# Sourcebook for Evaluating Global and Regional Partnership Programs

## Indicative Principles and Standards

IEG
INDEPENDENT EVALUATION GROUP

DAC NETWORK ON DEVELOPMENT EVALUATION

# THE INDEPENDENT EVALUATION GROUP

The Independent Evaluation Group (IEG) is an independent, three-part unit within the World Bank Group. IEG–World Bank is charged with evaluating the activities of the IBRD (The World Bank) and IDA, IEG–IFC focuses on assessment of IFC's work toward private sector development, and IEG–MIGA evaluates the contributions of MIGA guarantee projects and services. IEG reports directly to the Bank's Board of Directors through the Director-General, Evaluation.

The goals of evaluation are to learn from experience, to provide an objective basis for assessing the results of the Bank Group's work, and to provide accountability in the achievement of its objectives. It also improves Bank Group work by identifying and disseminating the lessons learned from experience and by framing recommendations drawn from evaluation findings.

# OECD/DAC Network on Development Evaluation

The Development Assistance Committee (DAC) Network on Development Evaluation is an international forum where bilateral and multilateral development evaluation experts meet regularly to improve evaluation practice and share experience. Its purpose is to increase the effectiveness of international development programs by supporting robust, informed, and independent evaluation.

The Evaluation Network is a subsidiary body of the DAC and presently consists of 30 representatives from OECD member countries and multilateral development agencies (Australia, Austria, Belgium, Canada, Denmark, European Commission, Finland, France, Germany, Greece, Ireland, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States, World Bank, Asian Development Bank, African Development Bank, Inter-American Development Bank, European Bank for Reconstruction and Development, UN Development Programme, International Monetary Fund).

Further information may be obtained from OECD, Development Co-operation Directorate, 2 rue André-Pascal, 75775 Paris Cedex 16, France, or dacevaluation.contact@oecd.org, web address: www.oecd.org/dac/evaluation.

# Sourcebook for Evaluating Global and Regional Partnership Programs

## Indicative Principles and Standards

Printed on Recycled Paper

This Sourcebook has been prepared by the Independent Evaluation Group of the World Bank under the auspices of the OECD/DAC Network on Development Evaluation.

# Contents

# Boxes

# Tables

# ACKNOWLEDGMENTS

Hurlbut provided editorial guidance. Hans Lundgren and Sebastian Ling, the coordinator and administrator of the DAC Evaluation Network, helped organize and run the stakeholder workshop. Rose Gachina, Thelma Alfred, Nathalie Bienvenu, Denise Moran, Petra Touam, and Yumiko Takahashi provided administrative and logistical support to the workshop. Caroline McEuen and Heather Dittbrenner edited the final draft.

The Sourcebook builds on work that has already been done by the DAC Evaluation Network, the United Nations Evaluation Group, the Evaluation Cooperation Group of the Multilateral Development Banks, evaluation associations, and others to develop principles, norms, and standards for evaluating development assistance programs, projects, and activities. It also draws on IEG's experience in reviewing GRPPs over the last few years, under the leadership of Uma Lele, as well as the feedback that was received at the stakeholder workshop.

This Sourcebook is one of a number of initiatives that have sought to understand and respond to the recent growth of GRPPs. Other contributions to the growing body of knowledge concerning global and regional partnership arrangements include (a) the work of Inge Kaul and her colleagues in the United Nations Development Programme, Office of Development Studies, who have produced three research publications on global public goods (1999, 2003, and 2006); (b) the UN Vision Project on Global Public Policy Networks (2000) led by Wolfgang Reinicke and Francis Deng; and (c) the International Task Force on Global Public Goods (2006) led by Tidjane Thiam, Ernesto Zedillo, and Sven Sandström.

At the present stage, the indicative principles and standards contained in this Sourcebook do not necessarily represent the official views of the World Bank, IEG, the DAC Evaluation Network, or DAC members.

# ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| AfDB | African Development Bank |
| AIDS | Acquired immunodeficiency syndrome |
| AsDB | Asian Development Bank |
| CGIAR | Consultative Group on International Agricultural Research |
| DAC | Development Assistance Committee (OECD) |
| DGF | Development Grant Facility (World Bank) |
| EU | European Union |
| EBRD | European Bank for Reconstruction and Development |
| ECG | Evaluation Cooperation Group (MDBs) |
| GEF | Global Environment Facility |
| GPR | Global program review (IEG) |
| GRPP | Global and regional partnership program |
| HIV | Human immunodeficiency virus |
| ICR | Implementation Completion Report (World Bank) |
| IDB | Inter-American Development Bank |
| IEG | Independent Evaluation Group, formerly OED (World Bank) |
| IFC | International Finance Corporation |
| IMF | International Monetary Fund |
| Logframe | Logical framework |
| M&E | Monitoring and evaluation |
| MDB | Multilateral Development Bank |
| NGO | Nongovernmental organization |
| OECD | Organisation for Economic Co-operation and Development |
| OED | Operations Evaluation Department, now IEG (World Bank) |
| SMART | Specific, measurable, attainable, relevant and time-bound |
| TDR | Special Programme on Research and Training in Tropical Diseases |
| TOR | Terms of reference |
| UNDP | United Nations Development Program |
| UNEG | United Nations Evaluation Group |

# OVERVIEW

## Origin of this Sourcebook

1.        At the March 30–31, 2006, meeting of the OECD/DAC Network on Development Evaluation, representatives of the World Bank's Independent Evaluation Group (IEG) presented their observations on the growing need to develop consensus principles and standards for the evaluation of Global and Regional Partnership Programs (GRPPs), based on their recent reviews of a sample of such programs and their evaluations. The meeting was attended not only by members from the evaluation units of 23 bilateral agencies and development cooperation ministries, but also by representatives of the African Development Bank (AfDB), the Asian Development Bank (AsDB), the European Bank for Reconstruction and Development (EBRD), the Inter-American Development Bank (IDB), the International Finance Corporation (IFC), the International Monetary Fund (IMF), and the United Nations Development Program (UNDP). Participants at the meeting expressed broad support for the development of such principles and standards, and requested that IEG play a leading role in developing them. The present Sourcebook of indicative principles and standards for evaluating GRPPs is the result of IEG's response to this request.

2.        An earlier draft of this Sourcebook was reviewed at a stakeholder consultative workshop held for this purpose in Paris on September 28–29, 2006. The workshop validated the approach of producing a free-standing and comprehensive document that presents, synthesizes, applies, and elaborates on existing evaluation principles and standards for the particular benefit of the governing bodies and management units of GRPPs. Workshop participants also provided comments that have substantially improved the operational relevance of the Sourcebook and called for the additional preparation of a companion document of guidance notes and good-practice examples for the particular benefit of evaluators of GRPPs.

## Purpose

3.        The purpose of the indicative principles and standards contained in this Sourcebook is to improve the independence and quality of program-level evaluations of GRPPs in order to enhance the relevance and effectiveness of the programs. The principal audiences are the governing bodies and management units of GRPPs, as well as professional evaluators involved in the evaluation of these programs. It is also hoped that these principles and standards will heighten

awareness and help advocate for improved evaluation of GRPPs among higher-level policy makers in both aid agencies and developing countries.

4.        Improving the results-based monitoring and evaluation (M&E) of GRPPs will also require much collaboration and consultation within the international development community. Both IEG and the DAC Evaluation Network hope that this Sourcebook will assist with that effort, and encourage (a) disseminating these indicative principles and standards widely to enhance the credibility and quality of GRRP evaluations; (b) monitoring their application and use by GRPP governing bodies, managers, and evaluators; and (c) continuing to share experience and fostering discussion among both commissioners and providers of evaluations, as well as other experts, on good practice in evaluation of GRPPs.

## Features of GRPPs and Implications for Evaluation

5.        GRPPs are an increasingly important modality for channeling and delivering development assistance to address pressing global/regional issues and concerns. For the purpose of this Sourcebook, GRPPs are programmatic partnerships in which:

- The partners contribute and pool resources (financial, technical, staff, and reputational) toward achieving agreed-upon objectives over time.
- The activities of the program are global, regional, or multi-country (not single-country) in scope.
- The partners establish a new organization with a governance structure and management unit to deliver these activities.

6.        Most GRPPs are specific to a certain sector or theme, such as agriculture, environment, health, finance, or international trade. Almost all advocate greater attention to specific issues or approaches to development in their sector, but on different scales:

- Some, generally small, programs are primarily policy or knowledge networks that facilitate communication, advocate policy change, and generate and disseminate knowledge and good practices in their sector.
- Other, somewhat larger, programs also provide country or local-level technical assistance to support national policy and institutional reforms and capacity strengthening, and to catalyze public or private investment in the sector.
- The largest programs also provide investment resources to support the provision of global, regional, or national public goods.

7.      Notwithstanding this diversity, GRPPs have many shared features that distinguish them from other common subjects of evaluation — projects, country-specific programs, and policies — and thus require special treatment in evaluation. These features and their implications for evaluation are summarized in Table 1.

8.      Among programmatic partnerships that meet the above definition, this Sourcebook is focused primarily on GRPPs that are engaged in international development and that provide public goods, whether through aid or trade mechanisms. The founding partners of these GRPPs have typically been international organizations (such as United Nations specialized agencies and the World Bank), bilateral aid agencies, and non-profit foundations engaged in development. Their objectives have been to promote a public interest in a particular area of development, even in the case of those programs with private sector partners (both commercial and non-commercial). Nonetheless, other types of GRPPs whose partners and objectives are more private in nature may also benefit from the principles and standards laid out in the Sourcebook.

9.      The term *donor* is used in this Sourcebook in the generic sense as referring to any organization or entity that makes a financial or in-kind contribution to a program that is reflected in the audited financial statements of the program. Thus, the term includes not only "official donors" but also developing countries that contribute annual membership dues, seconded staff, or office space, provided that these are formally recognized, as they should be, in the financial statements of the program. Donors can also be beneficiaries, especially in the case of programs that provide global public goods of direct or tangential benefit to both developed and developing countries. But in this Sourcebook, the term *donor* does not extend to beneficiary countries or groups that are providing counterpart contributions that are not formally recognized in the financial statements of the program.

10.     The term *stakeholders* refers to the parties who are interested in or affected, either positively or negatively, by the program. The term *partners* refers to stakeholders who are involved in the governance or financing of the program (including the members of the governing, executive, and advisory bodies), while the term *participant* refers to those involved in the implementation of the program (including the final beneficiaries). Both partners and participants are subsets of stakeholders. Stakeholders are often referred to as "principal" and "other," or "direct" and "indirect." While other or indirect stakeholders — such as taxpayers in both donor and beneficiary countries, visitors to a beneficiary country, and other indirect beneficiaries — may have interests as well, these are not ordinarily considered in evaluations unless a principal stakeholder acts as their proxy.

## Table 1. Indicative Features of GRPPs and Their Implications for Evaluation

| GRPP Feature | Implications for Evaluation |
|---|---|
| **GRPPs are programmatic partnerships** with multiple donors, partners, and other stakeholders, whose interests do not always coincide. There is joint decision making and accountability at the governance level. | • Identifying the various categories of stakeholders early in the planning for a GRPP evaluation, and taking account of their diverse interests, is very important in order to determine the appropriate degree of participation and consultation during the evaluation process.<br><br>• Assessing the continued relevance to principal stakeholders on both the supply and demand sides of the program is necessary, including confirmation that the program's objectives remain consistent with its authorizing environment and any applicable international conventions.<br><br>• Assessing the legitimacy and effectiveness of the governance and management arrangements is essential. Communications with and the flow of information to the various stakeholders are important determinants of legitimacy and effectiveness.<br><br>• The fiduciary standards of different donors and trustees need to be taken into account in drawing up an evaluation terms of reference and in assessing governance and management. The assessment of management should include some assessment of financial management, reporting, and compliance with donor requirements, since this can have a significant effect on mobilizing resources. |
| **GRPPs are global or regional in scope,** work in differing socio-political contexts, and operate at multiple levels — global, regional, national, and local. | • Soon after the launch of the program, management needs to establish a results-based M&E framework. Designing and implementing a multi-level M&E framework for a range of activities operating in diverse contexts is complex.<br><br>• In GRPP evaluations, decisions on evaluation scope, coverage, and sampling need to ensure adequate representativeness for validity of the findings.<br><br>• GRPP evaluations require a longer time frame and larger budget to achieve a sufficient level of data collection and stakeholder participation and consultation, because of the program's wide geographic scope, large number of beneficiaries, and multiple operational levels. |
| **Most GRPPs are housed in international organizations.** While most of these have separate governing bodies, their managers are employees of the host organization. | • GRPPs should have an evaluation policy that is approved by the governing body, in which the principles of independence and impartiality are agreed upon.<br><br>• Evaluations should be commissioned by and report to a governing body (not management) to ensure independence and impartiality and to guard against institutional bias. Governing bodies should ensure competitive and transparent bidding and selection. |
| **The results of GRPPs are the joint product of global / regional and country-level activities** and of parallel activities financed by other development agents. | • Assessing the effectiveness of a GRPP requires consideration of the program's inputs, outputs, outcomes, and impacts at all levels — global, regional, and national — ideally based on measurable indicators and a representative sample of activities at all levels.<br><br>• Attribution is often particularly difficult to discern in the case of a GRPP. |

| GRPP Feature | Implications for Evaluation |
|---|---|
| **The program usually evolves over time,** based on the availability of financing, and does not usually have a fixed end-point. | • The purpose, objectives, scope, and design of an evaluation need to take into account the maturity of the program. Evaluations are generally mid-term rather than ex-post, and are often planned to build on each other sequentially.<br><br>• GRPP evaluations should include an assessment of sources and uses of funds and the resource mobilization strategy.<br><br>• In a mature program, it may be necessary to assess strategies for devolution, exit, or alternative organizational and financing arrangements that are under consideration or under way. |
| **Governance and management are multi-layered and decision making is complex.** Continuity may be uncertain because the members of the governing body may rotate or otherwise change due to political circumstances. | • Assessment of the legitimacy and effectiveness of governance and management should analyze the respective roles of the governing body and management in various decision-making processes.<br><br>• Evaluators need to ascertain any changes in the membership criteria for the governing body or changes in actual representation.<br><br>• Feedback processes and dissemination plans for evaluation products need to be defined before the evaluation to include all relevant stakeholders. |
| **The decisions on activities to support are made through a programmatic process,** rather than fixed in advance as in a discrete project. | • The criteria and processes for allocating resources and choosing activities to support are important ingredients of both relevance and effectiveness, and need to be assessed. |
| **GRPPs are typically externally financed** with little capacity to earn income from their own resources. Total financing depends on individual donors' funding decisions. | • Assessing the sources and uses of funds and the relationship of the resource mobilization strategy to scale, effectiveness, and efficiency of the program is important in a GRPP evaluation.<br><br>• Causality may flow in both directions: the resource mobilization strategy has implications for effectiveness and efficiency, and the achievement of results and reports on them may influence the success of the resource mobilization strategy.<br><br>• In a mature program, it may be necessary to assess alternative financing arrangements (such as cost sharing), if these are under consideration or under way. |
| **GRPPs take several years to set up,** due to the need to reach consensus and establish the legal framework and governance arrangements. Sunk costs are relatively high at initial stages. | • Analysis of costs and benefits in an evaluation should ideally factor in start-up costs that were incurred prior to the formal legal establishment, and should include the costs incurred by the convening partners.<br><br>• At a minimum, GRPP evaluations should assess administrative costs relative to activity costs, and note any actual or expected economies of scale. For mature programs, it may be possible to compare the costs of individual activities to sectoral benchmarks, generic cost indicators, or the costs of other GRPPs in delivering similar activities. |
| **GRPPs are diverse** in size, age, sectoral focus and objectives, and in the types of activities supported (knowledge, technical assistance, investments). | • While some variation in evaluation approach and design is to be expected, some standards for evaluation of GRPPs are necessary to ensure credibility, and a minimum frequency is necessary to meet accountability objectives.<br><br>• The evaluation design, scope, coverage, and methodology may also differ according to the governing body's purpose in conducting an evaluation at a particular point in time, the maturity of the program, the portfolio size, and the type of activities supported. |

## Additional Background

11.    As a partner in most of these programs, the World Bank has joined the other members of GRPP governing bodies in commissioning regular evaluations of the relevance and contributions of GRPPs to development. As a donor to many of these programs through its Development Grant Facility (DGF), the World Bank has also introduced M&E requirements to ensure accountability and to promote continuous improvements in performance. In 2001, IEG initiated a comprehensive review of the World Bank's involvement in global programs, which included a review of the internal support and oversight functions and multiple case studies of individual programs in a variety of sectors. These case studies drew on external evaluations of the programs, supplemented by interviews with management and partners, and updated investigations of results.

12.    The result was two volumes that present IEG's findings.[1] The Phase 1 Report on global programs largely addressed the Bank's internal support and oversight processes for managing its global program portfolio. The Phase 2 Report presented additional findings from 26 case studies and made two recommendations for IEG that are relevant to the present initiative:

- IEG should review selected program-level evaluations conducted by Bank-supported global programs, like IEG reviews other evaluations of Bank support at the project and country levels. The findings of these Global Program Reviews (GPRs) would be reported to the World Bank's Executive Board and, after a pilot phase, disclosed to the public.

- IEG should work with the Bank's global partners to develop consensus standards for the evaluation of global programs.

13.    IEG has moved forward on both recommendations. First, it has developed a set of guidelines for its own GPRs in consultation with the Bank's units involved with GRPPs, operations policy, and trust fund management. These guidelines have built on the evaluation framework in IEG's Phase 2 Report and have incorporated lessons derived from the experience with three pilot GPRs completed by IEG in fiscal year 2006. IEG has also completed a review of Bank support for regional programs and initiated seven more GPRs to be completed in

---

1. Operations Evaluation Department (OED) of the World Bank, *The World Bank's Approach to Global Programs*, Phase 1 Report, 2002, and *Addressing the Challenges of Globalization: An Independent Evaluation of the World Bank's Approach to Global Programs*, Phase 2 Report, 2004. OED formally changed its name to the Independent Evaluation Group (IEG) of the World Bank in December 2005.

fiscal year 2007. Second, IEG discussed the need for consensus standards for the evaluation of GRPPs with the DAC Evaluation Network in March 2006, which resulted in the request for IEG to produce the present Sourcebook.

