have an interest in for review purpose, either as the lead or a supporting agency.

When the MOs of interest have been identified, the next step is the Preliminary Review described in Section 1.3. As the pilot test made clear, in order to effectively conduct the development effectiveness Assessment and to avoid attempting a meta-synthesis based on a poor sample of evaluations, it is first necessary to identify the position of each MO in the framework on development effectiveness (Table 2).

The appropriate sequencing of decisions and the rules to be applied would be as follows:

1. MOs to be considered for assessment in a given year to be drawn from the MOPAN list to ensure a coordinated approach; and
2. MOs of interest on the MOPAN list to undergo a Preliminary Review to determine the appropriate scenario from Table 2 which applies to each of the chosen MOs using the guidance on assigning these scenarios and selecting options detailed in Section 1.4.2 above.

The lead agency for the review should inform the management of the multilateral organization and its governing body of the proposed review. It is also important to contact the organization's evaluation department at an early stage to plan the exercise.

3.0 Selecting Evaluation Studies

The set of evaluations to be analyzed for each development effectiveness assessment will need to be carefully selected in order to provide adequate coverage of the MOs activity in support of development programming at country-level. It is important during the sampling process to meet with staff of the MO's evaluation function to set the context for the review. It allows the review team to develop a better understanding of the scope of evaluations conducted and would inform the sample selection process. This will also help to identify how the review can be beneficial to the MO.

The goal of the sampling strategy is to ensure that the evaluations chosen ensure coverage of a reasonable breadth of agency activities and a critical mass of MO investments in a given period. There is no set number of evaluation studies that should be chosen to achieve this coverage. In cases where the universe of available evaluations is small, all evaluations will be reviewed.

The most critical question for sampling evaluations reports for a given Multilateral Organization would seem to be what will be the main organizing principle for estimating the level of coverage provided by the sample of evaluation reports when compared to the distribution of a given MOs programming activities. Where country programs are a major focus of evaluation work and budget and expenditure data is presented on a geographic basis, geographic sampling is the most obvious choice. However, for UN organizations in particular, this is often not the case. Rather than propose a given level of coverage in the methodology guidelines, it may be best to describe a sampling objective and the process to be followed.

3.1 Objective of the Sampling Process

To select a sample of evaluation reports which provide reasonable coverage of MO programming over a specified time-frame with information available to allow for a comparison between programs covered by the evaluations in the sample and those implemented by the MO in question during the same time-frame.

3.2 Sampling Process

The sampling process involves the following steps:

1. Identify the universe of evaluation reports available from a given MO and published either internally or externally over a three- to four- year time-frame.³
2. Examine a subset of the available evaluation reports to determine how their scope of each evaluation's coverage is defined (if it is) in terms of geographic coverage (sub-national, country program level, regional programs or global programs); thematic coverage (focusing on, for example, gender equality, poverty alleviation or environmental sustainability); objectives coverage (focusing on a achievement of an overall MO objective or strategic priority); sector of coverage; or technical focus (implementation or roll-out of a specific intervention).
3. Examine the MO's annual reports, budgets, reports to the governing body, etc. In order to establish a profile of activities (normally expressed in financial terms) which can be compared to the set of parameters used to define the scope of published evaluations and thus to allow for an estimate of the percentage of MO activities covered by a given set of evaluations.
4. Identify the primary measure used to assess coverage of a given sample of evaluation reports. Possibilities include:
 - a) Geographic coverage (percentage of all expenditures in a given period accounted for by country, regional or global evaluation program evaluation reports under review);
 - b) Sector coverage (proportion of the main programming sectors supported by the MO as identified in budgets and annual reports which are accounted for in the evaluation sample);
 - c) Objectives/strategy coverage (the number of major, strategic MO objectives as identified in annual reports and budget documents which are addressed in the selected sample of evaluations compared to the number of strategic objectives identified);

³ The review team undertaking the meta-synthesis will review the universe of evaluation reports available from a given MO and determine if a sample (possibly including all the available reports) providing adequate coverage of the MOs programs and activities can be identified. If not, the MO in question can be assigned to Scenario Three as described in Table 2.

- d) Other estimates of thematic, technical or other dimensions of scope as identified in the evaluations and reported on an agency-wide basis either in the documents themselves or in annual reports and budgets (or other high level documents).

Of course, the sample of evaluations can be assessed in terms of coverage along more than one of the above dimensions. In the case of the sample drawn for the ADB pilot test for example, coverage was sought first on geographic and then on sector programming lines. The final decision on which measure of evaluation scope should be given priority will depend on how a given MO defines the scope of its evaluations and how it reports on the profile of its program activities; and

Tip

The review of evaluation quality should be carried out simultaneously with the review of the evaluations findings. Although this means that some evaluations, for which findings have already been rated, will be dropped from the sample because they fail to meet the quality requirements, it is still more efficient that conducting two separate reviews of the evaluation reports.

- e) Select a sample of evaluation reports using the criteria identified in D above, and use budgetary and other MO-wide data to estimate the proportion of MO activities covered by the evaluation sample.

The question of what level of coverage should be reached (i.e. the percentage of, for example, geographic programs covered) is a matter of judgment but it should be transparently presented in the development effectiveness report on a given MO.

4.0 Screening the Quality of Evaluations Chosen for Review

4.1 Criteria for Screening Quality of Evaluations

The purpose of screening the evaluations chosen for review is not to provide a comparable score across different organizations or conduct a detailed analysis of the quality of evaluations or the strength of the evaluation function in a given MO. Rather it is to determine if the findings and conclusions of the evaluation can be relied on as a reasonable guide to the development effectiveness of that element of MO programming

which is covered by the evaluation report under review.⁴

The purpose of screening the evaluations chosen for review is not to provide a comparable score across different organizations or conduct a detailed analysis of the quality of evaluations or the strength of the evaluation function in a given MO. Rather it is to determine if the findings and conclusions of the evaluation can be relied on as a reasonable guide to the development effectiveness of that element of MO programming which is covered by the evaluation report under review.⁵

The factors to be assessed during the screening process must be those which can be directly reflected *in the evaluation report itself* since many reports will not have appended their commissioning documents or Terms of Reference. The factors listed here are derived from the *DAC Quality Standards for Development Evaluation* and the *Standards in Evaluation in the UN System* published by the UNEG in 2005.

The pilot tests indicated that reviewing evaluations in order to assess quality can be very resource intensive if too many or overly complex quality criteria are used. With that experience in mind, the technical team has simplified the list of criteria.

The factors to be assessed when reviewing the selected evaluation reports are illustrated in Table 3. A copy of a working tool for implementing this screening guide is provided in Annex 1.

Tip

Given the importance of evaluation quality in assessing the MO's effectiveness, the quality review is a critical component of the overall review. As such, it is suggested that the review team leader plays a key role in ensuring the reliability of the quality reviews. For example, he or she could conduct a second review of a sample of evaluations, after the initial quality review by the team member. The team leader may also opt to review all evaluations that are to be excluded or are within five points of the cut-off point.

⁴ While the purpose and intent of the quality screening process is not to provide an overall assessment of the evaluation function of a given MO (that is the purpose of the DAC/UNEG Peer Review of the Evaluation Function) the distribution of evaluation quality scores for a given MO does provide interesting information on the quality of output of the function.

⁵ While the purpose and intent of the quality screening process is not to provide an overall assessment of the evaluation function of a given MO (that is the purpose of the DAC/UNEG Peer Review of the Evaluation Function) the distribution of evaluation quality scores for a given MO does provide interesting information on the quality of output of the function.

Table 3: Evaluation Quality Screening Guide

	Criteria to be Scored		
A	Purpose of the evaluation is clearly stated. The report describes why the evaluation was done, what triggered it (including timing in the project/program cycle) and how it was to be used.	3	
B	Evaluation objectives are stated. Evaluation objectives are clearly presented and follow directly from the stated purpose of the evaluation.	2	
C	The evaluation report is organized, transparently structured, clearly presented and well written. There is a logical structure to the organization of the evaluation report. The report is well written with clear distinctions and linkages made between evidence, findings, conclusions and recommendations.	3	
D	Subject evaluated is clearly described. Evaluation report describes the activity/program being evaluated, its expected achievements, how the development problem would be addressed by the activity and the implementation modalities used.	4	
E	Scope of the evaluation is clearly defined. The report defines the boundaries of the evaluation in terms of time period covered, implementation phase under review, geographic area, and dimensions of stakeholder involvement being examined.	4	
F	Evaluation criteria used to assess program effectiveness are clearly identified in the evaluation report and cover a significant number of the common criteria for assessing development effectiveness.	5	
G	Multiple lines of evidence are used. The report indicates that more than one line of evidence (case studies, surveys, site visits, key informant interviews) is used to address the main evaluation issues. One point per line of evidence, to maximum of four.	4	
H	Evaluations are well designed. The methods used in the evaluation are appropriate to the evaluation criteria and key issues addressed. Elements of good design include: an explicit theory of how objectives and results were to be achieved, specification of the level of results achieved (output, outcome, impact), baseline data (quantitative or qualitative) on conditions prior to program implementation, a comparison of conditions after program delivery to those before, and a qualitative or quantitative comparison of conditions among program participants and those who did not take part.	5	
I	Evaluation findings and conclusions are relevant and evidence based. The report includes evaluation findings relevant to the assessment criteria specified. Findings are supported by evidence resulting from the chosen methodologies. There is a clear logical link between the evidence and the	4	

	Criteria to be Scored		
	findings and conclusions are linked to the evaluation findings as reported.		
J	Evaluation report indicates limitations of the methodology. The report includes a section noting the limitations of the methodology and describes their impact on the validity of results and any measures taken to address the limitations (re-surveys, follow-ups, additional case studies, etc.	3	
K	Evaluation includes recommendations. The evaluation report contains specific recommendations that follow-on clearly from the findings and conclusions. Further the recommendations are specifically directed to one or more organizations and aimed at improving development effectiveness. (Objectives achievement, cross cutting themes, sustainability, cost efficiency or relevance).	3	
	Total	40	

4.2 Analysis of the Quality Screening Criteria

Once the team has finished screening the quality of the evaluations in the sample, the next task is to analyze the results in order to determine which evaluations should be retained in the sample for the review. This is a two-step process.

The first level of analysis is to determine which evaluations meet the overall threshold for inclusion. Based on the experience of the pilot and previous applications of this instrument, it has been determined that the cut-off point should be 25 points (or about 65% of the total available points). Any evaluations that do not receive at least 25 points should be excluded.

The second level of analysis focuses on the quality criteria that relate most directly to a review of development effectiveness. If an evaluation does not score well with respect to the design and implementation of the evaluation methodologies, then it is not likely to provide quality information on the MO's effectiveness. Therefore, all evaluations to be included in the sample for the review should attain a minimum of points for the key criteria related to the quality of the evaluation evidence. Based on the experience of the earlier reviews, it is suggested that an evaluation should obtain a minimum of 10 points out of the maximum 13 points for the Criteria G, H and I, to be included in the sample (See Table 4).

Table 4: Minimum Points Required for Key Quality of Evaluation Criteria

	Key Criteria Related to the Quality of Evaluations	Points
G	Multiple lines of evidence are used	4
H	Evaluations are well designed	5
I	Evaluation findings and conclusions are relevant and evidence based	4
	Total	13
	Minimum required	10

As a result, all evaluations that receive 25 points, with 10 of these points for Criteria G, H and I, should be included in the review.

5.0 Scaling and Classifying Evaluation Findings

5.1 Scale for Evaluation Findings

The technical team considered different options for scaling evaluation findings as they are found in the evaluation reports under review. The first was to follow the cumulative scaling guidelines used for both the MOPAN survey and for the MOPAN document review as detailed in their accompanying guide. MOPAN uses a six-point scale from a high of 6 (very strong) to a low of 1 (very weak) with each point on the scale defined as follows:

- (6) Very strong = The MO’s practice, behaviour or system is “best practice” in this area.
- (5) Strong = The MO’s practice, behaviour or system is more than acceptable yet without being “best practice” in this area.
- (4) Adequate = The MO’s practice, behaviour or system is acceptable in this area.
- (3) Inadequate = The MO’s practice, behaviour or system in this area has deficiencies that make it less than acceptable.
- (2) Weak = The MO has this practice, behaviour or system but there are important deficiencies.
- (1) Very weak = The MO does not have this practice, behaviour or system in place and this is a source of concern.

The MOPAN *Document Review Guide* goes further in that it specifies the elements which contribute in an additive way to each score for each of the Micro-Indicators to be assessed in the review of documents. This approach (while it is very refined in terms of

specifying six levels and very precise in its approach to assigning scores) would be very difficult to apply to the review of evaluation reports. The review of evaluations is not concerned with the elements which are present or not present in the evaluation report (which will be covered during the quality assessment and screening phase).⁶

Rather this approach focuses on the content and meaning of evaluation findings and conclusions in relation to the criteria used. In assessing the range of evaluation findings from positive to negative, it is not useful to take an additive approach since one line of evidence may be weak or inconclusive while another is demonstrably conclusive and supported by irrefutable evidence.

For that reason, similar efforts in the past⁷ have tended to adopt a rating scheme similar to those used by International Financial Institutions including the World Bank and the Asian Development Bank in their own processes of reviewing the effectiveness of projects and the production of Project Completion Reports. The scale used in the 2006 Review of Evidence of the Effectiveness of CIDA's Grants and Contributions was a six-point scale from Outstanding (6) to Highly Unsatisfactory (1); this was the same scale as used by the World Bank. The 2009 Review of the Effectiveness of CIDA's Multilateral Delivery Channel began with a five-point scale from Outstanding to Highly Unsatisfactory but subsequently collapsed the scale to four points as there were no observations meeting the criteria defined for Outstanding.

Building on that experience the Management Team indicated that a four-point scale would be appropriate. The scale is defined as:

(4) = Highly Satisfactory

(3) = Satisfactory

Tip

The key to an appropriate use of the operational guidelines is to see them as a tool for ensuring relevance to the organization and consistency across reviewers. As such the wording of the guidelines for each level (“highly satisfactory”, “satisfactory”, etc.) can be modified to include concepts and language that are both meaningful for the organization being reviewed and the particular review team.

⁶ A new version of the MOPAN document review methodology is expected in early June, any changes in that document will be considered as the Approach evolves.

⁷ Evaluation Directorate, CIDA. Review of Evidence of the Effectiveness of CIDA's Grants and Contributions, Ottawa 2006 and Review of the Effectiveness of CIDA's Multilateral Delivery Channel. Ottawa. 2009.

(2) = Unsatisfactory

(1) = Highly Unsatisfactory

Not addressed is coded as (0).

The MOPAN *Document Review Guide* provides a general definition of each of the levels, independent of the indicator being assessed. This is possible because the MOPAN process focuses on a comparative assessment of the MO's practice, systems or behaviours which can be readily expressed in terms of existence or non existence and, if they do exist, with acceptable, good or best practice.

This approach is not appropriate for classifying findings on evaluation criteria since the experience suggests that the level of a positive or negative finding tends to be defined in terms specific to the criteria under assessment and/or the methodologies used.

Objectives achievement for example can be addressed in terms of percentage of defined objectives achieved and/or in terms of reach (percentage of beneficiaries served) or timeliness (percentage of milestones reach on time).

The appropriate scoring levels are defined in terms specific to each sub-criterion. For any criteria not addressed in the evaluation report the coding used will be 0, not addressed.

5.2 Operational Guidelines for Classifying Evaluation Findings

The guidelines presented in Table 4 below are also provided in a more compact form for use by analysis team members in Annex 2. They are meant to be flexible in use and can be adapted to the realities of the structure, mandate and program of each MO by each team which undertakes a development effectiveness review. While a team is free to make reasonable modifications to the guidelines for classifying evaluation findings for each of the 19 sub-criterion, it should not modify the criteria and sub-criteria themselves (so that a sufficient level of consistency can be attained for future application across organizations). Although the focus of this review process is on development effectiveness, the guidelines include also references to the assessment of humanitarian effectiveness.⁸

⁸ In order to ensure the applicability of the review framework to organizations engaged in humanitarian action.