## Sources

14.     The principles and standards in this Sourcebook draw on three main sources: (a) existing evaluation principles, norms, standards, and guidelines that have been developed by international agencies and the evaluation networks to which they belong, other partners in GRPPs, and professional evaluators; (b) the guidelines that IEG uses for its own global program reviews; and (c) a forthcoming review of World Bank support for regional programs.[2]

15.     The aim has been to develop a set of principles and standards that are applicable to both *global* and *regional partnership* programs. This does not include regional (multi-country) investment projects supported by the World Bank and other donors, which have a substantially different character from partnership programs, and have proven more straightforward for the Bank and other donors to evaluate through their regular M&E processes. However, regional partnerships have some distinguishing features from global partnerships, and there has so far been less experience to draw upon in evaluating regional compared with global partnerships. Throughout the Sourcebook, the acronym "GRPP" is used when the principle or standard is intended to apply to both global and regional partnerships, and the adjectives "global" and "regional" are used when it is intended to apply to only one. It may be necessary at a later stage to make a more extensive effort to incorporate the distinguishing features of regional partnerships into this Sourcebook, as the experience in evaluating regional partnerships grows.

16.     Most of the indicative principles and standards in the Sourcebook are based on, draw on, or elaborate on principles and standards from sources that specifically cover evaluation of development assistance (Table 2). Many needed to be adapted or expanded to accommodate the special features of GRPPs highlighted in Table 1. The principles and standards of professional associations generally go into more detail on ethical and professional conduct of evaluations. This Sourcebook assumes that appropriate methods and criteria for selection of the evaluation team will result in choosing evaluators who will abide by these ethical and professional standards, and therefore it was not necessary to cite them all. Nonetheless, these were reviewed

---

2. *Regional Development Programs: An Independent Evaluation of World Bank Support*. This will be published in early 2007.

### Table 2. Principal Sources

OECD/DAC Principles for Evaluation of Development Assistance (1991)
OECD/DAC Evaluation Quality Standards (2006)
OECD/DAC Glossary of Key Terms in Evaluation and Results Based Management (2002)
OECD/DAC Guidance on Joint Evaluations (2006)
OECD Principles of Corporate Governance (1999, revised 2004)

UNEG Norms for Evaluation in the UN System (April 2005)
UNEG Standards for Evaluation in the UN System (April 2005)

ECG Good Practice Standards
ECG Template for Assessing the Independence of Evaluation Organizations

DGF Technical Note on Independent Evaluation: Principles, Guidelines and Good Practice (November 2003)
Global Environment Facility Monitoring and Evaluation Policy (February 2006)

U.S. Joint Committee on Standards for Educational Evaluation: Program Evaluation Standards (1994)
African Evaluation Association: African Evaluation Guidelines (2002)

American Evaluation Association Guiding Principles for Evaluators (Revised July 2004)
Canadian Evaluation Society Guidelines for Ethical Conduct
Council on Foundations, Evaluation Approaches and Methods (2003)

IEG, Addressing the Challenges of Globalization, An Independent Evaluation of the Bank's Approach to Global Programs, Phase 2 Report (2004)
IEG, Guidelines for Global Program Reviews (2006)

Note: Links to all these documents are available on the IEG Web site, http://www.worldbank.org/ieg/grpp. OECD/DAC, Organisation for Economic Co-operation and Development/Development Assistance Committee; UNEG, United Nations Evaluation Group; ECG, Evaluation Cooperation Group of the Multilateral Development Banks.

for completeness and are sometimes the source of amplifying footnotes.

17.     The various professional association guidelines have much in common. However, sometimes cultural differences influence the standards adopted. Particularly pertinent in this regard — and relevant for the evaluation of GRPPs — is the exercise that the African Evaluation Association undertook, beginning in 1998, to review the U.S. Program Evaluation Standards and to consider where modifications were desirable for evaluations of development programs in Africa. The resulting modified guidelines were published in 2002. Examples of cases where the Association felt the need to modify the U.S. Program Evaluation Standards are presented in Table 3. Those commissioning evaluations and selecting teams may wish to familiarize themselves with some of the cultural factors that emerged as important in considerations of methodology, evaluation practice, transparency, and participation and the need for members of the evaluation team to be sensitive to cultural norms.[3]

---

3. The African Evaluation Association is currently reviewing and revising its

Table 3. Comparison Between U.S. and African Evaluation Standards

| Standard | U.S. Program Evaluation Standards (1994) | African Evaluation Guidelines (2002) |
|---|---|---|
| Stakeholder identification | Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed. | Persons *and organizations* involved in or affected by the evaluation *(with special attention to beneficiaries at the community level)* should be identified *and included in the evaluation process*, so that their needs can be addressed *and so that the evaluation findings are utilizable and owned by stakeholders, to the extent that this is useful, feasible, and allowed.* |
| Values identification | The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear. | The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear. *The possibility of allowing multiple interpretations of findings should be transparently preserved, provided that these interpretations respond to stakeholders' concerns and needs for utilization purposes.* |
| Disclosure of findings | The formal parties to an evaluation should ensure that the full set of evaluation findings, along with pertinent limitations, are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results. | The formal parties to an evaluation should ensure that the full set of evaluation findings, along with pertinent limitations, are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results *as far as possible. The evaluation team and the evaluating institution will determine what is deemed possible, to ensure that the needs for confidentiality of national or governmental entities and of the contracting agents are respected, and that the evaluators are not exposed to potential harm.* |

## Use of Evaluation Terms

18.     This Sourcebook focuses on *results-based monitoring and evaluation.* It views the establishment of a results-based monitoring and evaluation system as a key enabling condition for effective evaluation of GRPPs. Throughout this Sourcebook, therefore, "monitoring and evaluation" is understood as "results-based monitoring and evaluation."[4] This is, however, distinct from "results-based management," which is defined by the DAC as "a management strategy focusing on performance and achievement of outputs, outcomes and impacts" (Table 4). Specifically, a results-based *measurement* system,

---

evaluation guidelines further, among other reasons, to reflect cultural aspects of evaluation.

4. While most modern monitoring and evaluation systems are results-based — meaning that these emphasize measuring outcomes and impacts and not just inputs and outputs — there may still be some monitoring and evaluation systems that are more narrowly focused on the monitoring of activities and inputs.

Table 4. Key Terms in Results-Based Management, Monitoring and Evaluation

| Term | Definition |
|------|------------|
| Results-based management | A management strategy focusing on performance and achievement of outputs, outcomes, and impacts. |
| Results chain | The causal sequence for a development intervention that stipulates the necessary sequence to achieve desired objectives — beginning with inputs, moving through activities and outputs, and culminating in outcomes, impacts, and feedback. In some agencies, reach is a part of the results chain between outputs and outcomes. |
| Inputs | The financial, human, and material resources used for a development intervention. |
| Results | The outputs, outcomes, or impacts (intended or unintended, positive or negative) of a development intervention. |
| Outputs | The products, capital goods and services that result from a development intervention. This may also include changes resulting from the intervention that are relevant to the achievement of outcomes. |
| Outcomes | The achieved or likely short-term and medium-term effects of the outputs of a development intervention. |
| Impacts | Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended. |
| Indicator | A quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor. |
| Performance monitoring | A continuous process of collecting and analyzing data to compare how well a policy, program, or project is being implemented against expected results. |

Source: OECD/DAC, *Glossary of Key Terms in Evaluation and Results Based Management*, 2002.

and related processes of monitoring and evaluation is a necessary, but not sufficient, condition for successful results-based *management*. Other aspects of management strategy — such as strategic planning and human resource performance management — are not dealt with in this Sourcebook.

19.     For the purpose of this Sourcebook, the terms "*principles*" and "*norms*" are used interchangeably. These are used in the same sense that they are used in the DAC *Principles for Evaluation of Development Assistance* (1991) and the United Nations Evaluation Group (UNEG) *Norms for Evaluation in the UN System* (2005). These are presented as guidance for reference, and not as binding on any parties. They are characteristic of *principles* in that they are either a presentation of facts or logical relationships observed over time that guide action; or they are prescriptions that are widely agreed to further the goals of professionalism, credibility, usefulness, and collaboration in evaluation. They are characteristic of *norms* in that they are based on shared purpose and frequent use. Principles and norms are expected to stay relatively stable over time and to be widely applicable.

20.    In contrast, a *policy statement* lays out actions and behaviors expected by those in authority of all members of a group or organization, and is therefore tailored to serve the specific objectives of a particular organization. The Sourcebook has drawn on evaluation policy statements of some GRPPs where these elucidate the application of widely accepted norms and principles to GRPPs. However, in these cases the extract from the policy statement has been relabeled a principle or a norm, since it would not be binding on other GRPPs.

21.    Both the DAC Evaluation Network and UNEG have also produced "*standards*" for evaluation in addition to their principles and norms (Table 2). In general, the standards are related to, and derived from, the principles and norms. Also, as noted earlier, some professional associations (of foundations, grant-makers, and evaluators) have adopted standards. Others have labeled similar prescriptions "*guidelines*." In this Sourcebook, the terms "standards" and "guidelines" are used interchangeably and considered to be at a lower level of generalizability than principles and norms. Examples of good practice are also presented under the same heading in some cases.

22.    In most cases, the DAC or UNEG designation of principle, norm, or standard has been maintained when these are cited and when they are the main source of the content of the proposed GRPP principle, norm, or standard. Any exceptions have been noted in a footnote. For easy reference, the sources that the principles and standards for GRPPs are based upon, draw on, or elaborate on are indicated in the margins.[5] For simplicity, where a principle, norm, or standard draws on more than one source, the sources are uniformly listed in the same order as in Table 2.

23.    For the benefit of users, this Sourcebook is a free-standing document, which builds on and applies existing evaluation principles, norms, and standards to GRPPs. Therefore, many of the 17 chapters contain principles and standards upon which the international devel-

---

5. Where the proposed principle, norm, or standard quotes directly from an existing principle, norm, or standard, quotation marks are used in the text. Where much of the text is the same, but paraphrased, the term "based on" is used in the margin note. In cases where the text relies substantially on a particular source, but also includes some original material supplemental to the source, the term "draws on" is used in the margin note. In cases where the text relies primarily on a source for one or more ideas, but also provides some additional rationale or description, the term "elaborates on" is used. In cases where the text makes adjustments to accommodate the special features of GRPPs, the phrase "applies to GRPPs" is included. While the authors have aimed to cite the main sources, it is sometimes possible that another source containing the same idea is not cited, due to lack of awareness or space constraints.

opment evaluation community has already reached consensus and which are included in the Sourcebook primarily for reasons of completeness, so that the users of the document do not have to continually refer to other principles and standards documents.

24.     The largest number of existing principles and standards that have been adapted and applied to the evaluation of GRPPs are contained in Chapters 2 through 7 on evaluation governance and process issues. The largest segment of material that is based on IEG's experience with reviewing GRPPs is contained in Chapters 9 through 15 on evaluation content and criteria issues. The remaining chapters — 8, 16, and 17 — are primarily checklists that have been derived from existing principles and standards documents. All chapters have benefited from the positive and constructive discussions at the stakeholder consultative workshop held in Paris on September 28–29, 2006.

# GLOSSARY

The following definitions reflect the use of these terms in the context of evaluating global and regional partnership programs (GRPPs). Therefore, these definitions do not necessarily reflect the use of these same terms in other contexts. Many of these definitions are based on the OECD/DAC *Glossary of Key Terms in Evaluation and Results Based Management*, 2002. The page or footnote references at the end of the various definitions indicate where in the Sourcebook the definition of the term is expanded or the use of the term is placed in a particular context, if applicable.

*Accountability:* As a criterion for assessing governance and management, the extent to which accountability is defined, accepted, and exercised along the chain of command and control within a program, starting with the annual general meeting of the members or parties at the top and going down to the executive board, the chief executive officer, task team leaders, implementers, and in some cases, to the beneficiaries of the program. [page 77]

*Baseline:* An analytical description of the situation prior to a development intervention, against which progress can be assessed or comparisons made.

*Cluster evaluation:* The simultaneous evaluation of more than one GRPP operating in the same sector, or operating collaboratively. [page 2, footnote 9]

*Cost-effectiveness:* The extent to which the program has achieved or is expected to achieve its results at a lower cost compared with alternatives. [page 65]

*Counterfactual:* The situation or condition that hypothetically would have prevailed if there had been no development intervention. [page 32, footnote 39]

*Devolution or exit strategy:* A proactive strategy to change the design of a program, to devolve some of its implementation responsibilities, to reduce dependency on external funding, or to phase out the program on the grounds that it has achieved its objectives or that its current design is no longer the best way to sustain the results which the program has achieved. [pages 87–88]

*Donor:* Any organization or entity that makes a financial or in-kind contribution to a program that is reflected in the audited financial statements of the program. Therefore, this includes not only "official donors" but also developing countries that contribute annual membership dues, seconded staff, or office space, provided that these are formally recognized in the financial statements of the program. [pages xvii and 76, footnote 71]

*Effectiveness (or efficacy):* The extent to which the program has achieved, or is expected to achieve, its objectives, taking into account their relative

importance. The term is also used as a broader, aggregate measure — encompassing relevance and efficiency as well — of the overall outcome of a development intervention such as a GRPP. [page 57 and footnote 55]

*Efficiency*: The extent to which the program has converted or is expected to convert its resources/inputs (such as funds, expertise, time, etc.) economically into results in order to achieve the maximum possible outputs, outcomes, and impacts with the minimum possible inputs. [page 65]

*Evaluability*: The extent to which an activity or program can be evaluated in a reliable and credible fashion. An evaluability assessment is a review of a given program, at the early stages of or preceding an evaluation, to determine, among other things, whether the program's objectives are adequately defined and its results verifiable. [pages 6–7]

*Evaluation:* The systematic and objective assessment of an ongoing or completed policy, program, or project, its design, implementation, and results. The aim is to determine the relevance and achievement of its objectives, and its developmental effectiveness, efficiency, impact, and sustainability. [page 1]

*Fairness:* As a criterion for assessing governance and management, the extent to which partners and participants, similarly situated, have equal opportunity to influence the program and to receive benefits from the program. [page 77]

*Financial management:* The processes that govern the recording and use of funds, including allocation processes, crediting and debiting of accounts, controls that restrict use, accounting, and periodic financial reporting systems. It also includes the processes which ensure that funds are used for the purposes intended — a fiduciary standard that is expected by the vast majority of donors. In cases where funds received accumulate over time, it would also include the management of the cash and investment portfolio. [pages 83–84]

*Formative evaluation:* An evaluation that is intended to improve performance, which is most often conducted during the implementation phase of programs or projects. [page 1, footnote 8]

*Governance:* The structures, functions, processes, and organizational traditions that have been put in place within the context of a program's authorizing environment to ensure that the program is run in such a way that it achieves its objectives in an effective and transparent manner. It is the framework of accountability and responsibility to users, stakeholders and the wider community, within which organizations take decisions, and lead and control their functions, to achieve their objectives. [page 71]

*Impact evaluation:* A systematic assessment of the effects — positive or negative, intended or unintended — of one or more development interventions on the final welfare outcomes of the affected individuals, households, and communities, and the extent to which these outcomes can be attributed to the development intervention(s). In its most rigorous form, an impact evaluation compares the welfare outcomes of the intervention(s) with an explicit counterfactual of what the outcomes would have been in the absence of the intervention(s). [page 95]

*Impacts:* Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.

*Impartiality:* In conducting an evaluation, the absence of bias in due process, in the scope and methodology, and in considering and presenting achievements and challenges. The principle applies to all members of the governing body, other donors and partners, management, beneficiaries, and the evaluation team. [page 15]

*Independent evaluation:* An evaluation that is carried out by entities and persons free from the control of those involved in policy making, management, or implementation of program activities. This entails organizational and behavioral independence, protection from interference, and avoidance of conflicts of interest. [pages 15–18]

*Indicator:* A quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor.

*Inputs:* The financial, human, and material resources used for a development intervention.

*Joint evaluation:* An evaluation that is conducted collaboratively by more than one partner. [page 2, footnote 9]

*Legitimacy:* As a criterion for assessing governance and management, the way in which governmental and managerial authority is exercised in relation to those with a legitimate interest in the program — including shareholders, other stakeholders, implementers, beneficiaries, and the community at large. [page 76]

*Logical framework or logframe:* A management technique that is used to develop the overall design of a program or project, to improve implementation monitoring, and to strengthen evaluation, by presenting the essential elements of the program or project clearly and succinctly throughout its cycle. It is a "cause and effect" model which aims to establish clear objectives and strategies based on a results chain, to build commitment and ownership among the stakeholders during the preparation of the program or project, and to relate the program's or project's interventions to their intended outcomes and impacts for beneficiaries. [page 13]

*Management:* The day-to-day operation of the program within the context of the strategies, policies, processes, and procedures that have been established by the governing body. [page 71]

*Monitoring:* The continuous assessment of progress achieved during program implementation in order to track compliance with a plan, to identify reasons for noncompliance, and to take necessary actions to improve performance. Monitoring is usually the responsibility of program management and operational staff. [page 1]

*Organizational capture:* A situation in which the host organization for a GRPP takes over and runs the program as if it were one of its own. [page 81, footnote 76]

*Outcomes:* The achieved or likely short-term and medium-term effects of the outputs of a development intervention.

*Outputs:* The products, capital goods and services that result from a development intervention. This may also include changes resulting from the intervention that are relevant to the achievement of outcomes.