Table 5: Operational Guidelines for Classifying Evaluation Findings

Common Criteria for Assessing Development Effectiveness	
1. Achievement of Development Objectives and Expected Results	
1.1 MO supported programs and projects achieve their stated development and/or humanitarian objectives and attain expected results	<p>(4) Highly satisfactory MO supported programs and projects achieve all or almost all significant development and/or humanitarian objectives at the output and outcome level.</p> <p>(3) Satisfactory MO supported programs and projects either achieve at least a majority of stated output and outcome objectives (more than 50% if stated) or that the most important of stated output and outcome objectives are achieved.</p> <p>(2) Unsatisfactory Half or less than half of stated output and outcome level objectives are achieved.</p> <p>(1) Highly unsatisfactory Less than half of stated output and outcome objectives have been achieved including one or more very important output and/or outcome level objectives.</p> <p>(0) Not addressed</p>
1.2 MO supported programs and projects have resulted in positive benefits for target group members	<p>(4) Highly satisfactory MO supported projects and programs have resulted in widespread and significant positive changes experienced by target group members as measured using either quantitative or qualitative methods (possibly including comparison of impacts with non-program participants). These benefits may include the avoidance or reduction of negative effects of a sudden onset or protracted emergency.</p> <p>(3) Satisfactory MO supported projects and programs have resulted in positive changes experienced by target group members (at the individual, household or community level). These benefits may include the avoidance or reduction of negative effects of a sudden onset or protracted emergency.</p> <p>(2) Unsatisfactory MO supported projects and programs result in no or very few positive changes experienced by target group members. These benefits may include the avoidance or reduction of negative effects of a sudden onset or protracted emergency.</p> <p>(1) Highly unsatisfactory Problems in the design or delivery of MO supported activities mean that expected positive benefits for target group members have not occurred or are unlikely to occur.</p>

Common Criteria for Assessing Development Effectiveness	
(0)	Not addressed
1.3 MO programs and projects made differences for a substantial number of beneficiaries and where appropriate contributed to national development goals	
(4)	Highly satisfactory MO supported projects and programs have contributed to positive changes in the lives of substantial numbers of beneficiaries. Further, they have contributed to the achievement of specific national development goals or have contributed to meeting humanitarian relief objectives agreed to with the national government and/or national and international development and relief organizations.
(3)	Satisfactory MO supported projects and programs have contributed to positive changes in the lives of substantial numbers of beneficiaries as measured quantitatively or qualitatively. These may result from development, relief, or protracted relief and rehabilitation operations and may include the avoidance of negative effects of emergencies.
(2)	Unsatisfactory MO supported projects and programs have contributed to positive changes in the lives of only a small number of beneficiaries (when compared to project or program targets and local or national goals if established).
(1)	Highly unsatisfactory MO supported projects and programs have not contributed to positive changes in the lives of beneficiaries as measured quantitatively or qualitatively.
(0)	Not addressed
1.4 Extent MO activities have contributed to significant changes in national development policies and programs (policy impacts), or needed system reforms	
(4)	Highly satisfactory MO activities have made a substantial contribution to either re-orienting or sustaining effective national policies and programs in a given sector or area of development disaster preparedness, emergency response or rehabilitation. Further, the supported policies and program implementation modalities are expected to result in improved positive impacts for target group members.
(3)	Satisfactory MO activities have made a substantial contribution to either re-orienting or sustaining effective national policies and programs in a given sector or area of development disaster preparedness, emergency response or rehabilitation.
(2)	Unsatisfactory MO activities have not made a significant contribution to the development of national policies and programs in a given sector or area of development, disaster preparedness, emergency response or rehabilitation. (Policy changes in humanitarian situations may include allowing access to the effected populations).

Common Criteria for Assessing Development Effectiveness
<p>(1) Highly unsatisfactory National policies and programs in a given sector or area of development (including disaster preparedness, emergency response and rehabilitation) were deficient and required strengthening but MO activities have not addressed these deficiencies.</p> <p>(0) Not addressed</p>
<p>2. Cross-Cutting Themes: Inclusive Development Which is Environmentally Sustainable</p>
<p>2.1 Extent MO supported activities effectively address the cross-cutting issue of gender equality</p> <p>(4) Highly satisfactory MO supported programs and projects achieve all or nearly all of their stated gender equality objectives.</p> <p>(3) Satisfactory MO supported programs and projects achieve a majority (more than 50%) of their stated gender equality objectives.</p> <p>(2) Unsatisfactory MO supported activities either lack gender equality objectives or achieve less than half of their stated gender equality objectives. (Note: where a program or activity is clearly gender-focused (maternal health programming for example) achievement of more than half its stated objectives warrants a satisfactory rating).</p> <p>(1) Highly unsatisfactory MO supported activities are unlikely to contribute to gender equality or may in fact lead to increases in gender inequalities.</p> <p>(0) Not addressed</p>
<p>2.2 Extent changes are environmentally sustainable</p> <p>(4) Highly satisfactory MO supported programs and projects are specifically designed to be environmentally sustainable and include substantial planned activities and project design criteria to ensure environmental sustainability. These plans are implemented successfully and the results are environmentally sustainable.</p> <p>(3) Satisfactory MO supported programs and projects include some planned activities and project design criteria to ensure environmental sustainability. These activities are implemented successfully and the results are environmentally sustainable.</p> <p>(2) Unsatisfactory MO supported programs and projects do not include planned activities or project design criteria intended to promote environmental sustainability. There is, however, no direct</p>

Common Criteria for Assessing Development Effectiveness

indication that project or program results are not environmentally sustainable.

OR

MO supported programs and projects include planned activities or project design criteria intended to promote sustainability but these have not been successful.

(1) Highly unsatisfactory

MO supported programs and projects do not include planned activities or project design criteria intended to promote environmental sustainability. In addition changes resulting from MO supported programs and projects are not environmentally sustainable.

(0) Not addressed

3. Sustainability

3.1 Benefits continuing or likely to continue after project or program completion or there are effective measures to link the humanitarian relief operations, to rehabilitation, reconstructions and, eventually, to longer-term developmental results

(4) Highly satisfactory

Highly likely that the program or project will result in continued benefits for the target group after completion. For humanitarian relief operations, the strategic and operational measures to link relief to rehabilitation, reconstruction and, eventually, development are credible. Further, they are likely to succeed in securing continuing benefits for target group members.

(3) Satisfactory

Likely that the program or project will result in continued benefits for the target group after completion. For humanitarian relief operations, the strategic and operational measures to link relief to rehabilitation, reconstruction and, eventually, development are credible.

(2) Unsatisfactory

There is a low probability that the program/project will result in continued benefits for the target group after completion. For humanitarian relief operations, efforts to link the relief phase to rehabilitation, reconstruction and, eventually, to development are inadequate. (Note, in some circumstances such linkage may not be possible due to the context of the emergency. If this is stated in the evaluation, a rating of satisfactory can be given)

(1) Highly unsatisfactory

There is a very low probability that the program/project will result in continued intended benefits for the target group after project completion. For humanitarian relief operations, the evaluation finds no strategic or operational measures to link relief, to rehabilitation, reconstruction and, eventually, to development.

(0) Not addressed

Common Criteria for Assessing Development Effectiveness

3.2 Extent MO supported projects and programs are reported as sustainable in terms of institutional and/or community capacity

(4) Highly satisfactory

Either MO programs and projects have contributed to significantly strengthen institutional and/or community capacity as required or institutional partners and communities already had the required capacity to sustain program results.

(3) Satisfactory

MO programs and projects may have contributed to strengthening institutional and/or community capacity but with limited success.

(2) Unsatisfactory

MO programs and projects may have failed to contribute to strengthening institutional and/or community capacity or, where appropriate, to strengthen local capacities for delivery of relief operations and/or for managing the transition to rehabilitation and/or development.

(1) Highly unsatisfactory

The design of MO supported programs and projects failed to address the need to strengthen institutional and/or community capacity as required. In the case of humanitarian operations, the design of programs and projects failed to take account of identified needs to strengthen local capacities for delivery of relief operations and/or for managing the transition to rehabilitation and/or development.

(0) Not addressed

3.3 Extent MO development programming contributes to strengthening the enabling environment for development

(4) Highly satisfactory

MO development activities and/or MO supported projects and programs have made a significant contribution to changes in the enabling environment for development including one or more of: the overall framework and process for national development planning; systems and processes for public consultation and for participation by civil society in development planning; governance structures and the rule of law; national and local mechanisms for accountability for public expenditures, service delivery and quality; and necessary improvements to supporting structures such as capital and labour markets. Further, these improvements in the enabling environment are leading to improved development outcomes.

(3) Satisfactory

MO development activities and/or MO supported projects and programs have made a notable contribution to changes in the enabling environment for development including one or more of: the overall framework and process for national development planning; systems and processes for public consultation and for participation by civil society in development planning; governance structures and the rule of law; national and local

Common Criteria for Assessing Development Effectiveness

mechanisms for accountability for public expenditures, service delivery and quality; and necessary improvements to supporting structures such as capital and labour markets.

(2) Unsatisfactory

MO development activities and/or MO supported projects and programs have not made a notable contribution to changes in the enabling environment for development.

(1) Highly unsatisfactory

For development programs, there were important weaknesses in the enabling environment for development (the overall framework and process for national development planning; systems and processes for public consultation and for participation by civil society in development planning; governance structures and the rule of law; national and local mechanisms for accountability for public expenditures, service delivery and quality; and necessary improvements to supporting structures such as capital and labour markets). Further, the MO activities and support provided to programs and projects failed to address the identified weakness successfully, further limiting program results.

(0) Not addressed

4. Relevance

4.1 MO supported programs and projects are suited to the needs and/or priorities of the target group

(4) Highly satisfactory

Methods used in program and project development (including needs assessment for relief operations) to identify target group needs and priorities (including consultations with target group members) and the program and project takes those needs into account and is designed to meet those needs and priorities (whether or not it does so successfully).

(3) Satisfactory

MO supported activity, program or project is designed taking into account the needs of the target group as identified through a process of situation or problem analysis (including needs assessment for relief operations) and the resulting activities are designed to meet the needs of the target group.

(2) Unsatisfactory

No systematic analysis of target group needs and priorities took place during the design phase of developmental or relief and rehabilitation programming or there is some evident mismatch between program and project activities and outputs and the needs and priorities of the target group.

(1) Highly unsatisfactory

Substantial elements of program or project activities and outputs were unsuited to the needs and priorities of the target group.

(0) Not addressed

Common Criteria for Assessing Development Effectiveness

4.2 MO supported development projects and programs align with national development goals

(4) Highly satisfactory

All MO supported development projects and programs are reported in the evaluation to be fully aligned to national development goals as described in national and sector plans and priorities, especially including the national poverty eradication strategy and sector strategic priorities.

(3) Satisfactory

Most MO supported development programs and projects are fully aligned with national plans and priorities as expressed in national poverty eradication and sector plans and priorities. Wherever MO supported programs and projects are reported in the evaluation as not directly supportive of national plans and priorities they do not run counter to those priorities.

(2) Unsatisfactory

Significant portion (1/4 or more) of the MO supported development programs and projects are not aligned with national plans and priorities, but there is no evidence that they run counter to those priorities.

(1) Highly unsatisfactory

Significant elements of MO supported development program and project activity run counter to national development priorities with a resulting loss of effectiveness.

(0) Not addressed

4.3 MO has developed an effective partnership with governments, bilateral and multilateral development and humanitarian organizations and NGOs for planning, coordination and implementation of support to development and/or emergency preparedness, humanitarian relief and rehabilitation efforts

(4) Highly satisfactory

MO has consistently achieved a high level of partnership during the evaluation period.

(3) Satisfactory

MO has improved the effectiveness of its partnership relationship with partners over time during the evaluation period and that this partnership was effective at the time of the evaluation or was demonstrably improved.

(2) Unsatisfactory

MO has experienced significant difficulties in developing an effective relationship with partners and that there has been significant divergence in the priorities of the MO and its partners.

(1) Highly unsatisfactory

MO experiences significant divergence in priorities from those of its (government, NGO or donor) partners and lacks a strategy or plan which will credibly address the

Common Criteria for Assessing Development Effectiveness	
	divergence and which should result in strengthened partnership over time.
(0)	Not addressed
5.	Efficiency
5.1	Program activities are evaluated as cost/resource efficient
(4)	Highly satisfactory MO supported (development, emergency preparedness, relief and rehabilitation) programs and projects are designed to include activities and inputs that produce outputs in the most cost/resource efficient manner available at the time.
(3)	Satisfactory Level of program outputs achieved (development, emergency preparedness, relief and rehabilitation) when compared to the cost of program activities and inputs are appropriate even when the program design process did not directly consider alternative program delivery methods and their associated costs.
(2)	Unsatisfactory MO supported programs and projects under evaluation (development, emergency preparedness, relief and rehabilitation) do not have credible, reliable information on the costs of activities and inputs and therefore the evaluation is not able to report on cost/resource efficiency. OR MO supported programs and projects under evaluation present mixed findings on the cost/resource efficiency of the inputs.
(1)	Highly unsatisfactory Credible information indicating that MO supported programs and projects (development, emergency preparedness, relief and rehabilitation) are not cost/resource efficient.
(0)	Not addressed
5.2	Evaluation indicates implementation and objectives achieved on time (given the context, in the case of humanitarian programming)
(4)	Highly satisfactory Nearly all stated output and outcome level objectives of MO supported programs and projects are achieved on time or, in the case of humanitarian programming, a legitimate explanation for delays in the achievement of some outputs/outcomes is provided.
(3)	Satisfactory More than half of stated output and outcome level objectives of MO supported programs and projects are achieved on time and that this level is appropriate to the context faced by the program during implementation, particularly for humanitarian programming.
(2)	Unsatisfactory

Common Criteria for Assessing Development Effectiveness

Less than half of stated output and outcome level objectives of MO supported programs and projects are achieved on time but the program or project design has been adjusted to take account of difficulties encountered and can be expected to improve the pace of objectives achievement in the future. In the case of humanitarian programming, there was a legitimate explanation for the delays.

(1) Highly unsatisfactory

Less than half of stated output and outcome level objectives of MO supported programs and projects are achieved on time, there is no credible plan or legitimate explanation found by the evaluation which would suggest significant improvement in on-time objectives achievement in the future.

(0) Not addressed

5.3 Evaluation indicates that MO systems and procedures for project/program implementation and follow up are efficient (including systems for engaging staff, procuring project inputs, disbursing payment, logistical arrangements etc.)

(4) Highly satisfactory

Efficiency of agency systems and procedures for project implementation represent an important organizational strength in the implementation of the program under evaluation.

(3) Satisfactory

Agency systems and procedures for project implementation are reasonably efficient and have not resulted in significant delays or increased costs.

(2) Unsatisfactory

Some deficiencies in agency systems and procedures for project/program implementation but does not indicate that these have contributed to delays in achieving project/program objectives.

(1) Highly unsatisfactory

Serious deficiencies in agency systems and procedures for project/program implementation that result in significant delays in project start-up, implementation or completion and/or significant cost increases.

(0) Not addressed

6. Using Evaluation and Monitoring to Improve Development Effectiveness

6.1 Systems and process for evaluation are effective.

(4) Highly satisfactory

Program being evaluated (along with similar programs and projects) is subject to systematic regular evaluations or describes significant elements of such practice.

(3) Satisfactory

Common Criteria for Assessing Development Effectiveness

Program being evaluated is subject to systematic and regular evaluations or describes significant elements of such practice. No mention of policy and practice regarding similar programs and projects. This may include specialized evaluation methods and approaches to emergency preparedness, relief and rehabilitation programming.