*Oversight:* One of the core functions of the governing body of a program: Monitoring the performance of the program management unit, appointing key personnel, approving annual budgets and business plans, and overseeing major capital expenditures. [page 72]

*Participants:* Stakeholders who are involved in the implementation of the program (including the final beneficiaries), but not in the governance of the program. [pages xvii and 22]

*Partners:* Stakeholders who are involved in the governance or financing of the program (including the members of the governing, executive, and advisory bodies). [pages xvii and 22]

*Path-dependence:* The dependence of institutional choices and economic outcomes on the path of previous choices and outcomes, rather than simply on current conditions. In path-dependent processes, institutions are self-reinforcing, history has an enduring influence, and choices are made on the basis of transitory conditions that persist long after these conditions change. [page 79, footnote 73]

*Performance monitoring:* A continuous process of collecting and analyzing data to compare how well a policy, program, or project is being implemented against expected results.

*Probity:* As a criterion for assessing governance and management, the adherence by all persons in leadership positions to high standards of ethics and professional conduct over and above compliance with the rules and regulations governing the operation of the program. [page 78]

*Public goods:* Goods which produce benefits that are non-rival (many people can consume, use, or enjoy the good at the same time) and non-excludable (it is difficult to prevent people who do not pay for the good from consuming it). If the benefits of a particular public good accrue across all or many countries, then the good is deemed a global or international public good. [page 52]

*Reach:* The beneficiaries and other stakeholders of a development intervention, or the degree to which the outputs of the intervention are extended to a broad range of beneficiaries in order to achieve more extensive results.

*Readiness assessment:* In relation to the establishment of a monitoring and evaluation system, a diagnostic tool for assessing the organizational capacity of a program and the political willingness of its governing body to monitor and evaluate the achievement of the program's goals and to develop a performance-based framework. [page 11]

*Relevance*: The extent to which the objectives and design of the program are consistent with (a) current global/regional challenges and concerns in a

particular development sector and (b) the needs and priorities of beneficiary countries and groups. [page 49]

*Resource mobilization:* The process by which resources are solicited by a program and provided by donors and partners. [page 83]

*Resources:* The inputs that are used in the activities of a program. Broadly speaking, the term encompasses natural, physical, financial, human, and social resources. [page 83]

*Responsibility*: As a criterion for assessing governance and management, the extent to which the program accepts and exercises responsibility to stakeholders who are not directly involved in the governance of the program and who are not part of the direct chain of accountability in the implementation of the program. [page 77]

*Results:* The outputs, outcomes, or impacts (intended or unintended, positive or negative) of a development intervention.

*Results-based management:* A management strategy focusing on performance and achievement of outputs, outcomes, and impacts.

*Results chain:* The causal sequence for a development intervention that stipulates the necessary sequence to achieve desired objectives — beginning with inputs, moving through activities and outputs, and culminating in outcomes, impacts, and feedback. In some agencies, reach is a part of the results chain between outputs and outcomes.

*Selection bias:* The distortion that arises in a statistical analysis due to the methodology that was used to collect the samples. For instance, the beneficiaries of a certain intervention may be selected (or self-selected) on the basis of certain characteristics. If these characteristics are unobserved, then only a randomized approach can in principle eliminate such selection bias. [page 97, footnote 89]

*Shareholders:* The subset of donors that are involved in the governance of the program. Therefore, this does not include individual (particularly anonymous) donors who choose not to be so involved, or who are not entitled to be involved if their contribution does not meet the minimum requirement, say, for membership on the governing body. [page 76, footnote 71]

*Stakeholder map:* A comprehensive list of the principal or direct stakeholders of a particular program, which also includes information on their perceived roles and responsibilities in relation to the program. [pages 9 and 22]

*Stakeholders:* The parties who are interested in or affected, either positively or negatively, by the program. Stakeholders are often referred to as "principal" and "other," or "direct" and "indirect." While other or indirect stakeholders — such as taxpayers in both donor and beneficiary countries, visitors to a beneficiary country, and other indirect beneficiaries — may have interests as well, these are not ordinarily considered in evaluations unless a principal stakeholder acts as their proxy. [pages xvii and 22]

*Subsidiarity:* As a criterion for assessing the relevance of a program, whether the activities of the program are being carried out at the most appropriate level — global, regional, national, or local — in terms of efficiency and responsiveness to the needs of beneficiaries. [page 52]

*Summative evaluation:* An evaluation study that is conducted at the end of an intervention (or a phase of that intervention) to determine the extent to which anticipated outcomes were produced during the period being evaluated. [page 1, footnote 8]

*Supervision:* One of the functions of program staff (or in some cases, contractors): Administering and monitoring the implementation of individual program activities. This includes contracting with implementing or executing agencies to implement individual activities and ensuring that they are reporting their progress in a timely way. [page 73]

*Sustainability:* When the term is applied to the **activities of a program**, the extent to which the benefits arising from these activities are likely to continue after the activities have been completed. When the term is applied to **organizations or programs** themselves, the extent to which the organization or program is likely to continue its operational activities over time. [page 87]

*Transparency:* As a criterion for assessing governance and management, the extent to which a program's decision-making, reporting, and evaluation processes are open and freely available to the general public. This is a metaphorical extension of the meaning used in the physical sciences — a "transparent" object being one that can be seen through. [pages 77–78]

*Triangulation:* The use of three or more theories, sources, or types of information, or types of analysis, to verify and substantiate an assessment. By combining multiple data sources, methods, analyses, or theories, evaluators seek to overcome the bias that comes from single informants, single methods, single observers, or single theory studies. [page 46, footnote 49]

*Value-for-money:* The extent to which a program has obtained the maximum benefit from the outputs and outcomes it has produced with the resources available to it. [page 65, footnote 59]

# INTRODUCTION

## 1. Definitions, Purposes, and Minimum Expectations for Credibility and Usefulness

### Principles and Norms

STAKEHOLDERS' INTEREST IN MONITORING AND EVALUATION

1.1 All principal stakeholders — partners, donors, management, employees, and direct beneficiaries — have an interest in M&E, both for accountability to political authorities and the general public and for learning from experience in order to improve the use of development resources.[6]

Based on DAC Principle I, para. 2

DEFINITIONS

1.2 Monitoring, the responsibility of the management and operational staff, is the continuous assessment of progress achieved during program implementation in order to track compliance with the plan, to identify reasons for noncompliance, and to take necessary actions to improve performance.

Based on DAC principle V, UNEG Norm 1, and GEF Policy, section 1.3

1.3 An evaluation is a systematic and objective assessment of an ongoing or completed policy, program, or project, its design, implementation, and results. The aim is to determine the relevance and achievement of its objectives, and its developmental effectiveness, efficiency, impact, and sustainability.[7]

Based on DAC Principle I, para. 5, and UNEG Norm 1

TYPES OF EVALUATION

1.4 Evaluations may be internally or externally led, and may adopt a formative or summative approach.[8] They may be aimed at a

Based on UNEG Norm 1, para. 1.6

---

6. Stakeholders are often referred to as "principal" and "other," or "direct" and "indirect." While other or indirect stakeholders — such as taxpayers in both donor and beneficiary countries, visitors to a beneficiary country, and other indirect beneficiaries — may have interests as well, these are not ordinarily considered in evaluations unless a principal stakeholder acts as their proxy.

7. Some evaluations also assess value-for-money, target group satisfaction, and additionality or value added. Descriptors that are sometimes used to distinguish evaluations from other types of reviews and assessments include: objective, credible, reliable, and drawing on evidence-based information.

8. A *formative evaluation* is "intended to improve performance [and is] most often conducted during the implementation phase of projects or programs." A *summative evaluation* is a "study conducted at the end of an in-

single program, to determine its contribution to one or more development objectives, or they may be *cluster evaluations* [9] to assess several programs operating in the same sector or country, or to evaluate collaborative efforts. An evaluation can be conducted at any time during the life of a program, at mid-point, at end-phase, or at end-point. If desirable for accountability or for learning lessons applicable to other development efforts, an impact evaluation of selected program activities may be conducted either during or after the closing of a program. (See also Chapter 15, Impact Evaluation.)

GENERAL PURPOSES OF MONITORING AND EVALUATION

Based on GEF Policy, section 1.3, para. 18

1.5    *Monitoring* provides initial information on progress toward achieving intended objectives, outcomes, and impacts — including productivity and other efficiency targets — and gives signals and information for proactive and reactive decision making by management. A good monitoring system for a GRPP combines information at all levels — the program, portfolio, and activity levels — to provide a comprehensive picture of performance to management and to facilitate decision making and learning.

Based on DAC Principle II, paras. 6 and 10, and UNEG Norm 13, para.13.1

1.6    The general purposes of *evaluation* of GRPPs are to improve the performance of the program in meeting its objectives and to provide a basis for accountability to donors, stakeholders, and the general public. Specifically, evaluation aims to improve the relevance of the program, to enhance achievement of results, to optimize resource use, and to address issues of target group satisfaction. With appropriate stakeholder participation, an evaluation can promote dialogue and improve cooperation between partners and participants, with the side-benefits of increasing beneficiary ownership of policy reforms or new types of interventions. With appropriate dissemination it can also contribute to organizational learning and knowledge building that may also benefit other programs and development efforts.

---

tervention (or a phase of that intervention) to determine the extent to which anticipated outcomes were produced. A summative evaluation is intended to provide information about the worth of the program." See OECD/DAC, *Glossary of Key Terms in Evaluation and Results Based management*, 2002.

9. In this Sourcebook the term *cluster evaluation* refers to the simultaneous evaluation of more than one GRPP operating in the same sector, or operating collaboratively. The term refers to the multiple subjects of the evaluation, rather than the collaboration of the evaluators. The term *joint evaluation* refers to evaluations that are conducted collaboratively by more than one partner — the same way in which the OECD/DAC uses the term. There is potential overlap between the two concepts: more than one agency may collaborate in evaluating either a single GRPP or a cluster of several GRPPs.

1.7    While evaluation contributes to ensuring accountability, its focus on relevance, results, and efficiency distinguishes it from ensuring accountability for the use of public funds in the accounting and legal senses, which generally requires in-depth examination by audit agencies. Optimum levels of oversight to assure accountability require both periodic evaluation and audits.

Based on DAC Principle II, para. 8, and UNEG Norm 1, para. 1.4f

INTENTIONALITY TO USE RESULTS OF EVALUATION

1.8    To achieve their purposes, evaluations must be used.[10] They should be timely, and accepted as relevant and useful for decision making on important matters. Feedback and dissemination to management, partners, and operational staff are essential in order to facilitate decision making and learning. Dissemination to other stakeholders in a clear and concise form is also desirable for transparency. Evaluation always requires an explicit response by the commissioners of the evaluation and the management of the program. After each evaluation is completed and for the benefit of future evaluations, the commissioners of the evaluation may also wish to review the results of the evaluation process and consider, among other things, if more funding or a different focus might have enhanced its usefulness.

Based on DAC Principles I, V, and X, UNEG Norm 12, and GEF Policy, section 3.3

## Standards and Guidelines

### MINIMUM CONDITIONS FOR CREDIBLE AND QUALITY MONITORING AND EVALUATION

1.9    Good quality monitoring systems use SMART indicators — "specific, measurable, attainable, relevant, and time-bound" — to track the use of inputs, the progress of activities, the outputs associated with key activities, and outcomes. For GRPPs, some indicators will be defined at the program level, some at the portfolio level (such as aggregate summary statistics), and some at the activity level. While objective data on inputs and results are always preferable, some data may also reflect subjective or summary assessments. Data collection is timely, of adequate periodicity to facilitate problem solving and support decision making, and is controlled by a quality-assurance system. Accountability for data collection and quality assurance is clear, and incentives are appropriate to ensure an acceptable level of quality. Monitoring reports to management and governing bodies are clear, accessible, and easy to understand, and include definitions and parameters.

Elaborates on and applies to GRPPs, DAC Standard 4.2, and GEF Policy, section 3.2, para. 56, and section 1.3, para. 17

---

10. The U.S. Program Evaluation Standards for professional evaluators include a set of seven "utility standards" to help ensure that evaluation will serve the information needs of intended users. The African Evaluation Association has adapted these standards to the African context.

Elaborates on
and applies DAC
Principle IV,
para. 18, and
UNEG Norm 8,
to GRPPs

1.10    The credibility and quality of *evaluation* of GRPPs depends on (a) the degree of independence of the evaluation process; (b) the degree of transparency of the evaluation process; (c) appropriate participation and consultation with relevant stakeholders; (d) the expertise and experience of the evaluators; (e) appropriately defined scope and methodology; and (f) clearly defined criteria for assessment. In addition, the budget must be sufficient for the chosen scope and methodology in order to avoid compromising the credibility or quality of the evaluation.

# EVALUATION GOVERNANCE ISSUES

## 2. Prerequisites and Enabling Conditions for Effective Evaluations

### Principles and Norms

INSTITUTIONAL RESPONSIBILITY FOR MONITORING AND EVALUATION

2.1     Appropriate institutional arrangements for managing M&E are a prerequisite for ensuring effective processes and for making full use of the information generated by M&E systems. Plans for proper monitoring and periodic evaluation should be built into the design of the program at inception. Institutional arrangements need to meet the requirements for (a) a policy and set of guidelines for M&E, including a disclosure policy; (b) impartiality and independence of evaluations; and (c) using M&E findings to improve future decision making and activities.

<div style="float:right">Based on DAC Principle III, para. 14, and UNEG Norm 2</div>

2.2     Monitoring is always the responsibility of the management and operational staff, and evaluation is the responsibility of the governing body or other unit separate from management. In most GRPPs, evaluations are commissioned by part-time governing bodies and conducted by independent teams of consultants or independent experts. In larger GRPPs, there may be a mandate and sufficient resources for a separate internal evaluation unit.[11] In either case, the body commissioning the evaluation takes responsibility for the quality of the final report and for disseminating the findings and recommendations, in different formats for different audiences, as appropriate.[12]

<div style="float:right">Applies DAC Principle V, para. 22, and GEF Policy, section 5.1, para. 72, to GRPPs</div>

MONITORING AND EVALUATION POLICY

2.3     Existing principles and norms issued by the DAC and the UNEG stress the need for a policy on evaluation approved by the

<div style="float:right">Elaborates on and applies DAC Principle III and UNEG Norm 3 to GRPPs</div>

---

11. Where there is a separate evaluation unit, the principle on competency should also apply: "Aid agencies/ [programs] need a critical mass of professional evaluation staff in order to have sufficient expertise in their various fields of activity and to ensure credibility of the process." (DAC Principle IV, para. 19.)

12. Having a governing body commission or manage an evaluation that includes an assessment of governance (that is, the performance of the governing body itself) poses a potential conflict of interest. In some cases, this may be resolved by designating a "higher body" or "external group" to which the evaluation team would report directly. See also paragraphs 3.5 and 3.6 on organizational independence and Chapter 12, Governance and Management.

governing body.[13] At a minimum, the governing body should make an early commitment in the policy to the following:

- To have the GRPP evaluated periodically and to provide adequate funding for conducting evaluations
- To agree on the purpose of regular evaluations
- To agree on how evaluation results will be disseminated and used to improve accountability, learning, decision making, and broader knowledge sharing (including among evaluators).

2.4    The policy should also address such issues as independence and impartiality of evaluation, desired stakeholder participation and consultation, and the openness of the evaluation process (including disclosure). The policy should define the respective roles of management and governing bodies in M&E, and explain how evaluations are expected to be planned, managed, budgeted, and reviewed. The policy should also include guidance on mandatory criteria and on processes for selecting external evaluation teams, if applicable. (See the standards section below for further guidance.)

2.5    The monitoring and evaluation policy should also refer to the need for an effective monitoring system, both to provide the information required for scheduled reporting to the governing body on the use of resources, the progress of activities, outputs, and outcomes, and to provide the information necessary for future evaluations.

### MONITORING AND EVALUATION FRAMEWORK

Elaborates on DAC Principle I, para. 4, and UNEG Norm 7, para. 7.2

2.6    A key enabling condition for effective evaluation is the early establishment of an M&E framework. Early after its launch, each GRPP should put in place an M&E framework, at least at the program level, which includes (a) clear and coherent objectives and strategies, (b) an expected results chain, (c) measurable indicators that meet the monitoring and reporting needs of the governing body and management, and (d) systematic and regular processes for collecting and managing data, including baseline data. (See standards and guidelines below for establishing an effective M&E framework.) An approximate date for the first independent evaluation should be set (generally 2–3 years after the launch of the program), and the evaluation budgeted.

2.7    An evaluation may have to include — or be preceded by — an assessment of the adequacy of the M&E framework and system, if there are doubts about their adequacy. Such an evaluability assess-

---

13. UNEG Norm 7 suggests that evaluation policy be built into early plans at inception of the program. This Sourcebook extends this to include monitoring as well as evaluation.

ment would determine whether the objectives of the GRPP have been clearly defined; whether *SMART* indicators have facilitated the collection of timely, relevant, and accurate data; whether information sources are accessible and reliable; and whether any serious constraints prevent an impartial evaluation process. The commissioners of the evaluation should decide whether the evaluability assessment should precede or be part of the evaluation itself.

## MONITORING AND EVALUATION PLANNING AND PROGRAMMING

2.8     M&E requirements for reporting, accountability, and learning need to be built into the regular planning processes in GRPPs from the start. (See standards and guidelines below for more information on establishing monitoring systems.)

Based on DAC Principle I, para. 4, and GEF Policy, section 3.1, para. 54

2.9     Planning and programming of evaluations must take into account the needs of the governing bodies and program managers, as well as those of other potential users of evaluation products (such as policy makers and activity implementers, whether public or private). There may also be external requirements for timing of evaluations.[14] Otherwise, evaluations should be timed to provide effective input into key decisions, such as approving a new phase, expansion, funding replenishments, restructuring governance, reaching out to new donors, and the like. Other important factors that influence timing include whether the program has reached a certain degree of maturity or completed a critical mass of supported activities. (See also paragraph 6.7.) Cluster evaluations that include other programs with similar objectives, or those operating in the same sector, may be explored. The evaluation plan should identify specific possibilities for interaction with stakeholders and for the participation of various groups of stakeholders. It should also include a dissemination plan.