(2) Unsatisfactory

No indication that programs and projects of this type (development, emergency preparedness, relief and rehabilitation) are subject to systematic and regular evaluations.

(1) Highly unsatisfactory

Evaluation practices in use for programs and projects of this type (development, emergency preparedness, relief and rehabilitation) are seriously deficient.

(0) Not addressed

6.2 Systems and processes for monitoring and reporting on program results are effective

(4) Highly satisfactory

Monitoring and reporting systems for the program are well-established and report regularly. The quality of regular reports is rated highly by the evaluation and results are reportedly used in the management of the program.

(3) Satisfactory

Monitoring and reporting systems for development and humanitarian programming as appropriate are well-established and report regularly.

(2) Unsatisfactory

While monitoring and reporting systems for the development and humanitarian programming exist, they either do not report on a regular basis or they are inadequate in frequency, coverage or reliability.

(1) Highly unsatisfactory

Absence of monitoring and reporting systems for the development and humanitarian programming. This would include the absence of adequate monitoring of outputs during the implementation of humanitarian programming.

(0) Not addressed

6.3 Results Based Management (RBM) systems are effective

(4) Highly satisfactory

RBM system is in place for the program and there is evidence noted in the evaluation that the system is used to make changes in the program to improve effectiveness.

(3) Satisfactory

RBM system is in place and produces regular reports on program performance.

(2) Unsatisfactory

Common Criteria for Assessing Development Effectiveness

While an RBM system is in place, or being developed, it is unreliable and does not produce regular reports on program performance.

(1) Highly unsatisfactory

No evidence of the existence of an RBM system for the program and no system is being developed.

(0) Not addressed

6.4 MO makes use of evaluation to improve development/humanitarian effectiveness

(4) Highly satisfactory

Report includes a management response (or has one attached or associated with it) describes a response to each major recommendation which is appropriate and likely to result in the organizational and programmatic changes needed to achieve their intent.

(3) Satisfactory

Report includes a management response (or has one attached or associated with it) that indicates which recommendations have been accepted.

OR

There is a clear indication that similar evaluations in the past have been used to make clearly identified improvements in program effectiveness.

(2) Unsatisfactory

Report includes a management response (or has one attached or associated with it) but it does not indicate which recommendations have been accepted.

OR

There is some, non-specific indication that similar evaluations have been used to improve program effectiveness in the past.

(1) Highly unsatisfactory

Report does not include a management response and does not have one appended to it or associated with it. There is no indication of how the evaluation results will be used. There is no indication that similar evaluations have been used to improve effectiveness in the past.

(0) Not addressed

6.0 Analysis of Effectiveness Results

6.1 Analysis of Quantitative Results

Once the review is completed, the scores and their supporting evidence should be entered into a database. In reporting the distribution of results for each sub-criterion, it is

necessary to identify the number and/or percentage, as appropriate, of evaluations that received a rating of highly satisfactory, satisfactory, unsatisfactory, highly unsatisfactory, as well as the number/percentage that do not address any given sub-criterion. It is important to be very transparent in reporting the number of evaluations that did not address any given sub-criterion, either in a tabular or graphical format. Analysis of the number of evaluations that did not address a sub-criterion is important and may, in fact, result in recommendations with respect to the coverage of the evaluations. If the coverage of the sub-criterion is too weak, results in this area should not be reported. For convenience in the later reporting of the findings, it is useful to convert the number of evaluations addressing a sub-criterion into qualitative ratings of “weak,” “moderate” and “high” coverage. Based on the experience of the previous reviews, it is suggested that if a sub-criterion is addressed in about three-quarters of the evaluations, the coverage is “strong”; between half and three-quarters of the evaluations, it is “moderate”; and in less than half the evaluations, it is “weak.”

Tip

At times it is difficult to identify the “bottom line” in the evaluation findings with respect to a given sub-criterion, particularly in evaluations that cover multiple components of the MO’s programming. Two sources of information can help to determine the “bottom line”, as defined by the evaluation authors – the executive summaries or summary reports of the evaluation and the evaluation recommendations. Although the analysis should be based on the full evaluation report, these can be used to identify the main findings with respect to development effectiveness.

Although the number of evaluation not addressing a sub-criterion is important, the analysis of the ratings (“highly satisfactory,” “satisfactory,” “unsatisfactory” or “highly unsatisfactory”) should be based on the number of evaluations that reported on the sub-criterion. It is expected that an MO will address all the sub-criteria across the range of evaluations, as they are indicators of the development effectiveness of the organization, as a whole. But it is not expected that each evaluation will address each sub-criterion, particularly given the range of types of evaluation that might be included in the sample (e.g. operational evaluations, policy evaluations, thematic evaluations, country program/portfolio evaluations).

The results for each sub-criterion are reportedly separately in a section dealing with the overall criteria. They are not rolled up into one rating for the criteria or an overall development effectiveness rating, since this would imply that the criteria and sub-criteria are mutually exclusive and of equal weight, which is not necessarily the case.

The analysis of the sixth criterion (“Using evaluation and monitoring to improve development effectiveness”) should be treated separately from the other criteria. It is an explanatory variable that contributes to an understanding of how the MO is learning and improving effectiveness, rather than being a direct measure of development effectiveness.

6.2 Factors Contributing to or Inhibiting Effectiveness

Under the general heading of Effectiveness, the DAC Criteria for Evaluating Development Assistance include the question:

“What were the major factors influencing the achievement or non-achievement of objectives?”

Tip

To facilitate the analysis of the contributing factors, the team could develop rolling codes for classifying and counting the factors. These codes would be first developed once a significant number of evaluations have been reviewed and continue to evolve as the reviews continue.

The experience of the pilot test indicates that many evaluations do not explicitly identify factors which contributed to (or impeded) the achievement of development objectives. Rather, many evaluation reports tend to interweave the actual evaluation evidence cited for a particular finding with some indications of what may have contributed to the findings. This makes the identification of factors (and quantifying the results) a more challenging analytical problem.

The pilot test team addressed this challenge by producing a grid for each evaluation which included (for every assessment criteria) not only the classification of the finding (highly satisfactory, satisfactory, etc.) but the main supporting evidence (complete with page references). This evidence then became the main source of information on contributing factors (especially where contributing factors were not specifically identified and reported on in the evaluation).

By reviewing the evidence for each finding in every evaluation, the analytical team is able to list the contributing factors wherever they are cited in evaluation evidence. These can then be used to compile frequency counts and to identify the most frequently cited factors contributing to (or impeding) positive results.

Another approach to the analysis of the programming context and how evaluations are used to improve program results would be to examine the recommendations made in each evaluation report and to classify them using an analysis grid. One such coding

format would classify each recommendation made in an evaluation report under one of the following headings:

- Program or project design;
- Results framework or statement of expected results;
- Financial resource issues;
- Human resource issues;
- Implementation challenges;
- Objectives (realism, clarity, etc.);
- Oversight/governance;
- Organizational or program strategy;
- Policy-related (gender equality, poverty alleviation, environment, etc.); and,
- Risk management.

7.0 Organizing and Carrying Out the Development Effectiveness Assessment

7.1 Lead and Supporting Agencies and Their Roles

In order to effectively carry out an assessment of the development effectiveness of a given MO, it will be necessary to identify both the lead and supporting agencies (if available) and to specify their roles. The optimal choices in this area seem to be based mainly on practical considerations. If the approach is to be applied to more than two MOs in future years, it seems unlikely that three or more Network members can provide resources to each one. On the other hand, there is certainly value in having more than one agency conduct the review.

In principle, as with MOPAN, there could be a lead and a supporting agency for each development effectiveness assessment but this would not exclude either a larger grouping (a lead agency, with two supporting agencies for example) or a single agency undertaking a review if there was no interest in support from other agencies.