Elaborates on DAC Principle VIII, para. 30, and GEF Policy, sections 2.8 and 3.3

2.10     Ideally, an evaluation plan calling for broad stakeholder participation would meet all the accountability needs of donors and other partners and would obviate the need for individual stakeholders to undertake separate evaluations. In practice, however, some individual donors may need to conduct separate or joint evaluations with other donors in order to meet their reporting obligations to their authorities, thereby giving rise to multiple evaluations. In such cases, donors and other partners should, at a minimum, share their plans for evaluations and schedule them to facilitate potential complementarity and appropriate programming, since joint, or at least coordinated evaluations

Elaborates on and applies to GRPPs, DAC Principle VII and DAC *Guidance for Managing Joint Evaluations*

---

14. For example, the World Bank's Development Grant Facility (DGF) requires that an independent evaluation be conducted every three-to-five years for the programs that it supports.

tend to achieve greater benefits than separate evaluations in terms of efficiency, consensus-building, and joint learning.[15]

| | |
|---|---|
| Elaborates on DAC Principle VIII, para. 30 | 2.11　　The M&E plan, including provisions for any single-donor or joint evaluations, requires the support and endorsement of the governing body. Before any evaluation commences, the full governing body should approve the management and reporting arrangements and the terms of reference (TOR).[16] |

### RESOURCES AND BUDGETING

| | |
|---|---|
| Based on DAC Principle X, para. 42, and GEF Policy, section 3.1, para. 54 | 2.12　　The governing body must allocate adequate resources for the effective implementation and operation of the M&E plan. Planning for monitoring and evaluation should be an explicit part of planning and budgeting for the program as a whole. This includes allocating staff and budget resources to establish feedback mechanisms in order to ensure that the results of evaluations are utilized in future policy and program development. |

### QUALITY CONTROL

| | |
|---|---|
| Based on DAC Standard 8.2 | 2.13　　Quality control must be exercised throughout the evaluation process. Depending on the scope and complexity of the evaluation, quality control is carried out either internally or through an external body, peer review, or reference group. (See also paragraph 2.30.) |

## Standards and Guidelines

### ADDITIONAL FUNCTIONS OF THE INSTITUTIONAL ARRANGEMENTS FOR MONITORING AND EVALUATION

| | |
|---|---|
| Draws on UNEG Standard 1.1 | 2.14　　Ideally, the institutional arrangements for monitoring and evaluation should also (a) promote a culture that values M&E as a basis for learning and improving the effectiveness of the program; (b) provide adequate financial resources for M&E; (c) ensure that accountability for M&E and its follow-up is clear; and (d) enable capac- |

---

15. OECD/DAC recently issued guidelines for joint evaluations, which are available on their Web site: *Guidance for Managing Joint Evaluations*, 2006, and *Joint Evaluations: Recent Experiences, Lessons Learned and Options for the Future*, 2005.

16. The U.S. Program Evaluation Standards and the African Evaluation Guidelines cite among "feasibility standards" the importance of assessing the "political viability" of the evaluation up front: "The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted."

ity strengthening in M&E, cooperation, and shared learning with other organizations or programs.

## SPECIAL ARRANGEMENTS FOR DRAWING UP THE M&E POLICY AND EVALUATION PLAN

2.15    Given the large size of some governing bodies, and the possibility that members may not have evaluation experience or expertise, it may be necessary to set up a subcommittee to draw up the evaluation policy and M&E plan. This subcommittee may in turn wish to seek help from an external peer group of representatives of various stakeholders or from expert consultants. The full governing body should approve the final M&E policy and evaluation plan. The use of an external peer group may also be appropriate for overseeing and managing especially complex or difficult evaluations.

## PARTICIPATORY M&E AND EARLY IDENTIFICATION OF STAKEHOLDERS

2.16    The M&E policy and evaluation plan should give serious consideration to participatory methodologies. The very process of developing a participatory monitoring system tends to enable and inform the planning of participatory evaluations. It also contributes to capacity strengthening and can lay the foundation for sharing experience across activities supported by the program. This, in turn, will enhance program improvement and subsequent evaluation. A high degree of stakeholder participation in prior evaluations is also likely to enhance the quality of subsequent evaluations.

2.17    A comprehensive list of stakeholders, or "stakeholder map," which also includes information on their perceived roles and responsibilities, is an indispensable prerequisite for determining the participation and consultation process to be followed in conducting evaluations. This map should be updated periodically, and included in the TOR for each evaluation, along with the desired participation and consultation process, to guide the evaluation team. (See also paragraphs 4.7–4.9, and 4.17–4.20.)

## MONITORING AND EVALUATION POLICY

2.18    The approved GRPP evaluation policy should include:

Based on UNEG Standard 1.22

- A clear explanation of the concept and role of evaluation within the organization
- A clear definition of the roles and responsibilities of the governing body, senior management, task team leaders, and evaluation professionals, if applicable
- Clear guidelines on the process of planning for evaluations
- Clear guidelines on how evaluations are organized, managed, and budgeted

- Clear guidelines on the follow-up of evaluations
- A clear statement on and guidelines for disclosure and dissemination.

2.19    The evaluation policy, and related policies on risk management and audit plans, should also be agreed with the host organization in which the program is located, if applicable.

EVALUATION PLANNING AND PROGRAMMING [17]

2.20    Evaluation planning and programming needs to take into account the maturity of the program, since its maturity will likely affect both the purpose of the evaluation and its scope and methodology. The following may serve as a general guideline:

- **Program in Early Stages (first 2–3 years):** Important purposes of an early — usually the initial — evaluation would be to assess the appropriateness of the program design and to review the governance and management arrangements. The evaluation should also review the relevance and clarity of the objectives, identify constraints that make achievement of specific objectives difficult or impossible, and recommend adjustments if necessary.

- **Established Program (over 5 years old):** The evaluation should address inputs, the progress of activities, and outputs. The recommendations will typically focus on ways to increase the effectiveness and efficiency of the program.

- **Mature Program:** At this stage, the program should be operating smoothly and meeting the expectations expressed at its initiation. The evaluation will typically pay particular attention not only to outputs, but also to outcomes, as well as to sustainability and other strategic issues such as growth, devolution, or exit.

2.21    It may be advantageous to undertake evaluations for established and mature programs in the form of "cluster evaluations" done jointly with other programs operating in the same sector. A simultaneous assessment of their objectives, strategies, and activities may identify a potential for achieving joint results more effectively. Complementary or mutually reinforcing inputs, activities, and outputs may also be identified, yielding recommendations for improving collaboration and reducing duplication.

---

17. See also Chapter 6, Planning for Scope and Methodology.

DESIGN OF THE M&E FRAMEWORK

2.22    The design of an M&E framework at the activity level will vary by program, in terms of both detail and breadth of applicability. Ratings of progress and performance may or may not be used. For most GRPPs, particularly those that do not involve investments, it may not be cost-effective to require an M&E plan or independent evaluation for each separate activity.[18] Rather, for most GRPPs, broad indicators will typically be drawn up for monitoring all similar activities. A subset of activities, defined by size or type, may require periodic progress reports and a completion report at the end. For most programs these will be self-evaluations. These reports will help provide essential information not only for monitoring, but also for evaluations as well.

2.23    If the governing body and/or management decide that an impact evaluation (of longer-term results) will be required in the future, it is important to plan for this from the outset, so that appropriate indicators (possibly for control groups as well as beneficiary groups), data-gathering, quality control mechanisms, and recording of baseline data can be funded and implemented. It may be advisable to contract with an expert evaluation team for the planning of the future impact evaluation, including the collection of baseline data. (See Chapter 15, Impact Evaluation.)

INITIAL STEPS IN ESTABLISHING AN M&E FRAMEWORK [19]

2.24    At the inception of the program, when preparing the budget for the early years, the development of an M&E framework should be adequately funded and provision made for staffing or engaging consultants for its planning and implementation.

2.25    It may be advisable to carry out a "readiness assessment" before the establishment of the system. This is a diagnostic tool for assessing the organizational capacity of the program and the political willingness of its governing body to monitor and evaluate the achievement of the program's goals and to develop a performance-based framework. The three main objectives of a readiness assessment are (a) to determine whether incentives are appropriate for the suc-

Draws on the book, *Ten Steps to a Results-Based Monitoring and Evaluation System*, as applicable to GRPPs

---

18. The Global Environment Facility has such requirements for its investment grants, since it views the benefits as exceeding costs, and has a critical mass of staff available to carry out the requirements.

19. This section draws on the book by Jody Zall Kusek and Ray C. Rist, 2004, *Ten Steps to a Results-Based Monitoring and Evaluation System*. However, this section does not cover all ten steps, since the conduct of evaluation, the use of findings, and sustaining the system lie beyond the initial steps and for the most part is covered in other parts of this Sourcebook.

cessful launch of an M&E system; (b) to analyze roles and responsibilities in order to be able to assign accountability for the functioning of the M&E system; and (c) to assess existing capacity.

2.26    At a minimum, and to provide an adequate foundation for accountability, the monitoring of inputs (such as expenditures and staff time) should begin immediately after the launch of the program. A classification scheme for expenditures needs to be drawn up after careful discussion. It should differentiate between (a) administrative and activity expenditures; (b) expenditures at the country, regional, and global levels; and (c) other relevant distinctions (such as categories of expenditure, or expenditures by different agencies, and the like) A decision needs to be made whether to record and/or track non-financial contributions/expenditures.

2.27    The next step toward developing an M&E framework is agreement on the objectives of the program. For many programs, consensus objectives will have been defined at meetings preceding the funding of the program and incorporated in the program charter. For other programs, the funding follows agreement on the need to address a particular issue, but the proper response to it, and thus the objectives of the program, are not agreed until later. It may be necessary not only to list objectives, but also to establish an "objectives hierarchy," should trade-offs emerge later in implementation.

2.28    Once objectives are agreed, a strategy and a set of interventions or activities are agreed upon, generally with the expectation that these will generate specified outputs and outcomes. The early establishment of such an expected results chain is an important step in building an M&E framework. Also important is the early identification of stakeholders and categories of beneficiaries to enable monitoring of welfare outcomes.

2.29    Next, additional indicators that measure the progress of activities, outputs, outcomes, and impacts, and any relevant external factors that affect results must be defined and provision made for their measurement. Other indicators that might be required to assess the achievement of the objectives — such as the success of participatory measures, or the degree in which the program is successful in responding to opportunities and in learning from experience — also need to be defined. Indicators of participation should cover not just participation rates (inputs) but also the expected outcomes such as learning, awareness, behavioral change, and so on. The measurement of the agreed indicators could be the responsibility of activity-level staff in the course of implementation, it could rely on government statistics, or it could be done periodically by specialized program staff (for example, by means of surveys of target group satisfaction).

2.30     Accountability for data collection, monitoring, and reporting needs to be clear, and a quality-assurance system put in place. Once this is done, plans should be made to collect baseline data and ensure that these are stored in an accessible place.

2.31     As the program evolves, strategic plans prepared by the program management and approved by the governing body may define specific expected outcomes and implementation targets. Management should also agree with the governing body on a desirable periodicity of reporting, which allows regular tracking of progress against targets and expected outcomes.

2.32     Finally, feedback mechanisms are needed to ensure that the comments of management, the governing body, implementing staff, and data-gatherers themselves are taken into account to help to continuously improve the M&E system. The system may also need to be adapted periodically in response to changing objectives and strategies, or changes in the external environment.

## USE OF A LOGICAL FRAMEWORK

2.33     A logical framework, or logframe, is a management technique that is used to develop the overall design of a development project, to improve implementation monitoring, and to strengthen evaluation, by presenting the essential elements of the project clearly and succinctly throughout its cycle. It is a "cause and effect" model that has been widely used by the bilateral and multilateral donor community since the 1970s in order to establish clear objectives and strategies based on a results chain, to build commitment and ownership among the stakeholders during the preparation of a project, and to relate the project's interventions to their intended outcomes and impacts for beneficiaries.[20]

Draws on the World Bank's *LogFrame Handbook*

2.34     Developing a logframe is as appropriate for GRPPs that have no fixed end-point as it is for projects with an expiration date, as long as the logframe is kept updated and relevant, and incorporates input from new stakeholders as they become involved in the program. But developing a logframe for a GRPP may be more complex because of the larger number of stakeholders and the different levels of program components and activities. Ideally, a single consensus logframe would be developed for the overall program, which would form the basis for the implementation and monitoring of the program and subsequent program-level evaluations. It may also be desirable to supplement this with more narrowly focused logframes for different stakeholder groups, who may have different objectives and who may

---

20. For more information, see World Bank, 2004, *The LogFrame Handbook: A Logical Framework Approach to Project Cycle Management*.

wish to add additional indicators that reflect their particular interests in the program.

### MONITORING SYSTEMS AND INDICATORS

2.35    In planning and adopting monitoring systems, GRPPs should use indicators that are SMART — specific, measurable, attainable, relevant, and time-bound.[21]

### PEER REVIEW OR REFERENCE GROUP

2.36    It may be useful to establish a peer review or reference group composed of technical experts in the sector concerned and/or M&E experts. This group would provide substantive guidance to the monitoring and/or evaluation process (such as providing inputs on the TOR and providing quality control of draft reports).

### STAKEHOLDER STEERING OR LEARNING GROUP

2.37    "When feasible, a core learning group or steering group composed of representatives of the various stakeholders in the evaluation may be created. This group's role would be to act as a sounding board, and to facilitate and review the work of the evaluation. In addition, this group may be tasked with facilitating the dissemination and application of the results and other follow-up action," particularly with their own constituents.

### KNOWLEDGE MANAGEMENT AND DISSEMINATION

2.38    GRPPs should support knowledge sharing by ensuring the highest standards in accessibility and presentation of M&E products, by using a range of channels to reach target audiences, by participating in knowledge management, and by sharing activities with other relevant organizations.

---

21. In the case of indicators, the GEF Policy also adds "realistic" for R, and "timely, traceable, and targeted" for T. The book, *Ten Steps to a Results-Based Monitoring and Evaluation System* also cites useful criteria under the acronym "CREAM": Indicators should be **Clear** (precise and unambiguous); **Relevant** (appropriate to the subject at hand); **Economic** (available at a reasonable cost); **Adequate** (providing a sufficient basis to assess performance); and **Monitorable** (amenable to independent validation). See also Salvatore Schiavo-Campo, 1999, "'Performance' in the Public Sector," p. 85.

# 3.  Independence and Impartiality in Conducting Evaluations

## Principles and Norms

### INDEPENDENCE AND IMPARTIALITY AS A PREREQUISITE FOR CREDIBILITY OF EVALUATION

3.1     To ensure its credibility, the evaluation process should be independent from any process involving program policy making, management, or activity implementation, as well as impartial. Impartiality is the absence of bias in due process, in the scope and methodology, and in considering and presenting achievements and challenges. The principle of impartiality applies to all members of the governing body, other donors and partners, management, beneficiaries, and the evaluation team. And the requirements for independence and impartiality are present at all stages of the evaluation process, including planning, budgeting and financing, formulation of mandate and scope, drafting of TOR, selection and approval of evaluation teams, conduct of the evaluation, formulation of findings and recommendations, and review and finalization of the report (and other products arising from the evaluation).

*Elaborates on and applies DAC Principles III and VI and UNEG Norm 5 to GRPPs*

3.2     A well-defined policy on monitoring and evaluation should be established during the setting up of the program to systematize the evaluation function and to ensure that these requirements are met. The policy should also provide for adequate budgets and funding for evaluations which are separate from regular program management funds. (See also paragraphs 2.3–2.5.) The requirements for independence and impartiality are particularly important for GRPPs, since the majority of programs are housed in (or hosted by) one of the partner organizations, and the program staff may be formally employed by that organization. Independence and impartiality are thus required to guard against bias and ensure that the views of all stakeholders are taken into account. While independence is essential for credibility, it is not a guarantee of a quality evaluation product.

### ORGANIZATIONAL INDEPENDENCE

3.3     "The evaluation function has to be located independently from the other management functions so that it is free from undue influence and so that unbiased and transparent reporting is assured." Accordingly, the members of an evaluation unit or team should not have been directly responsible for setting the policy, design, or overall management of the program, nor expect to be in the near future. Members of an evaluation unit or team evaluating a GRPP should report to a unit separate from program management. This would nor-

*Elaborates on and applies to GRPPs, UNEG Norm 6, paras. 6.1 and 6.3, and ECG Template for Assessing the Independence of Evaluation Organizations*

mally be the commissioner of the evaluation, usually the governing body.[22] Members of the unit or team should be insulated from political pressures from either donors or beneficiary groups and should not participate in political activities that could affect independence.

3.4    The larger GRPPs may set up and finance separate internal evaluation units.[23] To preserve the independence of these units, they should report directly to the governing body, not line management. To give credence to the evaluation function, the head of the unit should be sufficiently high in rank.

3.5    The majority of GRPPs rely on teams of external consultants for periodic evaluation work. Ideally, the governing body, which is separate from program management, should commission the evaluation, approve the TOR, select the team, and ultimately approve the evaluation report in order to ensure ownership of the findings and follow-up. However, it may not be feasible for the governing body to actively *manage* the evaluation process, or for the entire governing body to review the evaluation in detail, since the governing body may have limited time and evaluation expertise. In these cases, the governing body may entrust these functions to a subcommittee on oversight and evaluation in order to preserve the principle of independence. The governing body should ratify the composition of such a subcommittee, which would ideally have representation from each of the different categories of stakeholders on the governing body. It might also include external members with evaluation expertise — from outside both the program and the governing body.[24]

---

22. In some cases, the evaluation team has reported to host organizations. This is a second-best solution, since the host organization is only one of the partners on the governing body to which the program is accountable. When the host organization bears too much responsibility for the evaluation, this may reduce the incentive for other partners to participate fully and effectively, or the ability of the host organization to look at the weaknesses of the program objectively. (See also paragraph 12.27.)

23. This is the case with the Global Environment Facility. Where there is a separate evaluation unit, an additional requirement for ensuring independence is that unit staff are protected by a personnel system in which compensation, training, tenure, and advancement are based on merit, and where budgetary resources are determined in accordance with a clear policy parameter. (See the ECG Template for Assessing the Independence of Evaluation Organizations.)

24. For example, expertise could be drawn from the evaluation units of one or more of the partner organizations, as long as that unit is independent of their line management, and as long as staff who participate in the evaluation of a GRPP do not subsequently participate in reviews or meta-evaluations of this particular evaluation.

3.6    For small GRPPs that do not have the resources to establish a formal oversight subcommittee, a less structured peer review or advisory panel may be a lower-cost alternative. At a minimum, such an external panel should have at least one member with adequate stature and evaluation expertise to ensure impartiality. Panel membership could be voluntary, with members drawn from the academic and research communities.

## BEHAVIORAL INDEPENDENCE AND PROTECTION FROM INTERFERENCE

3.7    In addition to organizational independence, behavioral independence must be assured. For large GRPPs with internal evaluation units, whether or not they report to line management, it is advisable to have an external peer review process. This could involve an evaluator from a peer organization who would be able to provide impartial comments and judgments with respect to the process and the evaluation findings.

*Based on UNEG Norm 6, paras. 6.4 and 6.5, ECG Template for Assessing the Independence of Evaluation Organizations, and GEF Policy, section 3.3*

3.8    The evaluation team, whether internal or external, should be able to work freely and without interference. It should be assured of cooperation and access to all relevant information. Team members should be able to express their views in a free manner. Vested interests on the part of either the program management and commissioners of evaluations or the evaluation team should not be allowed to interfere with the conditions for an impartial and independent evaluation. Provisions for phased payments for external consultants need to be accompanied by assurances that review of interim products for payment are based on an objective confirmation of delivery of expected products, rather than findings.

## AVOIDANCE OF CONFLICTS OF INTEREST [25]

3.9    Any conflict of interest should be addressed openly and honestly at any stage of the evaluation process at which it arises, so that it does not undermine the evaluation outcome. For a large GRPP with an internal evaluation unit, where there is a "revolving door" practice within the organization (that is, evaluation staff have the opportunity to move into positions within program operations, and vice versa), steps should be taken to minimize potential conflicts of interest.[26]

*Elaborates on DAC Standard 6.1, UNEG Standards 2.1 and 3.15, paras. 5 and 34, and ECG Template for Assessing the Independence of Evaluation Organizations*

---

25. This section has been placed under the heading of *principles and norms* even though it draws primarily on existing standards, since avoiding conflicts of interest is an important factor in determining the degree of independence.

26. For instance, incoming staff (to the evaluation unit) should declare potential conflicts of interest if they are assigned to an activity in which they had prior involvement. Outgoing staff should not be transferred — for a minimum period of, say, five years — to activities they have previously evaluated in order to reduce the likelihood for partiality when an activity being evaluated presents opportunities for future job placements/advancements.

3.10    Evaluators, both internal and external, should declare any conflict of interest to the commissioners before embarking on an evaluation project, and at any point where such a conflict occurs. Evaluators should also report — to those who commissioned or are managing the evaluation — any conflict of interest that they discover on the part of other participants in the evaluation, such as stakeholders consulted. If a potential conflict of interest arises, and if the managers of the evaluation identify and/or accept special means to diminish its implications for independence and impartiality, both the initial conflict and the actions taken should be disclosed to the governing body and the program management. As a general rule, conflicts of interests and how they were dealt with should be disclosed in the final report.[27]

### THE NEED FOR BALANCE

3.11    The need for impartiality and for the absence of bias requires that evaluations give a comprehensive and balanced presentation of the strengths and weaknesses of the program being evaluated. To the extent possible, the evaluation should reflect the views of all partners and participants — including donors, implementers, and beneficiaries — regarding the relevance and effectiveness of the activities being evaluated. When interested parties have different views, these should be reflected in the evaluation analysis and reporting.

## Standards and Guidelines

### SPECIAL CONSIDERATIONS IN ENSURING INDEPENDENCE AND IMPARTIALITY

3.12    For large GRPPs with internal evaluation units, it has been argued that certain ways of organizing the evaluation function might strengthen independence and impartiality, but weaken the potential linkage between evaluation findings and follow-up decisions. If some evaluation functions must be attached to line management, the staff exercising such functions should report to a sufficiently high level of the management structure, or to a management committee, to help avoid compromising the independence of the evaluation process and its results.

3.13    In GRPPs where the provision for financing of evaluations has not yet been systematized, one donor partner or group of donor partners

---

27. Members of GRPP evaluation teams should not be currently employed by any of the governing partners, except by one of their evaluation units if this unit is independent of their line management. If an evaluation team, after being selected, recruits a team member who is an employee or consultant of one of the governing partners, the potential for conflict of interest should be carefully considered. One result might be that the individual serves as a resource person, as opposed to a fully independent member of the core evaluation team.

has often paid for the evaluation directly. In these cases, in order to have a balanced and unbiased evaluation product that will have ownership by the governing body and broader stakeholders, care should be taken to ensure that the financiers do not have undue influence over the evaluation process (including the drafting of the TOR and the selection of consultants). Regardless of the funding source, the procedures described in paragraphs 3.5 and 3.6 of using an oversight committee or an external panel endorsed by the governing body should be followed.

3.14    Given that GRPPs are a fairly new but growing phenomenon, the pool of evaluation candidates with the experience and technical knowledge required to evaluate the program may be small, and the only candidates with the necessary skills may have had prior involvement with the program in question. But hiring such candidates may pose a conflict of interest and compromise the independence of the evaluation. (See paragraphs 7.15 and 7.16 on measures to prevent or mitigate such a situation.)

REVIEW OF DRAFT EVALUATION REPORTS

3.15    To improve the probability of behavioral independence and protection from interference, the governing body and the program management should agree early in the program on the procedures for reviewing the draft evaluation report. It is highly recommended that these procedures be uniform for each evaluation and laid out in advance in an evaluation policy. (See paragraphs 2.4 and 3.1.) Or they could be allowed to evolve, for instance, as the governing body gains experience working with the management team. In either case, the agreed-upon procedures should be stated in the evaluation TOR. (See also paragraph 16.4.)

3.16    To ensure organizational and behavioral independence, the evaluation team should report to the governing body (or to an oversight committee or external panel, as discussed in paragraphs 3.5 and 3.6). The management of the program should also be given the opportunity to review the draft evaluation report in order to correct any factual errors and to comment on the findings and recommendations. But this should be done in such a way that maintains the behavioral independence of the evaluation team and provides for transparency (to the governing body) regarding any changes that management proposes (Box 1). The evaluation team must have the ability to express its findings without undue interference, while providing for quality assurance and promoting efficient, open discussion. In all cases, the evaluation team must retain the discretion to accept or reject any of the changes that management proposes. Under no circumstances, should management be perceived as or be allowed to "clear" the evaluation report, or impose any amendments on it.[28]

---

28. This having been said, the evaluation team has the strong incentive, for its own credibility, to correct all errors of fact or interpretation in the report.

> **Box 1. Possible Alternatives for Reviewing the Draft Evaluation Report**
>
> - Provide the draft first to the commissioner of the evaluation for comment. This is usually the governing body, or a subcommittee thereof. The report may also be provided to any technical advisory committee at the same time, or shortly thereafter. Under this alternative, management would only receive the draft report after the commissioner of the evaluation has had a chance to comment.
>
> - Provide the draft to the governing body and management at the same time. Then the governing body can choose to read the first draft at this stage or wait until management has reviewed it and provided comments and/or corrections. But this procedure may stretch the capacity of the governing body, whose members may feel that they are getting more information than they need. And the evaluation team may find it confusing to receive comments (possibly conflicting) from both the governing body and management at the same time.
>
> - Provide the draft to management first, and have management copy their comments to the governing body. After reading management's comments, the governing body may request a copy of the first draft of the evaluation if they so desire, and they are free to comment from that point on. In this case, the team can manage comments in sequence.
>
> - Provide the draft to management first and let the governing body know that management has provided comments to the evaluation team. Also let the governing body know that both the first draft and management comments are available on request. In this case, transparency is on demand.

### DESCRIPTION OF DEGREE OF INDEPENDENCE IN EVALUATION REPORTS

Based on DAC
Standards 6.1
and 6.2

3.17    The evaluation report should indicate the degree of the independence of the evaluators from the policy, operations, and management functions of the commissioners, implementers, and beneficiary groups. It should also indicate the level of transparency and impartiality observed in the commissioning, contracting, definition of scope of work, and selection of evaluators. Conflicts of interest and the ways in which they were dealt with should be addressed openly and honestly. It would also be good practice for the evaluation team, whether internal or external, to report on pressures or obstructions encountered during the evaluation process that could have affected — or did affect — their independence or objectivity.[29] Some of the above information would come from the commissioners of the evaluation, and some from the evaluation team. (See also paragraphs 5.7 and 5.8, and Chapter 17, Final Reports and Other Evaluation Products.)

---

29. If it were to become common practice that evaluators report on such pressures encountered during the course of their work to their own community of peers (such as a professional network of evaluators), the program and its constituents would be less inclined to exert such pressures.

# PARTICIPATION AND TRANSPARENCY IN MONITORING AND EVALUATION PROCESSES

## 4. Participation and Inclusion

### Principles and Norms

#### BUILDING PARTICIPATION INTO THE EVALUATION PROCESS

4.1 Participation in a program-level evaluation involves a continuum that ranges from consultation at key points of decision making to full collaboration at all stages. Planning for a sufficient level of participation and consultation in the evaluation process should take place at the programming stage, since this materially affects the time frame and budget for the evaluation.

4.2 Participation in the evaluation process should also be considered as part of the program's monitoring and evaluation policy, since program-level evaluations are not conducted in isolation, but often build on earlier ones and set expectations for future ones.

#### CONSULTATION OF STAKEHOLDERS ESSENTIAL

4.3 The M&E policy should establish a minimum standard for participation and inclusion in program-level evaluations. At a minimum, consultation with an identified set of stakeholders is essential at key stages of the evaluation process — planning, design, conduct, and follow-up — in order to improve credibility, to enhance programmatic learning, and to sharpen the quality of program results.

Draws on UNEG Norm 10

4.4 Indicators to assess (a) participation levels, (b) quality of participation, and (c) effectiveness of participation in enhancing program results should be built into the M&E frameworks of GRPPs.

#### PURPOSE OF PARTICIPATION IN EVALUATION

4.4 Participatory evaluation is a learning process (in and of itself) that can increase programmatic learning and ownership of the program. Participation *adds value* — the more participatory the process, the more value can potentially be added to the program, as learning is extended to the program, its implementers, and its beneficiaries during the evaluation process. Ultimately, participation in evaluation facilitates consensus-building and ownership of evaluation findings, conclusions, and recommendations.

4.5      Developing participatory monitoring systems can *enable* participatory evaluation. Participation in programmatic monitoring builds the capacity of implementers and beneficiaries, which helps to sustain programmatic results after program financing ceases.

4.6      Having stakeholders participate in M&E — particularly program participants from developing countries — can provide an opportunity for learning by doing and can strengthen skills and capacities in beneficiary groups. Such opportunities for participation and capacity strengthening should be identified at the time that the M&E framework is first established, and again when individual evaluations are planned.[30]

### IDENTIFICATION OF STAKEHOLDERS

4.7      For the purposes of this Sourcebook, the term *stakeholders* refers to the parties who are interested in or affected, either positively or negatively, by the program. The term *partners* refers to stakeholders who are involved in the governance or financing of the program (including the members of the governing, executive, and advisory bodies), while the term *participant* refers to those involved in the implementation of the program (including the final beneficiaries). Both partners and participants are subsets of stakeholders. Stakeholders are often referred to as "principal" and "other," or "direct" and "indirect."[31]

4.8      The program's governing body and management unit should assist evaluators with the identification of a core group of representative stakeholders, paying attention to avoid capture by special interests, or individual groups. The program should also help evaluators to identify "excluded" stakeholder groups, where these exist. The complete list of stakeholders, or "stakeholder map," should also point out the agreed-upon or perceived roles and responsibilities of the stakeholders identified. This mapping exercise should be a routine programmatic function, updated regularly, and reproduced transparently in the evaluation TOR.[32]

---

30. The need for beneficiaries to play a leading role in monitoring traditional development assistance is central to the Paris Declaration on Aid Harmonization, March 2, 2005.

31. While other or indirect stakeholders — such as taxpayers in both donor and beneficiary countries, visitors to a beneficiary country, and other indirect beneficiaries — may have interests as well, these are not ordinarily considered in evaluations unless a principal stakeholder acts as their proxy.

32. This UNEG standard has been elevated to a principle. The U.S. Program Evaluation Standards and the African Evaluation Association Guidelines cite among "utility standards" the importance of "stakeholder identification" in

4.9    While evaluators have the right to engage in wider consultations than those specified in the TOR, the process concerning who has been consulted or included in the evaluation process and how they have been chosen should be transparent.

## CAREFUL CONSIDERATION NEEDED TO DETERMINE THE DEGREE OF PARTICIPATION IN THE EVALUATION

4.10    "While not all evaluations can be participatory to the same degree, it is important that consideration be given to participation of stakeholders, as such participation is increasingly recognized as a critical factor in the subsequent use of findings, conclusions, recommendations, and lessons. Also, including certain groups of stakeholders may be necessary for a complete and fair assessment."[33]

UNEG Standard 4.10, para. 17

## ADDITIONAL BENEFITS OF STAKEHOLDER PARTICIPATION IN THE M&E PROCESS

4.11    Broader participation of stakeholders further enhances the quality and credibility of M&E and the likelihood of appropriate follow-up action. Whenever possible, both partners and participants should be involved in the evaluation process. Participation of implementers and beneficiaries is particularly important since they are responsible for sustainability of program outcomes after the program's involvement ceases. Where there are countries with more developed M&E cultures involved in the implementation of GRPP activities, and where M&E has been institutionalized within relevant government ministries, GRPP evaluations should support these systems by considering inclusion of specialist representation from these countries in the evaluation process.

Elaborates on and applies DAC Principle VI, para. 2.3, to GRPPs

4.12    The participation of, or at least consultation with, other program participants is also important, so that their perspective as contributors, users, and beneficiaries may be incorporated in the analysis and findings. Such participants may include organizations (governmental, nongovernmental, or private), households, or individuals. The nature of their participation in M&E depends on their role in the activities funded. Participatory approaches to M&E are particularly important in activities which affect the incomes and livelihoods of local groups, especially disadvantaged populations in and around activity sites (such as indigenous communities, women, and poor households).

Elaborates on GEF Policy, section 2.8, para. 47

---

order to assess their needs, include them in the evaluation process, and increase the likelihood of stakeholder ownership of the evaluation findings.

33. This UNEG standard has been raised to a principle because of the additional logistical complexity of facilitating the participation of stakeholders in GRPPs that are operating at multiple levels — global, regional, national, and local.

Draws on DAC
Principle VI

4.13    When considering methodologies, evaluators should always find ways to seek the views of representatives of beneficiary groups in evaluating the *effectiveness and reach* of programs; in assessing the quality of services to their constituents; and, if practical, in interpreting the analytical results and in reviewing the findings. (See also Chapter 10, Effectiveness.) Input from actual beneficiaries concerning the goods and services delivered by the program always enhances the credibility and quality of an evaluation, even if sought only on a sample basis.

SEEKING AND INCORPORATING STAKEHOLDERS' COMMENTS

Based on DAC
Standard 8.1

4.14    Key stakeholders should be given the opportunity to comment on findings, conclusions, recommendations, and lessons learned.[34] Stakeholder comments should be sought, should be disclosed, and should be reflected appropriately in the final evaluation.

## Standards and Guidelines

PARTICIPATION IN PLANNING OF THE M&E FRAMEWORK

Draws on and
applies DAC
Standard 4.3
and UNEG
Standard 3.11 to
GRPPs

4.15    Both partners and participants should be given the opportunity to provide early input into the M&E framework and processes, and into the programming of specific evaluations. Depending on the size and nature of the GRPP, participation and inclusion processes may add to the cost of programmatic M&E because of the multiple levels, diverse activities, broad geographical scope, and large number of beneficiaries that are characteristic of many GRPPs.

4.16    While involving stakeholders in evaluations of GRPPs may be perceived as costly, moderated e-discussions could be used throughout various stages of an evaluation process to seek inputs and views from beneficiaries in the concerned regions who have access to the appropriate technologies. For those without such access, it may be necessary to use more direct means, such as local contact points, and to consult with them and keep them informed at key stages (such as design, mid-point, and draft recommendations).

---

34. This DAC Standard has been elevated to a principle, since this "right to comment" should be stated in the program's M&E policy. The identification of *key* stakeholders — usually all those on the governing bodies, and sometimes the implementing partners — is left to the discretion of the governing body.

4.17    It is the obligation of the program's governing body and management units to provide the evaluation team with a clear "stakeholder map," including the roles and responsibilities of those identified. In particular, the program should clearly articulate how it differentiates its program partners (those who convened and govern the program) from the program's wider participant base, according to their respective roles in the governance and implementation of the program. GRPP management should also be prepared to identify other beneficiaries in their stakeholder base, particularly as the GRPP evolves and seeks to measure their welfare outcomes.