When more than one Network member agency is undertaking a development effectiveness assessment, the report to the Network (represented by either the Management Group or the Task Team) should be made on behalf of the lead agency

with operational support from other participating agencies. Thus the lead agency provides team leadership and provides the lead author for the draft and final reports. The reports themselves would be issued by the lead agency.

Finally, when a lead and supporting agencies are identified to undertake a particular development effectiveness assessment, they will begin by carrying out the Preliminary Review identified in Section 1.3 above.

Tip

In order to ensure that the team leader has sufficient evidence from the evaluations for reporting, the reviewers need to be aware of the analysis plan before beginning the review of the evaluation reports.

7.2 Carrying out a Meta-Synthesis of Evaluations

The process of carrying out the meta-synthesis of evaluations when such action is called for under Scenario 2 and Option 2 of the Framework should be characterized by the following essential elements.

1. Each Assessment should be led by an Evaluation Manager from one of the participating agencies.
2. The evaluations to be reviewed should be selected based on sampling criteria agreed in Section 4.0.
3. The quality assurance screening and the review of the findings reported in the chosen evaluations should be carried out by a small team of two to four analysts working under the supervision of a senior evaluation specialist.
4. The most important challenge in carrying out the meta-synthesis review of evaluations (as identified in the pilot test process) is maintaining consistency among the analysts as they review and classify evaluation findings under each criterion for each evaluation. The pilot test results indicate that a good level of consistency and inter-reviewer reliability can be achieved by undertaking the following steps:

Tip

To ensure consistency among the reviewers, it is suggested that the team hold regular meetings or conference calls during the process to compare notes on the reviews.

- a) The team of evaluation analysts undergoes training on the application of the analysis grid and the summary of contributing or impeding factors. This is best done in workshop setting for a full day.
 - b) The team members and the senior evaluation specialist/team leader carry out a review of the same evaluation report(s) at specific points in the process. Each time they will meet (in person or by telephone) and discuss the resulting classification of evaluation findings for each of the criteria and agree on the most appropriate response. At the beginning of the process this should be done for two evaluations. This will increase the reviewers' confidence that they are able to apply the criteria in a way which is consistent with practice by other team members. The same process should be repeated at the approximate mid-point in the evaluation review process.
 - c) The senior evaluation specialist should review all completed assessment grids and discuss any anomalies in classification with the reviewer in question. If needed he/she can re-score the evaluation reports in question.
5. At the end of the review process, the senior evaluation specialist reviews a print-out of all the results and ensures that there is consistency between the ratings and the evidence provided by the reviewers.

Tip

To address the level of inter-reviewer reliability, a tool could be used to measure quantitatively the difference between individual reviewer ratings and the consensus rating of the team for the evaluations they have reviewed by each member of the review team. As more evaluations are reviewed jointly, the gap between individual reviewers and the consensus rating should decrease. These ratings would be useful data to report in order to reflect the level of inter-reviewer reliability and strengthen the credibility of the report.

8.0 Development Effectiveness Assessment Report

The structure and content of the report on each MO should include fairly detailed information on, for example choice of the MO, sampling of evaluations, development effectiveness Assessment process and limitations. The latter will include, for example, difficulties in obtaining evaluations or criteria which are systematically either not covered by evaluations or covered poorly.

The findings sections of the report on each MO will cover each of the key dimensions of development effectiveness as well as their related assessment criteria. The report will also include an analysis of factors contributing to or impeding development effectiveness as cited in the evaluations reviewed.

The overall length of the Development Effectiveness Assessment Reports on each MO should be no more than 30 pages with a five page executive summary and any necessary technical annexes.

The basic structure of the report will be as follows:

Acknowledgment (1 page)

Acronyms (1 page)

Executive Summary (3 pages)

Introduction (1 page)

Methodology (2-3 pages)

- Sampling process
- Size and representativeness of chosen sample
- Quality Assurance Results
- Coverage of the specified evaluation criteria
- Review process and lessons learned
- Limitations

Findings on Development Effectiveness (15-20 pages)

- Achievement of Development Objectives/Expected Results
- Cross Cutting Themes: Gender Equality and Environment
- Sustainability
- Relevance

Tip

Sufficient time for the analysis and reporting should be allowed, as the qualitative analysis is particularly time-consuming. In addition, if errors in the ratings are identified by the senior evaluation specialist during the qualitative review, it is important to have the time to perform data analysis again, with the corrections made.

Tip

The report provides a more interesting story of the MO's performance if it includes boxes to highlight specific examples from the evaluations.

- Efficiency
- Using Evaluation and Monitoring to Improve Development Effectiveness

Factors Contributing to Development Effectiveness (2-3 pages)

Recommendations

Annex 1: Terms of Reference

Annex 2: Evaluation Review Grid

Annex 3: List of Evaluations Reviewed

Tip

Previous reviews suggest that the analysis of the contributing factors can usefully be integrated with the quantitative findings on development effectiveness in order to tell a more complete story about the MO's performance.

Annex 1: Evaluation Quality Screening Scoring Guide for Multilateral Organizations

Evaluation Title: _____

Reviewer: _____

	Points for criteria scored	Maximum Points	Score
A	<p>Purpose of the evaluation:</p> <ul style="list-style-type: none"> - why the evaluation was done (1) - what triggered the evaluation (including timing in the project/program cycle) (1) - how evaluation is to be used (1) 	3	
B	<p>Evaluation objectives</p> <ul style="list-style-type: none"> - evaluation objectives are clearly stated (1) - objectives logically flow from purpose (1) 	2	
C	<p>Organization of the evaluation</p> <ul style="list-style-type: none"> - logical structure to the organization of the evaluation (1) - evaluation is well written (1) - clear distinction between evidence, findings, conclusions, and recommendations (1) 	3	
D	<p>Subject evaluated is clearly described</p> <p>Evaluation describes:</p> <ul style="list-style-type: none"> - the activity/program being evaluated (1) - the program's expected achievements (1) - how the program addresses the development problem (1) - the implementation modalities used (1) 	4	
E	<p>Scope of the evaluation</p> <p>Evaluation defines the boundaries of the evaluation in terms of:</p> <ul style="list-style-type: none"> - time period covered (1) - implementation phase under review (1) - geographic area (1) - dimensions of stakeholder involvement being examined (1) 	4	
F	<p>Evaluation criteria</p> <p>Evaluation criteria include:</p> <ul style="list-style-type: none"> - the achievement of development objectives and expected results (including impacts) (1) - cross-cutting issues: inclusive development which is gender sensitive and environmentally sustainable (1) 	5	

	Points for criteria scored	Maximum Points	Score
	<ul style="list-style-type: none"> - the sustainability of benefits and positive results achieved (1) - the relevance of MO activities and supported projects and programs (1) - the efficiency of MO operations in support of projects and programs (1) 		
G	<p>Multiple lines of evidence</p> <ul style="list-style-type: none"> - one point (1) for each line of evidence used (case studies, surveys, site visits, interviews, etc.), up to a maximum of four points (4) 	4	
H	<p>Evaluation design</p> <p>Elements of a good evaluation design include:</p> <ul style="list-style-type: none"> - an explicit theory of how objectives and results were to be achieved (1) - specification of the level of results achieved (output, outcome, impact) (1) - baseline data (quantitative or qualitative) on conditions prior to program implementation (1) - comparison of conditions after program delivery to those before (1) - a qualitative or quantitative comparison of conditions among program participants and a control group (1) 	5	
I	<p>Evaluation findings and conclusions are relevant and evidence based</p> <p>Evaluation report includes:</p> <ul style="list-style-type: none"> - evaluation findings relevant to the assessment criteria (1) - findings that are supported by the chosen methodology (1) - a clear logical link between the evidence and the finding (1) - conclusions which are linked to the evaluation findings as reported (1) 	4	
J	<p>Evaluation limitations</p> <ul style="list-style-type: none"> - statement of the limitations of the methodology (1) - impact of limitations on evaluation (1) - remedies of limitations (1) 	3	
K	<p>Evaluation Recommendations</p> <ul style="list-style-type: none"> - evaluation contains recommendations that flow from findings and conclusions (1) - recommendations are directed to one or more organization (1) - recommendations are aimed at improving development effectiveness (1) 	3	
	Total (required to have a minimum of 25 points)	40	
	Total for Criteria G, H and I (required to have minimum of	13	