4.18    Evaluators should be aware that many programs view the meaning of the term *partner* as broader than the members of the governing, executive, and advisory bodies, and also including some of the following:

- Institutional partners with whom the program conducts joint or parallel activities at the global/regional level
- Financial partners not involved in governance
- Participants at the annual forum or general meeting who may or may not have voting rights, but are otherwise not regularly involved in governance
- Beneficiary countries
- Implementing partners of all types, including other international organizations, government agencies, the private sector, and international and local NGOs.

4.19    Because of the variation in the use of these terms, the evaluation TOR should identify the various categories of stakeholders. If this identification is absent from the TOR, evaluators should aim to clarify this before undertaking the evaluation.

4.20    In regional partnership programs, sovereign countries are often the principal *partners* represented on the governing body, since the success of regional programs hinges to a greater extent (than in global programs) on beneficiary country ownership, capacity for collective decision making, and cooperative implementation of program activities. But if beneficiary countries are not represented on the governing, executive, or advisory bodies, it is particularly important that evaluators find ways to include beneficiary country representatives in the evaluation process.

Draws on IEG's forthcoming review of regional programs

STAKEHOLDER LEARNING OPPORTUNITIES

4.21    "The evaluation approach must consider learning and participation opportunities (e.g., workshops, learning groups, debrief-

UNEG Standard 3.11, para. 23

ings, participation in the field visits) to ensure that key stakeholders are fully integrated into the evaluation learning process."

| | |
|---|---|
| UNEG Standard 3.11, para. 24 | 4.22 "When feasible, a core learning group or steering group composed of representatives of the various stakeholders in the evaluation may be created. This group's role would be to act as a sounding board, and to facilitate and review the work of the evaluation. In addition, this group may be tasked with facilitating the dissemination and application of the results and other follow-up action, particularly with their own constituents." |

### REPORTING ON PARTICIPATION AND CONSULTATION

| | |
|---|---|
| Elaborates on DAC Standard 4.3, UNEG Norm 11, para. 11.2, and UNEG Standard 4.18 | 4.23 The rationale for the degree of participation chosen for the evaluation should be included in the evaluation report, possibly in the preface prepared by the commissioners of the evaluation. The list of people interviewed or the characteristics of those surveyed should always be included in the evaluation report. In addition, the criteria for determining those consulted should be presented. This would include the choice of countries and locations for site visits or case studies, sampling methodology if applicable, and criteria for choosing those interviewed. Those interviewed should be given a chance to review any quotations attributed to them. |
| Draws on DAC Standard 8.1 and UNEG Norm 8, para. 8.2 | 4.24 At the time of the issuance of the final report, any substantial differences of view that remain should be transparently presented. In disputes about facts that can be verified, the evaluators should investigate and change the draft where necessary. A separate section of the report, or an annex, may be set aside for views of particular stakeholders, if needed. In the case of differing opinions or interpretation, care should be taken that the reporting of stakeholders' comments does not conflict with the rights and welfare of evaluation informants. |

# 5.    Transparency and Disclosure

## Principles and Norms

### RATIONALE FOR TRANSPARENCY AND DISCLOSURE IN EVALUATION

5.1    The transparency of the evaluation process is crucial to its credibility and legitimacy. It can facilitate consensus-building and ownership of the findings, conclusions, and recommendations among stakeholders.

*Based on DAC principle IV, para. 20, and UNEG Norm 10*

5.2    The provision of evaluation information to the public is necessary for the evaluation to achieve one of its main purposes — to provide a basis for accountability and responsibility.

*Based on DAC Principle II, para. 6*

### OPENNESS OF THE EVALUATION PROCESS

5.3    To promote transparency and legitimacy, both the monitoring and the evaluation processes should be as open as possible.[35] Particularly for evaluations, the whole process should be open, with results made widely available, not only to stakeholders, but also to other concerned entities in the same sectors, academia, research institutions, civil society, and the public. Evaluations should be conducted, and evaluation findings and recommendations presented, in a manner that is easily accessible and understood by target audiences.

*Based on DAC Principles IV and V, UNEG Norms 10 and 13, and GEF Policy, section 5.2*

5.4    The process of selecting evaluation teams should be as transparent as possible. For GRPPs, all governing partners should have the chance to review and approve the selected evaluation team. Formally recording the approval of the evaluation team at the end of the selection process can also help to prevent related disputes later on. (See also Chapter 7, Evaluation Team Selection.) As a general rule, conflicts of interest relating to evaluation team members should be disclosed in the final evaluation report, even if measures are taken to mitigate their effects.

5.5    Information should be provided up front to evaluation informants about the scope and limits of confidentiality that will be observed during the evaluation process as well as their prospective access to information on evaluation results downstream.

*Elaborates on UNEG Standard 2.7*

---

35. On issues of transparency, the principles for monitoring and evaluation diverge to some extent. The evaluation process always needs to be open to ensure credibility and independence and to support accountability. However, monitoring may take place in different ways to support management's needs. Only selected monitoring information, for example, may be made available to governing bodies. Monitoring of staff performance should always be confidential. Thus, management should have some discretion in determining the appropriate disclosure of monitoring information. (See also footnote 72 on page 78.)

Based on UNEG Norm 3

5.6     Each GRPP should develop a policy on transparency, disclosure and dissemination which may be a part of the program's overall evaluation policy. (See paragraphs 2.3 to 2.5 and 2.18.) The policy should specify that:

Elaborates on UNEG Norm 2, GEF Policy, section 3.3, para. 62c, and U.S. Program Evaluation Standard P6

- Transparency is required, with respect to both the stakeholders of the program and the general public. Clear communications are necessary with stakeholders at all stages of the evaluation with respect to its purpose, the criteria applied, and the intended use of the findings. However, the policy might differentiate the types of information to be made available to different categories of stakeholders at different stages of the evaluation. The policy should provide for processes to ensure that the full set of evaluation findings, along with pertinent limitations, is made accessible to any persons with expressed legal rights to receive the results.[36]

- Governing body members in particular have a responsibility to facilitate the relay of evaluation findings and results to their constituencies. Any policy statement concerning disclosure and dissemination of an evaluation should include a dispute-resolution clause in the event that one or more governing parties abrogate the policy once the evaluation findings are put forward.

Elaborates on UNEG Norms 2, 4, and 10

- Annual and multi-year M&E plans and work programs, TORs for evaluations, and evaluation findings should be made available to stakeholders and to the public on a timely basis.

Elaborates on DAC Standard 5.1, and UNEG Norm 10 and Standard 1.4

- Final evaluation reports should be public documents. The assumption should be that evaluators have respected cases where anonymity of respondents needed to be preserved. However, the policy statement should ideally lay out review procedures to ensure, before dissemination, that the reasonable protection and confidentiality of particular stakeholders have been respected when required.

- The policy on disclosure and dissemination should also cover treatment of interim findings and draft reports, which may depend on the degree of emphasis on participation and consultation with stakeholders.

---

36. The African Evaluation Guidelines modified the U.S. Program Evaluation Standard to add the following: "The evaluation team and the evaluating institution will determine what is deemed possible, to ensure that the needs for confidentiality of national or governmental entities and of the contracting agents are respected, and that the evaluators are not exposed to potential harm."

5.7     The final evaluation report should provide details about the evaluation process, so that key important parameters of the evaluation process are disclosed. The commissioners of the evaluation should prepare a preface to the final evaluation report that provides the background information that should rightly come from them. This includes who commissioned the evaluation, how it was managed, who funded it, to whom the evaluation team reported, and how the draft report was reviewed — all of which is important in assessing the degree of independence of the evaluation. In addition, the governing body may wish to disclose, for accountability purposes, other key information such as the method and criteria used for selecting the evaluation team and the budget or final cost of the evaluation, subject to any prohibitions that may arise from the applicable donors' legal restrictions. (See also paragraph 3.17 and Chapter 17, Final Reports and Other Evaluation Products.)

5.8     The evaluation team may also wish to provide some additional information on the evaluation process from their own perspective, such as factors they think may have hindered the independence or quality of the evaluation — for example, political obstructions; limitations on access to information; and restrictions on budget, travel, or sampling, and the like.

## NORMS FOR DISSEMINATION TO FACILITATE KNOWLEDGE SHARING AND LEARNING

5.9     As stated in Chapter 4, participatory evaluation is a learning process (in and of itself) that can increase programmatic learning and ownership of the program. Whether or not evaluation is participatory, systematic and targeted dissemination of evaluation results to stakeholders is essential in facilitating learning and ensuring improved planning and implementation of GRPP activities, and an explicit budget should be set aside for this. In addition, wider dissemination, beyond key stakeholders, should be considered by commissioners of the evaluation and program management in order to contribute to broader knowledge sharing and development effectiveness — such as dissemination to donors outside the program, other GRPP governing bodies and management units, and other international organizations involved in development.

*Elaborates on DAC Principle X, para. 41*

5.10    To have an effect on decision making, easy accessibility to evaluation findings is crucial. Feedback loops should be established to policy makers, operational staff, beneficiaries, and the general public. Documentation emanating from M&E should be made available in easily consultable and readable form, including editions in local languages as necessary and feasible.

*Based on DAC principle V, UNEG Norm 10, and GEF Policy, section 3.3*

5.11    In particular, findings and lessons from M&E activities should be made available to stakeholders directly involved in activity formulation and implementation at the country level. Dissemination strategies should be based on user needs and priorities, and use the latest technologies and approaches, where feasible.

5.12    Evaluation results may be disseminated in several ways apart from the evaluation report itself. The dissemination strategy should be tailored to the audience: for example, annual reports that provide a synthesis of findings for stakeholders; abstracts/summaries that provide a synopsis of findings in appropriate languages for country or local participants; and workshops that are conducted by local implementers in areas affected by the programs.

5.13    All evaluation products should be complemented by a disclosure statement made by the governing body that details the disclosure policy that applies to their dissemination. It is also good practice to disclose dissemination plans in the evaluation report.

## Standards and Guidelines

### TAILORING COMMUNICATIONS TO AUDIENCE

5.14    Communications to a given category of stakeholders should always include all important results that may bear on the interest of these stakeholders. Evaluators should also strive to present results as clearly and simply as possible so that stakeholders can easily understand the evaluation process and results.

5.15    The final report should be posted on the program Web site in order to be available to the public both for accountability and for knowledge-sharing purposes. However, other means of disseminating findings, lessons, and recommendations that are more accessible to key stakeholders should also be considered. Ideally, those interviewed and/or consulted should receive copies of the final report and other products, as appropriate, and in an easily accessible form, which may be a hard copy. Other means, such a workshops and/or constituent meetings in local languages, also need to be considered if it is important to communicate results to poor beneficiaries.

### RESPONSES TO EVALUATION RECOMMENDATIONS

5.16    The response of the governing body and/or management unit to the evaluation — which reports what will be done, who will do it and by when — should be made public when available. Ideally, this should be posted on the Web site alongside the evaluation report in order to make transparent the effect that the evaluation has had on program strategy and plans.

# PLANNING AND CONDUCT OF EVALUATIONS

## 6.  Planning for Scope and Methodology

### Principles and Norms

ENSURING QUALITY OF EVALUATION

6.1  Each evaluation should employ processes that are quality oriented, and use appropriate methodologies for data collection, analysis and interpretation.[37]

*Based on UNEG Norm 8*

COVERAGE OF EVALUATION AND TERMS OF REFERENCE

6.2  Each evaluation should be planned and a TOR drawn up to:

*Elaborates on DAC Principle IX, para. 32*

- Define the rationale, purpose, and scope of the evaluation, including the specific objectives of the evaluation and the main audience for its findings. The purpose will almost always include an assessment of performance and results to date, and will also answer other strategic questions relating to governance and management, financing, scope, or a particular policy. (See also Chapters 16 and 17 for complete checklists with respect to TORs and the content of evaluation reports.)

- Define clearly the subject of the evaluation — a GRPP or subset of its activities, or possibly several GRPPs in the same sector — as well as the contextual factors and issues that need to be understood and that affect the methodology chosen. In the case of a GRPP, essential contextual background information includes the circumstances surrounding the origin of the program, its maturity, its objectives, the coverage and range of activities supported, the identification of stakeholders, trends in expenditures, and expected outcomes and impacts for specific target groups.[38]

- Define the criteria by which the program will be judged. In

---

37. The U.S. Program Evaluation Standards and the African Evaluation Guidelines have "accuracy standards" related to information gathering, analysis, drawing of conclusions, reporting, and meta-evaluation.

38. This treatment should include the "raison d'être" of the program — that is, why global or regional collective action was deemed necessary or useful, and what additional features the partnership brings to the program.

addition to the standard criteria of relevance, effectiveness, efficiency, and sustainability, the legitimacy and effectiveness of the governance and management arrangements will almost always need to be covered.

- Define the evaluation issues and questions that will be addressed in the evaluation, such as the continued rationale for the program, the achievement of objectives, factors influencing the achievement or non-achievement of objectives (both internal and external to the program), and other outcomes and impacts (both intended and unintended). Aspects of the performance of the host organization and/or the program's partners could also be addressed where the host organization is performing some governance or management functions on behalf of the program and where the partners have made specific commitments to the program (such as pledges to provide funding). However, the inclusion of these in the TOR should be cleared with the host organization and the partners, respectively. (See also paragraph 12.9.)

- If desired, allow some evaluation issues and questions to remain open until key stakeholders have been consulted during the course of the evaluation, in case additional issues are raised from those stakeholders' perspectives. Any substantial changes in the objectives and scope of the evaluation should be communicated to the commissioner of the evaluation, and an amended TOR or budget approved as required.

- Define the methods and techniques to be used to address the issues identified, including proposed methodologies for gathering of information (existing or new), analysis of the information, and interpretation of the results of the analysis. For GRPPs, it is normally necessary to collect data on results at the program, portfolio, and activity levels. It is particularly important to have a representative sample of activities, since GRPPs have a large scope and multiple beneficiaries. An attempt should be made to establish causal relationships in accordance with an existing results chain or logframe, if available, while acknowledging the complexities and identifying assumptions and limitations. If a comparison to a counterfactual is to be attempted, stakeholder consultation is desirable to agree on the counterfactual.[39]

- Determine the resources and time required to complete the evaluation.

---

39. The *counterfactual* is the situation or condition that hypothetically would have prevailed if there had been no development intervention. OECD/DAC, *Glossary of Key Terms in Evaluation and Results Based Management*, 2002.

## Standards and Guidelines

6.3     The *rationale* of the evaluation describes why and for whom the evaluation is undertaken and why it is undertaken at a particular point in time. The *purpose* of the evaluation is usually in line with the learning and accountability functions of evaluations — such as (a) to contribute to improving the effectiveness of the program; (b) to consider a continuation, discontinuation, or change in scope of the program; or (c) to account for aid expenditures to stakeholders and taxpayers. The *objectives* of the evaluation specify what the evaluation aims to achieve — such as (a) to ascertain the results of the program; (b) to assess the relevance, effectiveness, and efficiency of the program; and (c) to provide findings, conclusions, and recommendations with respect to specific aspects of the program. As noted in Chapter 2, both the purpose and objectives of the evaluation are likely to be different at different stages of program maturity. (See Tables 5 and 6 for more detailed guidelines.)

*Based on DAC Standards 1.1, 1.2 and 1.3*

6.4     The evaluation objectives should follow from the purpose of the evaluation. They should be realistic and achievable — taking into consideration the scope of the program, the number of activities, and the quality of data on the one hand, and the overall time frame, resources, and level of participation and consultation expected on the other hand. The objectives of the evaluation should be clear and agreed upon by all partners involved.

*Elaborates on UNEG Standard 3.5*

SCOPE OF THE EVALUATION

6.5     The scope of the evaluation should be clearly articulated by defining the time period covered by the evaluation, the interventions and activities to be included in the evaluation, and any delimitation on geographical or target group coverage. Any restrictions in scope should generally be justified by the rationale and purpose of the evaluation and explained in the evaluation report. In general, restrictions in scope are not justified for evaluations conducted for the purpose of satisfying accountability needs.

*Elaborates on DAC Standard 2.1 and UNEG Standard 3.5*

6.6     GRPPs are distinguished by the expectation of benefits arising from the partnership, over and above the benefits associated with the discrete activities supported (whether global, regional, country, or local activities). These additional benefits of the partnership may arise from the large scale, from joint activities enabled by the partnership, or from the cross-fertilization and enrichment of knowledge among the large number of partners. Thus, the scope of a GRPP evaluation should ideally encompass the achievement of these additional expected benefits, not just the benefits of the discrete activities supported.

6.7    The purpose and scope of the evaluation are likely to be different at different stages of program maturity. Tables 5 and 6 provide broad guidelines for determining the evaluation scope and questions at different stages of the program.

Table 5. Sample Issues to Feature in the Scope of an Evaluation at Various Stages of the GRPP

| Stage/Timing | Sample Issues to Be Examined |
|---|---|
| A. Early Stage (first 2–3 years) | Design of the program: Is it appropriate or in need of adjustment? |
| | Institutional structures of the program: Are governance and management arrangements in place and functioning as planned? |
| | Resource mobilization: Have the needed resources been mobilized for governance, management, and M&E needs, and is there a strategy for growth to support a growing portfolio? |
| | Mechanism for monitoring and evaluation: Have appropriate M&E mechanisms been set up? |
| | Program performance: Using process indicators, are input, activity, and output targets being met? |
| | Capacity building: Are steps being taken to close the gaps in the capacity of national or local institutions, as applicable? |
| B. Established Stage (over 5 years old) | Operations: Are these functioning as designed? |
| | Sources and uses of funds: Are the inflows of funds stable or growing? Are the allocation processes and reviewing of proposals (if applicable) working as planned? |
| | Targets: Are these being met? |
| | Capacity strengthening: Are national institutions being strengthened? Is technology being transferred? |
| | Outcomes and impacts: Are expected goals being met? |
| | Strategic direction: Given the above, are the program's strategic directions correct and on course? For growth, outreach to new donors and partners, broadening the target area, devolution, etc.? |
| C. Mature Stage | Outputs: Are outputs matching planned expectations? |
| | Impacts: Are there measurable indictors of the impacts of the program? |
| | Capacity optimization: Are national personnel and institutions capable of performing effectively? |
| | Sustainability: What measures have been taken to ensure the sustainability of the program with regard to financial, institutional, and other resources? |
| | Strategic direction, devolution, and possible exit: What arrangements have been made for the planned transfer of implementation responsibilities or withdrawal of external resources — financial, technical, etc.? |

*Source*: Both this and the following table have been adapted from material provided by Dr. Adetokunbo Lucas, based on his experience in evaluating GRPPs and in serving on the governing bodies of GRPPs.