Assessing the Development Effectiveness of Multilateral Organizations: Guidance on the Methodological Approach

	Points for criteria scored	Maximum Points	Score
	10 points)		

Annex 2: Operational Guidelines for Classifying Evaluation Findings⁹

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
Common Development Evaluation Assessment Criteria				
1. Achievement of Development Objectives and Expected Results				
1.1 MO supported programs and projects achieve their stated development and/or humanitarian objectives and attain expected results.	Less than half of stated output and outcome objectives have been achieved including one or more very important output and/or outcome level objectives.	Half or less than half of stated output and outcome level objectives are achieved.	MO supported programs and projects either achieve at least a majority of stated output and outcome objectives (more than 50% if stated) or that the most important of stated output and outcome objectives are achieved.	MO supported programs and projects achieve all or almost all significant development and/or humanitarian objectives at the output and outcome level.
1.2 MO supported programs and projects have resulted in positive benefits for target group members.	Problems in the design or delivery of MO supported activities mean that expected positive benefits for target group members have not occurred or are unlikely to occur.	MO supported projects and programs result in no or very few positive changes experienced by target group members. These benefits may include the avoidance or reduction of negative effects of a sudden onset or protracted	MO supported projects and programs have resulted in positive changes experienced by target group members (at the individual, household or community level). These benefits may include the avoidance or reduction of negative	MO supported projects and programs have resulted in widespread and significant positive changes experienced by target group members as measured using either quantitative or qualitative methods (possibly including comparison of impacts with non-program participants). These

⁹ These operational guidelines are meant to be flexible and can be adapted to the context of the MO and the review team. However, while a team is free to make reasonable modifications to the guidelines for each of the sub-criterion, it should not modify the criteria and sub-criteria themselves.

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
		emergency.	effects of a sudden onset or protracted emergency.	benefits may include the avoidance or reduction of negative effects of a sudden onset or protracted emergency.
1.3 MO programs and projects made differences for a substantial number of beneficiaries and where appropriate contributed to national development goals.	MO supported projects and programs have not contributed to positive changes in the lives of beneficiaries as measured quantitatively or qualitatively.	MO supported projects and programs have contributed to positive changes in the lives of only a small number of beneficiaries (when compared to project or program targets and local or national goals if established).	MO supported projects and programs have contributed to positive changes in the lives of substantial numbers of beneficiaries as measured quantitatively or qualitatively. These may result from development, relief, or protracted relief and rehabilitation operations and may include the avoidance of negative effects of emergencies.	MO supported projects and programs have contributed to positive changes in the lives of substantial numbers of beneficiaries. Further, they have contributed to the achievement of specific national development goals or have contributed to meeting humanitarian relief objectives agreed to with the national government and/or national and international development and relief organizations.
1.4 MO activities contributed to significant changes in national development policies and programs (including for disaster preparedness, emergency response and rehabilitation) (policy impacts) and/or to needed system reforms.	National policies and programs in a given sector or area of development (including disaster preparedness, emergency response and rehabilitation) were deficient and required strengthening but MO activities have not	MO activities have not made a significant contribution to the development of national policies and programs in a given sector or area of development, disaster preparedness, emergency response or rehabilitation. (Policy	MO activities have made a substantial contribution to either re-orienting or sustaining effective national policies and programs in a given sector or area of development disaster preparedness, emergency response or rehabilitation.	MO activities have made a substantial contribution to either re-orienting or sustaining effective national policies and programs in a given sector or area of development disaster preparedness, emergency response or rehabilitation. Further, the supported policies and program implementation

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
	addressed these deficiencies.	changes in humanitarian situations may include allowing access to the effected populations).		modalities are expected to result in improved positive impacts for target group members.
2. Cross Cutting Themes: Inclusive Development Which can be Sustained				
2.1 Extent MO supported activities effectively address the cross-cutting issue of gender equality.	MO supported activities are unlikely to contribute to gender equality or may in fact lead to increases in gender inequalities.	MO supported activities either lack gender equality objectives or achieve less than half of their stated gender equality objectives. (Note: where a program or activity is clearly gender-focused (maternal health programming for example) achievement of more than half its stated objectives warrants a satisfactory rating).	MO supported programs and projects achieve a majority (more than 50%) of their stated gender equality objectives.	MO supported programs and projects achieve all or nearly all of their stated gender equality objectives.
2.2 Extent changes are environmentally sustainable.	MO supported programs and projects do not include planned activities or project design criteria intended to promote environmental	MO supported programs and projects do not include planned activities or project design criteria intended to promote	MO supported programs and projects include some planned activities and project design criteria to ensure environmental sustainability. These	MO supported programs and projects are specifically designed to be environmentally sustainable and include substantial planned activities and project design criteria to ensure environmental

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
	sustainability. In addition changes resulting from MO supported programs and projects are not environmentally sustainable.	environmental sustainability. There is, however, no direct indication that project or program results are not environmentally sustainable. OR MO supported programs and projects include planned activities or project design criteria intended to promote sustainability but these have not been successful.	activities are implemented successfully and the results are environmentally sustainable.	sustainability. These plans are implemented successfully and the results are environmentally sustainable.
3. Sustainability				
3.1 Benefits continuing or likely to continue after project or program completion or there are effective measures to link the humanitarian relief operations, to rehabilitation, reconstructions and, eventually, to longer-term developmental results.	There is a very low probability that the program/project will result in continued intended benefits for the target group after project completion. For humanitarian relief operations, the evaluation finds no strategic or operational measures to link relief, to	There is a low probability that the program/project will result in continued benefits for the target group after completion. For humanitarian relief operations, efforts to link the relief phase to rehabilitation, reconstruction and, eventually, to	Likely that the program or project will result in continued benefits for the target group after completion. For humanitarian relief operations, the strategic and operational measures to link relief to rehabilitation, reconstruction and, eventually, development	Highly likely that the program or project will result in continued benefits for the target group after completion. For humanitarian relief operations, the strategic and operational measures to link relief to rehabilitation, reconstruction and, eventually, development are credible. Further, they are likely to succeed in securing continuing benefits for target group members.

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
	rehabilitation, reconstruction and, eventually, to development.	development are inadequate. (Note, in some circumstances such linkage may not be possible due to the context of the emergency. If this is stated in the evaluation, a rating of satisfactory can be given)	are credible.	
3.2 Extent MO supported projects and programs are reported as sustainable in terms of institutional and/or community capacity.	The design of MO supported programs and projects failed to address the need to strengthen institutional and/or community capacity as required. In the case of humanitarian operations, the design of programs and projects failed to take account of identified needs to strengthen local capacities for delivery of relief operations and/or for managing the transition to rehabilitation and/or development.	MO programs and projects may have failed to contribute to strengthening institutional and/or community capacity or, where appropriate, to strengthen local capacities for delivery of relief operations and/or for managing the transition to rehabilitation and/or development.	MO programs and projects may have contributed to strengthening institutional and/or community capacity but with limited success.	Either MO programs and projects have contributed to significantly strengthen institutional and/or community capacity as required or institutional partners and communities already had the required capacity to sustain program results.
3.3 Extent MO development	For development programs, there were	MO development activities and/or MO	MO development activities and/or MO	MO development activities and/or MO supported projects