Table 6. Schematic Representation of a Life-Cycle Approach to Determining the Scope of an Evaluation

| Issues | Program Stage | | |
|---|---|---|---|
| | Early | Established | Mature |
| Program Design | +++ | ++ | + |
| Governance Structures | +++ | ++ | + |
| Management Structures | +++ | ++ | + |
| Resource Mobilization Strategy | +++ | +++ | ++ |
| Inputs | +++ | +++ | ++ |
| Activities | +++ | +++ | +++ |
| Outputs | ++ | +++ | +++ |
| Reach | ++ | +++ | +++ |
| Outcomes | + | ++ | +++ |
| Impacts | + | + | +++ |
| Sustainability | + | ++ | +++ |
| Devolution or Exit Strategy | + | ++ | +++ |
| Miscellaneous Topical Issues | ? | ? | ? |

Priority: High +++ , Medium ++ , Low + , To be determined ?

## FACTORS AFFECTING THE CHOICE OF METHODOLOGY

6.8    The objectives and scope of the evaluation are critical references in determining the evaluation methodology. The issues and questions to be addressed, the type of information sought, the logframe (if available), the nature of the results chain, and the quality of data already available also affect choice of methodology. Within this broad context, evaluation methodologies should be sufficiently rigorous to ensure a complete, fair, and unbiased assessment of the GRPP. Budgets should be sufficiently flexible to allow for this level of rigor and quality. Another important factor that affects the methodology is the level of participation and consultation of stakeholders desired. The evaluation TOR should include a "stakeholder map" of the various stakeholders and their roles, and indicate the level and type of participation expected. (See paragraph 4.8.)

*Elaborates on UNEG Standards 3.5 and 3.7*

6.9    GRPPs are highly diverse, and evaluation questions and methodologies need to be tailored to the specific sectors in which they operate. Where applicable, evaluation questions should consider whether private sector behavior or commercial market forces have influenced results, and what would be the best way of capturing these effects, taking into account the different incentives involved.

6.10    Care should be taken in planning the methodology to consider up front whether and how issues relating to gender and underrepresented groups are to be addressed.

*Based on UNEG Standard 3.7*

## ENSURING AN APPROPRIATE CHOICE OF METHODOLOGY

6.11    Since most GRPPs do not have a specialized evaluation unit or professional evaluators on the governing body, it is advisable to plan for the input of expert evaluators in determining the methodology before the TOR and contracts are finalized. Evaluators from the specialized evaluation units of one or several partners may be called upon to contribute, or consultants may be hired specifically to advise on planning the evaluation (in which case they should not be eligible to compete to conduct the evaluation). Alternatively, a skilled peer group could be called on to advise the governing body and review the TOR to ensure that appropriate parameters of evaluation design are specified.

6.12    While the evaluation methodology should be planned up front and specified in the TOR, the processes of planning and managing evaluations may also provide for the teams that are bidding to conduct the evaluation to make specific proposals, or for the team, once selected, to provide further details of the methodology in an inception report in the early stage of the evaluation process. Under all circumstances, it is helpful for the commissioner of the evaluation to interact with the evaluation team before work begins in order to clarify expectations. Any amendments to the scope and methodology that result from these processes should be endorsed by the governing body commissioning the evaluation and reflected in a formally revised TOR, which is then attached to the final evaluation report. The evaluation team should be held to the final TOR.

## ABSENCE OF AN ADEQUATE M&E FRAMEWORK

6.13    If the evaluation team finds that the objectives of the program are unclear, that an M&E framework is lacking, or that needed data are absent or of questionable quality, the team may advise that an evaluability assessment be conducted as a prerequisite to the evaluation. (See paragraph 2.7.) Also, if there is an M&E system, but it is not producing the quality of information necessary for an evaluation, a process audit might be advisable to see what constraints have been limiting its usefulness. In either case, there is a need to define a time frame for a decision (a) to create the conditions for a successful evaluation or (b) to go ahead with an evaluation, while acknowledging the inevitable limitations.

6.14    If it is decided to go ahead and evaluate with limitations, the evaluation team could agree with the commissioner of the evaluation to reconstruct a results framework and baseline information, as needed. This may require changing the TOR, the time frame, or the budget for the evaluation. (See also paragraphs 9.5 and 9.6.)

6.15    Key actions to create enabling conditions for quality evaluations are outlined in Chapter 2. These include (a) establishing or adapting the M&E framework to meet the requirements for evalua-

tion of the GRPP (such as collecting information on all interventions at all levels); (b) assigning clear responsibilities for M&E; and (c) ensuring that the responsibilities for monitoring and evaluation are separate in order to assure independence of evaluation.

## USE OF EXISTING EVALUATIVE INFORMATION

6.16    Ideally, the evaluation of the performance and results of a GRPP relies not only on information at the program level, but also on summary portfolio information and on more detailed information at the country and activity levels in order to validate overall findings. When using information from the country or activity level, evaluators should always make explicit whether they have merely accepted information or ratings provided by management or by the country (which would generally be only self-evaluations), or whether they have undertaken to validate such information through independent assessments based on field observations.

Draws on IEG's experience with reviewing GRPPs

## EVALUATION CRITERIA

6.17    The most common criteria for evaluating development assistance, endorsed by both the DAC and the UNEG, apply equally to GRPPs. These are relevance, effectiveness (or efficacy), efficiency, impact, and sustainability. The UNEG Standards also list additional criteria that may be used: value-for-money (which is an aspect of efficiency) and target group satisfaction.[40]

Based on DAC Standard 2.3 and UNEG Standard 3.6

6.18    To aid the reader, the criteria should be defined in unambiguous terms in the report. If one of the usual criteria specified above is not applied, this should be explained in the evaluation report, along with any additional criteria that were used. (See Chapters 9 to 15 for more guidance on evaluation criteria and questions relevant to GRPPs.)

Based on DAC Standard 2.3

## CONSIDERING POSSIBILITY OF PEER REVIEW

6.19    Depending on the scope and complexity of the evaluation, it may be useful to establish a peer review or reference group composed of experts in the technical topics covered by the evaluation, or experts in evaluation itself. This group would provide substantive guidance to the evaluation process (such as providing feedback on the methodology, analysis, and interpretation of results) and provide quality control of the draft reports.

Based on UNEG Standard 3.12

---

40. One multilateral agency uses the criterion of additionality or value added. There may be trade-offs in achieving results relating to different criteria (such as between effectiveness and efficiency), which may need to be taken into account in assessing the performance of the program in relation to each criterion. (See also Chapter 12, Governance and Management.)

# 7. Evaluation Team Selection and Contracting Process

## Principles and Norms

### IMPORTANCE OF CAREFUL SELECTION OF EVALUATION TEAM

7.1     "The credibility of evaluation depends on the expertise and independence of the evaluators and the degree of transparency of the evaluation process."

DAC Principle IV, para. 18

### SELECTION CRITERIA

7.2     Evaluators must be competent.[41] They must have the basic set of skills for conducting evaluation studies and managing evaluation team members.

Based on UNEG Norm 9

7.3     Commissioners of evaluation should endeavor to ensure that evaluators selected are impartial and unbiased.

Based on GEF Policy, section 3.3, para. 62b

## Standards and Guidelines

### SELECTION PROCESS AND CRITERIA

7.4     "Evaluators should be selected on the basis of competence, and by means of a transparent process."

UNEG Standard 3.13, para. 27

7.5     Evaluators should accurately represent their level of skills and knowledge; they should decline to conduct evaluations that fall outside the limits of their professional training and competence.

Based on UNEG Standard 2.1, para. 6

7.6     "Evaluators should declare any conflict of interest to clients before embarking on an evaluation project and at any point where such conflict occurs."

UNEG Standard 2.1, para. 5

### COMPETENCIES

7.7     Evaluators should have relevant educational background, qualifications, and training in evaluation, preferably an advanced university degree or equivalent background in the relevant disci-

Based on UNEG Standard 2.2, para. 7

---

41. The U.S. Program Evaluation Standards and the African Evaluation Guidelines also stipulate that the persons conducting the evaluation should be trustworthy in order to ensure credibility and acceptance.

plines, with specialized training in areas such as evaluation, project management, and advanced statistical research.[42]

<table>
<tr><td>UNEG Standard 2.3</td><td>7.8      "Evaluators should have professional work experience relevant to evaluations."</td></tr>
<tr><td>Based on UNEG standard 2.4</td><td>7.9      Evaluators need to be familiar with, and have specific technical knowledge of, the methodology or approach that will be needed for the evaluation.</td></tr>
</table>

METHOD OF SELECTION

Draws on DAC *Guidance for Managing Joint Evaluations*, and World Bank procurement guidelines and practices

7.10     Selecting an evaluation team that is acceptable to all partners is a challenge in joint evaluations, or in evaluations commissioned by a governing body made up of donors with diverse policies and procedures. Commissioners of evaluation should follow rules on selection of consultants as stipulated (a) in their charter and/or evaluation policy, (b) in the rules and procedures of trustees or host organizations, and (c) in administration agreements with donors. Any conflict among these should be transparently discussed and resolved, with the results disclosed to all relevant parties.

7.11     Competitive methods of selection should be favored, with justification provided if a non-competitive selection method is adopted. Competitive bidding is better for transparency, value-for-money, and competition on substance. Competitive bidding processes differ, and many joint or GRPP evaluations have followed the European Union, United Nations, or World Bank rules and procedures. A prequalification exercise may be used to identify consultants, who are then invited to submit a full bid. Criteria to encourage participation of local experts may also be included in the selection criteria.

7.12     The method of selection and any justification required should be disclosed in the evaluation report. This may be in an annex that also describes other aspects of the evaluation process, such as the ways in which independence was ensured.

7.13     All bidders for an evaluation contract should be notified of results. Good practice is to post results publicly.

---

42. Various standards and guidelines of professional evaluation societies also include "competency" among their standards of ethics. For example, the Canadian Evaluation Society Guidelines for Ethical Conduct state: "Evaluators should apply systematic methods of inquiry appropriate to the evaluation; evaluators should possess or provide content knowledge appropriate for the evaluation; evaluators should continuously strive to improve their methodological and practice skills." The American Evaluation Association Guiding Principles for Evaluators also adds "cultural competency."

## TIME FRAME FOR SELECTION OF CONSULTANTS

7.14    As an approximate guide, a minimum of three to four months will be needed from the publication of an invitation to bid to the completion of the negotiations with the evaluation team, in order to allow for consensus to be reached among the partners.

Based on DAC *Guidance for Managing Joint Evaluations*

## AVOIDING CONFLICTS OF INTEREST

7.15    The selection process should ensure that all candidates disclose their prior involvement with the program and agree not to be involved in the implementation of the recommendations. But the pool of candidates from which to draw evaluators with the required technical skills, knowledge, and experience may be limited because of the unique aspects of GRPPs and their relative newness in international development. This increases the potential for conflicts of interest because qualified candidates may have had some manner of prior involvement with the program.

7.16    To avoid compromising the independence of the evaluation under such circumstances, an oversight committee or external panel could help the governing body select the evaluation team and ensure that there are always some professional and unbiased evaluators on the team. The governing body, oversight committee, or external panel should work out mutually acceptable ways of mitigating conflicts of interest when these arise. (See also paragraphs 3.5 and 3.6 on institutional arrangements for independence.)

7.17    If a potential conflict of interest arises during the course of the evaluation, the managers of the evaluation should identify and implement ways of diminishing its implications for independence and impartiality. They should also disclose the initial conflict and the actions taken to the governing body and to program management.

7.18    As a general rule, all conflicts of interest, and any actions taken to mitigate them, should be disclosed in the final evaluation report. This includes the disclosure by evaluators who had prior involvement in the program.

## SIZE AND COMPOSITION OF THE EVALUATION TEAM

7.19    The number of evaluators in a given team depends on the budget and scope of the evaluation and the degree to which a multidisciplinary team is required.

Based on UNEG Standard 3.13, para. 26

7.20    Evaluation teams should possess a mix of evaluation skills and technical or sectoral/thematic knowledge relevant to the particular evaluation. At least one member of the team evaluating a GRPP

Elaborates on DAC Standard 4.5 and UNEG Standard 2.1, para. 2

should have knowledge or experience with multidonor programs, including the governance and financing issues associated with them.[43]

7.21    The lead evaluator or team leader should ensure the overall integrity of the team's performance. He or she should possess core evaluation competencies — that is, the qualifications, skills, experience, and attributes generally expected of evaluation professionals — and the ability to manage potential conflicts of interests that arise when the technical/sector experts on the team have had prior involvement with the program.

7.22    There has sometimes been a suggestion to include staff from partner agencies in a GRPP or joint evaluation. This can facilitate communications and strengthen ownership of the findings, but may lead to conflicts of interest that undermine the neutrality and credibility of the evaluation. Possible ways of increasing participation while minimizing conflicts of interest are (a) to accord observer status only; (b) to include nationals who are not employees of the agency being evaluated; or (c) to include staff of the independent evaluation office of the agency being evaluated, if there is such an office, and its degree of independence can be verified.[44]

7.23    To the degree possible, the composition of evaluation teams should be gender balanced and geographically diverse, and include professionals from the countries or regions concerned. In particular, the evaluation of activities in beneficiary countries should make the best possible use of local expertise, both technical and evaluative.

WRITTEN AGREEMENTS

7.24    The responsibilities of the parties who agree to conduct an evaluation should be set forth in a written agreement. The agreement obligates the contracting parties to fulfill all the agreed upon conditions or to renegotiate the contract. Such an agreement reduces the likelihood that misunderstandings will arise between the contracting parties and makes it easier to resolve them if they do arise. The

---

43. "Consultants often join together within a consortium when bidding for a large joint evaluation. This can be useful in bringing together team members with varied knowledge and expertise." OECD/DAC, *Guidance for Managing Joint Evaluations*, 2006.

44. While (c) is preferable, a combination of (a) and (c) is also possible. However, the staff of an independent evaluation office who participate in the evaluation of a GRPP should not subsequently participate in reviews or meta-evaluations of this particular evaluation.

agreement should specify what is to be done (by both parties), by whom and when, and any details on how it is to be done.[45]

7.25    The agreement will generally refer to the TOR, which will provide details, at least in the following areas: financing, time frame, persons involved, reports to be produced, content, methodology, and procedures to be followed. (See Chapter 16, Terms of Reference.)

7.26    The written agreement and/or TOR generally provide for various stages in the process of the evaluation, along with a timeline. Including a stage where evaluators produce an inception report to be reviewed by the commissioners of the evaluation can provide an opportunity (a) to tap the expertise of the evaluators in refining methodologies in response to new information; (b) to allow consideration of more participatory methods; and (c) to clarify expectations on consultation of stakeholders and reporting of progress; and (d) to resolve any other issues that have come up.

Elaborates on UNEG standard 3.10, para. 21

7.27    The relationships between the evaluation team and the commissioner(s) of an evaluation must be characterized by mutual respect and trust from the outset. Commissioners of the evaluation and the evaluation team should aim to clarify early in the evaluation process any matters such as confidentiality, privacy, communications, ownership of findings and reports, and referrals on matters of misconduct discovered, which may not be covered completely in written agreements.

Elaborates on UNEG Standard 3.10, para. 22

---

45. One issue that frequently comes up is whether to use *lump sum* agreements or *negotiated* contracts. There is also the issue of whether each of these options should include allowance for reimbursable expenses. Another legal question that often arises is whether to allow termination in the case of poor performance through a *cancellation* clause or an *option* clause that requires the commissioner to explicitly request the continuation of the work at certain points in the process.

# 8.   Ethical and Professional Conduct of Evaluations [46]

## Principles and Norms

### OVERALL INTEGRITY AND ETHICS

8.1     "Evaluators must have personal and professional integrity. Evaluators must respect the right of institutions and individuals to provide information in confidence and ensure that sensitive data cannot be traced to its sources. Evaluators must take care that those involved in evaluations have a chance to examine statements attributed to them."

UNEG Norm 11, paras. 11.1 and 11.2

8.2     Evaluators must be sensitive to the beliefs, manners, and customs of the social and cultural environments in which they work, including issues of discrimination and gender inequality.

Based on UNEG Norm 11, paras. 11.3 and 11.4

8.3     "Evaluators sometimes uncover evidence of wrong-doing. Such cases must be reported discreetly to the appropriate investigative body."

UNEG Norm 11, para. 11.5

## Standards and Guidelines

### HONESTY

8.4     Evaluators should accurately represent their levels of skills and knowledge. Evaluators should practice within the limits of their professional training and competence.[47]

Based on UNEG Standard 2.1, para. 6

8.5     Evaluators should declare any conflict of interest to clients at any point where such conflict occurs.

Based on UNEG Standard 2.1, para. 5

### ACCOUNTABILITY

8.6     Evaluators must ensure the honesty and integrity of the entire evaluation process. Evaluators have a responsibility to ensure that evaluation activities are independent, impartial, and conducive to producing accurate results.

Based on UNEG Standard 2.5, para. 15

---

46. The principles and standards in this section reflect mainly those advanced by DAC, UNEG, and the GEF. Professional evaluation associations have also developed their own ethical and propriety standards and guidelines.