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
<p>programming contributes to strengthening the enabling environment for development.</p>	<p>important weaknesses in the enabling environment for development (the overall framework and process for national development planning; systems and processes for public consultation and for participation by civil society in development planning; governance structures and the rule of law; national and local mechanisms for accountability for public expenditures, service delivery and quality; and necessary improvements to supporting structures such as capital and labour markets). Further, the MO activities and support provided to programs and projects failed to address the identified weakness successfully, further limiting program results.</p>	<p>supported projects and programs have not made a notable contribution to changes in the enabling environment for development.</p>	<p>supported projects and programs have made a notable contribution to changes in the enabling environment for development including one or more of: the overall framework and process for national development planning; systems and processes for public consultation and for participation by civil society in development planning; governance structures and the rule of law; national and local mechanisms for accountability for public expenditures, service delivery and quality; and necessary improvements to supporting structures such as capital and labour markets.</p>	<p>and programs have made a significant contribution to changes in the enabling environment for development including one or more of: the overall framework and process for national development planning; systems and processes for public consultation and for participation by civil society in development planning; governance structures and the rule of law; national and local mechanisms for accountability for public expenditures, service delivery and quality; and necessary improvements to supporting structures such as capital and labour markets. Further, these improvements in the enabling environment are leading to improved development outcomes.</p>

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
4. Relevance				
4.1 MO supported programs and projects are suited to the needs and/or priorities of the target group	Substantial elements of program or project activities and outputs were unsuited to the needs and priorities of the target group.	No systematic analysis of target group needs and priorities took place during the design phase of developmental or relief and rehabilitation programming or there is some evident mismatch between program and project activities and outputs and the needs and priorities of the target group.	MO supported activity, program or project is designed taking into account the needs of the target group as identified through a process of situation or problem analysis (including needs assessment for relief operations) and the resulting activities are designed to meet the needs of the target group.	Methods used in program and project development (including needs assessment for relief operations) to identify target group needs and priorities (including consultations with target group members) and the program and project takes those needs into account and is designed to meet those needs and priorities (whether or not it does so successfully).
4.2 MO supported development projects and programs align with national development goals:	Significant elements of MO supported development program and project activity run counter to national development priorities with a resulting loss of effectiveness.	<u>Significant portion</u> (1/4 or more) of the MO supported development programs and projects are not aligned with national plans and priorities, but there is no evidence that they run counter to those priorities.	<u>Most</u> MO supported development programs and projects are fully aligned with national plans and priorities as expressed in national poverty eradication and sector plans and priorities. Wherever MO supported programs and projects are reported in the evaluation as not directly supportive of national plans and priorities they do not run	<u>All</u> MO supported development projects and programs are reported in the evaluation to be fully aligned to national development goals as described in national and sector plans and priorities, especially including the national poverty eradication strategy and sector strategic priorities.

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
			counter to those priorities.	
4.3 MO has developed an effective partnership with governments, bilateral and multilateral development and humanitarian organizations and NGOs for planning, coordination and implementation of support to development and/or emergency preparedness, humanitarian relief and rehabilitation efforts.	MO experiences significant divergence in priorities from those of its (government, NGO or donor) partners and lacks a strategy or plan which will credibly address the divergence and which should result in strengthened partnership over time.	MO has experienced significant difficulties in developing an effective relationship with partners and that there has been significant divergence in the priorities of the MO and its partners.	MO has improved the effectiveness of its partnership relationship with partners over time during the evaluation period and that this partnership was effective at the time of the evaluation or was demonstrably improved.	MO has consistently achieved a high level of partnership during the evaluation period.
5. Efficiency				
5.1 Program activities are evaluated as cost/resource efficient:	Credible information indicating that MO supported programs and projects (development, emergency preparedness, relief and rehabilitation) are not cost/resource efficient.	MO supported programs and projects under evaluation (development, emergency preparedness, relief and rehabilitation) do not have credible, reliable information on the costs of activities and inputs and therefore the evaluation is not able to report on cost/resource	Level of program outputs achieved (development, emergency preparedness, relief and rehabilitation) when compared to the cost of program activities and inputs are appropriate even when the program design process did not directly consider alternative program delivery methods and their associated costs.	MO supported (development, emergency preparedness, relief and rehabilitation) programs and projects are designed to include activities and inputs that produce outputs in the most cost/resource efficient manner available at the time.

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
		efficiency. OR MO supported programs and projects under evaluation present mixed findings on the cost/resource efficiency of the inputs.		
5.2 Evaluation indicates implementation and objectives achieved on time (given the context, in the case of humanitarian programming)	Less than half of stated output and outcome level objectives of MO supported programs and projects are achieved on time, there is no credible plan or legitimate explanation found by the evaluation which would suggest significant improvement in on-time objectives achievement in the future.	Less than half of stated output and outcome level objectives of MO supported programs and projects are achieved on time but the program or project design has been adjusted to take account of difficulties encountered and can be expected to improve the pace of objectives achievement in the future. In the case of humanitarian programming, there was a legitimate explanation for the delays.	More than half of stated output and outcome level objectives of MO supported programs and projects are achieved on time and that this level is appropriate to the context faced by the program during implementation, particularly for humanitarian programming.	Nearly all stated output and outcome level objectives of MO supported programs and projects are achieved on time or, in the case of humanitarian programming, a legitimate explanation for delays in the achievement of some outputs/outcomes is provided.
5.3 Evaluation indicates that MO systems and procedures for	Serious deficiencies in agency systems and procedures for	Some deficiencies in agency systems and procedures for	Agency systems and procedures for project implementation are	Efficiency of agency systems and procedures for project implementation represent an

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
project/program implementation and follow up are efficient (including systems for engaging staff, procuring project inputs, disbursing payment, logistical arrangements etc.)	project/program implementation that result in significant delays in project start-up, implementation or completion and/or significant cost increases.	project/program implementation but does not indicate that these have contributed to delays in achieving project/program objectives.	reasonably efficient and have not resulted in significant delays or increased costs.	important organizational strength in the implementation of the program under evaluation.
6. Using Evaluation and Monitoring to Improve Development Effectiveness				
6.1 Systems and process for evaluation are effective.	Evaluation practices in use for programs and projects of this type (development, emergency preparedness, relief and rehabilitation) are seriously deficient.	No indication that programs and projects of this type (development, emergency preparedness, relief and rehabilitation) are subject to systematic and regular evaluations.	Program being evaluated is subject to systematic and regular evaluations or describes significant elements of such practice. No mention of policy and practice regarding similar programs and projects. This may include specialized evaluation methods and approaches to emergency preparedness, relief and rehabilitation programming.	Program being evaluated (along with similar programs and projects) is subject to systematic regular evaluations or describes significant elements of such practice.
6.2 Systems and processes for monitoring and reporting on program results are effective	Absence of monitoring and reporting systems for the development and humanitarian programming. This would	While monitoring and reporting systems for the development and humanitarian programming exist,	Monitoring and reporting systems for development and humanitarian programming as appropriate are well-	Monitoring and reporting systems for the program are well-established and report regularly. The quality of regular reports is rated highly by the evaluation

Criteria	(1) Highly Unsatisfactory	(2) Unsatisfactory	(3) Satisfactory	(4) Highly Satisfactory
	include the absence of adequate monitoring of outputs during the implementation of humanitarian programming.	they either do not report on a regular basis or they are inadequate in frequency, coverage or reliability.	established and report regularly.	and results are reportedly used in the management of the program.
6.3 Results Based Management (RBM) systems are effective	No evidence of the existence of an RBM system for the program and no system is being developed.	While an RBM system is in place, or being developed, it is unreliable and does not produce regular reports on program performance.	RBM system is in place and produces regular reports on program performance.	RBM system is in place for the program and there is evidence noted in the evaluation that the system is used to make changes in the program to improve effectiveness.
6.4 MO makes use of evaluation to improve development/humanitarian effectiveness	Report does not include a management response and does not have one appended to it or associated with it. There is no indication of how the evaluation results will be used. There is no indication that similar evaluations have been used to improve effectiveness in the past.	Report includes a management response (or has one attached or associated with it) but it does not indicate which recommendations have been accepted. OR There is some, non-specific indication that similar evaluations have been used to improve program effectiveness in the past.	Report includes a management response (or has one attached or associated with it) that indicates which recommendations have been accepted. OR There is a clear indication that similar evaluations in the past have been used to make clearly identified improvements in program effectiveness.	Report includes a management response (or has one attached or associated with it) describes a response to each major recommendation which is appropriate and likely to result in the organizational and programmatic changes needed to achieve their intent.