47. DAC principles and standards also refer to the fact that evaluators should continually seek to maintain and improve their competencies in order to provide the highest level of performance in their evaluations.

| Based on UNEG Standard 2.8, para. 22 | 8.7 Evaluators are responsible for their performance and their product — that is, the clear, accurate, and fair presentation of their report's limitations, findings, and recommendations. |

### CONDUCTING THE EVALUATION WITHIN THE ALLOTTED TIME AND BUDGET

| Based on DAC Standard 9.2 and UNEG Standard 2.8 | 8.8 "An evaluation [should be] conducted and results made available in a timely manner in relation to the purpose of the evaluation. Unenvisaged changes to time frame and budget [should be] explained in the report." Any departure from the planned implementation and products of the evaluation should be explained in the final report. |

### PROFESSIONALISM WITH COST-EFFECTIVENESS

| UNEG Standard 3.15, para. 32 | 8.9 "Evaluations should be conducted in a realistic, diplomatic, cost-conscious, and cost-effective manner."[48] |

| Based on UNEG Standard 3.15, paras. 33 and 34 | 8.10 Evaluations must be accurate and well-documented, and deploy transparent methods that provide valid and reliable information. Key findings should be substantiated through triangulation.[49] |

| Based on UNEG Standard 3.15, para. 35 | 8.11 Evaluators should carefully consider and openly present openly the values, assumptions, theories, methods, results, and analyses that significantly affect the evaluation, from its initial conceptualization to the eventual use of findings. |

### RESPECT FOR STAKEHOLDERS

| DAC Standard 7.1, consistent with UNEG Standards 2.6, 2.7, and 3.15, para. 31 | 8.12 "The evaluation process shows sensitivity to gender, beliefs, manners and customs of all stakeholders. The rights and welfare of participants in the evaluation are protected. Anonymity and confidentiality of individual informants should be protected when requested and/or as required by law."[50] |

---

48. Consistent with this, the U.S. Program Evaluation Standards and the African Evaluation Guidelines have set "feasibility standards" intended to ensure that the evaluation is realistic, diplomatic, and frugal, as well as "prudent."

49. *Triangulation* refers to "the use of three or more theories, sources, or types of information, or types of analysis, to verify and substantiate an assessment. By combining multiple data sources, methods, analyses, or theories, evaluators seek to overcome the bias that comes from single informants, single methods, single observers, or single theory studies." OECD/DAC, *Glossary of Key Terms in Evaluation and Results Based Management*, 2002.

50. UNEG Standards 2.6 and 3.15 also point out that the findings of evaluations might sometimes negatively affect the interests of some stakeholders, so that evaluators need to discuss and be sensitive to this possibility in their contact with stakeholders.

8.13    In conducting interviews and arranging consultation meetings, evaluators should provide maximum possible notice, minimize demands on time, and respect people's rights to privacy.

Based on UNEG Standard 2.6, para. 18

## ACKNOWLEDGING DISAGREEMENTS WITHIN THE EVALUATION TEAM

8.14    "Evaluation team members should have the opportunity to dissociate themselves from particular judgments and recommendations. Any unresolved differences of opinion within the team should be acknowledged in the report."

DAC Standard 7.2

## WRONG-DOING, FRAUD, AND MISCONDUCT

8.15    Evaluators should anticipate the possibility of discovery of wrong-doing, fraud, or misconduct, and clarify up front to whom such cases should be reported. Most GRPPs will not have a separate ethics, integrity, or investigation office. Clarification needs to be sought therefore, about procedures to follow if such a case is discovered and which of the following should be initially informed: the commissioner of the evaluation, the program manager, the chair of the governing body, a representative of a trustee agency, or a country authority. As an initial step, the nature of the case should be reported (without revealing the evidence, the identity of any individuals wholly or partially responsible, and the details of the case.) Confidentiality must be preserved until the appropriate authority is identified to whom the report should be made.

Elaborates on and applies UNEG Standard 3.16, para. 38, to GRPPs

# EVALUATION CONTENT AND CRITERIA

## 9. Relevance

### Principles and Norms

DEFINITION

9.1 *Relevance* is the extent to which the objectives and design of the program are consistent with (a) current global/regional challenges and concerns in a particular development sector and (b) the needs and priorities of beneficiary countries and groups. Shortcomings in relevance occur when the supply or the demand for the program is not well founded; when the program's activities are competing with or substituting for activities that individual donors, beneficiary countries, or other GRPPs could do more efficiently; or when the program's design and implementation are inappropriate for achieving its objectives.

*Based on DAC Glossary and IEG evaluation criteria*

NEED FOR GRPP EVALUATIONS TO ASSESS RELEVANCE

9.2 All GRPP evaluations should assess the relevance of GRPP objectives and design. The relevance of a GRPP typically arises from the interplay between global/regional challenges on the one hand and beneficiary needs and priorities on the other, since the interests of all partners and participants do not always coincide. Indeed, the divergence of benefits and costs between the global/regional and country levels, or the inability of existing institutional arrangements to reflect shared interests is often a reason for financing the provision of global/regional public goods.[51]

*Draws on IEG's experience with reviewing GRPPs*

9.3 The assessment of relevance includes assessing whether the objectives and the design of the program are still appropriate at the time of the evaluation, given that circumstances may have changed since the program was started or its objectives last revised. The assessment may also include the relevance of the program in relation to specific priorities, sector strategies, operational policies, and guidelines of the program's partners, if this is specified in the TOR.

---

51. It should be recognized that donor countries can also be important beneficiaries of global public goods programs such as the Consultative Group on International Agricultural Research (the outputs of which are also being used in donor countries), the Multilateral Fund for the Implementation of the Montreal Protocol (which has reduced emissions of ozone-depleting substances for the benefit of all), and global health programs that are mitigating the spread of infectious diseases such as HIV/AIDS, tuberculosis, and malaria.

# Standards and Guidelines

## ARTICULATION OF CURRENT OBJECTIVES, STRATEGIES, AND ACTIVITIES

9.4    Building on previous evaluations (where applicable), an evaluation of a GRPP should articulate the current objectives and design of the program as well as changes that have occurred since the inception of the program and during the evaluation period. This would include a description of the objectives, strategies, and major activities of the program — for example, the extent to which the program is engaged in facilitating the operation of a network, in generating and disseminating knowledge, in advocating an approach to development in a sector, or in financing or delivering technical assistance or investments.

## LACK OF CLEARLY ARTICULATED OBJECTIVES OR STRATEGIES

Draws on IEG's experience with reviewing GRPPs

9.5    The evaluation needs to be based on a clear statement of the objectives and strategies of the program. In cases (a) where the objectives and strategies have not been well articulated, (b) where these have changed during the evaluation period, or (c) where their articulation in historical program documents is different from that in the TOR, evaluators will need to construct a clear and agreed-upon statement of the objectives and strategies in consultation with the governing body (or oversight subcommittee or external panel) that is overseeing the evaluation. The evaluators may even propose constructing a logical framework for the program in consultation with the program management.[52]

9.6    If the two parties agree to create a logframe for the purpose of the evaluation, this should be done in such a way that does not compromise the independence of the evaluation. Although logframes are common in project evaluation, placing responsibility for the creation of a logframe on the evaluators themselves is more problematic for GRPPs. Many GRPPs have extensive authorizing environments, and the construction of a logframe should ideally be a participatory exercise among all the partners and participants in order to enhance accountability for results.

## IMPLICIT OBJECTIVES OF THE PROGRAM, IF ANY

9.7    The evaluators should also attempt to ascertain the extent to which the program has objectives that have not been explicitly articu-

---

52. See paragraphs 2.33 and 2.34. All GRPPs should be designed with some form of logical framework, agreed upon by program partners, that includes an articulation of the program's objectives and indicators to measure the achievement of its objectives. This expectation is in line with the development community's commitment to provide development assistance in accordance with a results agenda.

lated, such as influencing the approaches of other donors and organizations operating in the sector. It may be necessary for the evaluators to assess the relevance and the achievement of these objectives as well, in order to capture the full range of outcomes of the program. Particularly where these implicit objectives are well understood and agreed on by the program's partners, it is important to hold the program accountable for their achievement (or lack thereof) and to recommend that the program adopt a more explicit and complete statement of its objectives.

### ASSESSING THE RELEVANCE OF THE OBJECTIVES OF GRPPS

9.8     The relevance of the objectives should be assessed against each of the following four criteria.

9.9     **The existence of an international consensus that global/regional collective action is required.** Such a consensus can be articulated in a variety of ways, such as formal international conventions, less formal international agreements reached at major international meetings and conferences, or formal and informal standards and protocols promoted by international organizations, NGOs, and others. This criterion may be viewed as relevance *from the supply side*. Sponsorship of a GRPP by a number of significant international organizations generally enhances its relevance from the perspective of their membership (donor and beneficiary countries) and from the perspective of the profession (technical experts), but these alone are not sufficient. There needs to be a consensus not only on the need for action but also on the definition of the problem, on priorities, and on strategies for action. What is the authorizing environment for the program? Was the assessment of the global/regional public policy gap that led to the creation of the program correct? For continuing relevance, evidence should be presented that the original consensus that led to the creation of the program is still present, and that the program is still needed to address specific global/regional public concerns. For those programs (such as global and regional environment programs) that are implementing international conventions, to what extent are their objectives and strategies still sufficiently aligned with the objectives of these conventions (which constitute their authorizing environment)? For donor-driven programs, to what extent has there been a plan in place to increase the relevance of the program to beneficiaries over time?

9.10     **Alignment with beneficiary needs, priorities, and strategies.** Relevance to beneficiaries should be assessed against their priorities, strategies, and political and institutional contexts as articulated in the countries' own Poverty Reduction Strategies and donors' participatory strategies (such as World Bank Country Assistance Strategies and UN Development Assistance Frameworks). This may be viewed as

51

relevance *from the demand side*. Where beneficiary countries are signatories to the international conventions or declarations that gave birth to the programs, this enhances relevance. But even donor and supply-driven programs may acquire beneficiary ownership over time by demonstrating positive outcomes and impacts. Obtaining evidence of beneficiary ownership of the program is particularly important if the representation of beneficiaries in the governance or implementation of the program has been deficient in the past or present.

9.11 **Consistency with the subsidiarity principle.** This principle concerns the most appropriate level — global, regional, national, or local — at which particular activities should be carried out in terms of efficiency and responsiveness to the needs of beneficiaries. This may be viewed as relevance *in the vertical sense*. In general, GRPPs are an appropriate level for activities for which the benefits of collective action relative to the transaction costs of operating the global or regional partnership exceed the net benefits arising from individual donors' using their normal instruments. The activities of GRPPs should not be competing with or substituting for activities that individual donors or countries could do more efficiently by themselves. Evaluators should pay particular attention to those programs that, on the face of it, are primarily supporting the provision of national or local public goods. For programs that are providing global or regional public goods that cannot or will not be provided by individual countries or entities acting alone, consistency with the subsidiarity principle is more straightforward (Box 2).

---

**Box 2. What Are Global and Regional Public Goods?**

Public goods produce benefits that are non-rival (many people can consume, use, or enjoy the good at the same time) and non-excludable (it is difficult to prevent people who do not pay for the good from consuming it). If the benefits of a particular public good accrue across all or many countries, then this is deemed a global or international public good.

In their pure form, true global public goods are rare. Therefore, the International Task Force on Global Public Goods, 2006, adopted a practical definition, as follows: "International public goods, global and regional, address issues that: (a) are deemed to be important to the international community, to both developed and developing countries; (b) typically cannot, or will not, be adequately addressed by individual countries or entities acting alone; and, in such cases (c) are best addressed collectively on a multilateral basis."

This definition implies that information and knowledge about development — an output of many global programs — are not necessarily global public goods. There is, for instance, no shortage of knowledge now being disseminated globally on the Internet. Useful knowledge also tends to be contextual, and its global public goods characteristics must be verified through empirical research.

---

9.12    **The absence of alternative sources of supply.** This may be viewed as relevance *in the horizontal sense*. Such an analysis could be done from several perspectives. First, what is the comparative advantage, value added, or core competency of the program relative to other GRPPs with similar or complementary objectives? Is the program providing additional funding, advocacy, or technical capacity that is otherwise unavailable to meet the program's objectives? Is the program providing these things more efficiently than other GRPPs? Second, to what extent are the goods and services being provided or supported by the program in the nature of public goods? Are there alternative and more efficient ways in which these could be delivered? Is the program providing goods and services that could be provided by the private sector under regular market conditions?

## ASSESSING RELEVANCE OF THE DESIGN OF GRPPS

9.13    This concerns the extent to which the strategic approach and the priority activities of the program are appropriate for achieving the objectives of the program. Is the balance between the various types of activities appropriate in light of the program's resources, the needs and priorities of beneficiaries in the sector, the subsidiarity principle, and alternative sources of supply? Is the geographic coverage of the program consistent with the objectives of the program, such as addressing extreme poverty or the particular needs of fragile states? Are the strategies of the program still appropriate for achieving the objectives, given recent developments in the sector, such as the development of new technologies?

Draws on IEG's experience with reviewing GRPPs

9.14    GRPPs support diverse types of activities. While almost all advocate greater attention to — as well as improved donor coordination in relation to — specific issues or specific approaches to development in their sector, they are doing so on different scales:

- Some, generally small, programs are primarily policy or knowledge networks that facilitate communication, advocate policy change, and generate and disseminate knowledge and good practices in a particular area of development.
- Other, somewhat larger, programs also provide country or local-level technical assistance to support national policy and institutional reforms and capacity strengthening, and to catalyze public or private investment in the sector.
- The largest programs also provide investment resources to support the provision of global, regional, or national public goods.

9.15    For each type of activity (networking, advocacy, knowledge creation, technical assistance, or investments), the evaluators should assess the validity of the assumptions underlying the expected rela-

tionship between the activities and the achievement of the objectives. The expected outcomes and impacts may be achieved either through command and control within bureaucracies, through voluntary exchange in markets, through a common interest in collective action, or through some combination of these. The expected outcomes and impacts will also depend on the nature of the goods or services being provided (whether excludable or rival), the motivations and the capacities of the partners and participants, and the rules that govern their interactions. For instance, in cases where the interests of donor and beneficiaries may diverge (such as the preservation of biodiversity of global importance), the assessment of relevance needs to ask whether the program is providing appropriate incentives (such as incremental-cost financing) to overcome these divergent interests.

9.16    Assessing the relevance of the design of the program is greatly facilitated if the program has formally articulated a results chain or logical framework along with qualitative or quantitative indicators. To what extent do the results chain and accompanying indicators capture the distinct contributions of each type of activity to the program's objectives? Does the results chain clearly identify the extent to which the achievement of the objectives depends on the behavior of organizations and individuals — whether public or private, and functioning in bureaucracies, markets, or collectivities?

9.17    For programs that are providing global/regional public goods, an important consideration in designing a program is the manner in which the individual efforts of the partners contribute or add up to the collective outcome for the program as a whole — that is, whether the collective outcome equals the "best shot," "summation," or "weakest link" of the individual efforts.[53] For best shot aggregation technologies (such as an AIDS vaccine), the individual partners should pool their efforts, because the collective outcome equals that of the best individualized effort. For summation technologies (such as mitigating climate change), the collective outcome equals the sum of the individual efforts. Therefore, one partner's contribution (or lack thereof) can substitute for (or nullify) another partner's contribution. For weakest link technologies (such as the eradication of an infectious disease), the smallest provision (or lack thereof) determines the collective outcome. If one necessary partner does not do anything, the disease will not be eradicated.

---

53. For a current treatment of these different aggregation technologies, see Scott Barrett, 2006, "Making International Cooperation Pay: Financing as a Strategic Inventive," in Inge Kaul and Pedro Conceição, eds., *The New Public Finance: Responding to Global Challenges.*

9.18     Under the rubric of the Millennium Development Goals and the Paris Declaration, both donor and recipient countries have declared their commitment to harmonize and align aid delivery. Therefore, the design of the GRPP should not detract from efforts to align donor activities and strengthen beneficiary country capacity for planning, budgeting, and sectoral performance assessment.[54] In addition, the design should not contradict the operational policies and guidelines of the program's partners in relation to special considerations such as environmental management, indigenous peoples, gender equality, etc.

### ADDITIONAL CONSIDERATIONS FOR REGIONAL PROGRAMS

9.19     Regional partnership programs are often sub-regional in scope, with a contiguous geographic dimension to them such as a body of water (like the Aral Sea or Lake Victoria), a river system (like the Nile), or a transport or power system. More than for most global programs, these programs exist for the specific purpose of resolving collective action dilemmas among the participating countries regarding the use of the common resource. Therefore, it is important for evaluators to assess both individual country ownership of the program and the appropriateness of the incentives for cooperation that have been built into the design of the program. Experience has shown that the absence of either can have serious consequences for the effectiveness of the program.

Draws on IEG's forthcoming review of regional programs

9.20     For these regional programs, the assessment of relevance needs to ask to what extent there has been an adequate assessment of the costs and benefits to the countries individually, particularly in programs where countries have to make difficult trade-offs, such as water sharing or usage agreements. Has there been sufficient analysis of the political context and the inter-partner relationships that enable the development of trust, confidence measures, and conflict resolution mechanisms? Has there been an assessment of the capacity of the countries to implement their part of the regional programs? Has the design of the program taken into account how the partnership expects to transfer some or all of its functions to national institutions and structures over time? Is there a plan for sustainability and a clear understanding of the time period and the extent for which external financing will be needed? (See also Chapter 14, Sustainability, Risk, and Strategies for Devolution or Exit.)

---

54. See the Paris Declaration on Aid Effectiveness, March 2, 2005.

# 10. Effectiveness (or Efficacy)

## Principles and Norms

### DEFINITION

10.1    *Effectiveness (or efficacy)* is the extent to which the program has achieved, or is expected to achieve, its objectives, taking into account their relative importance.[55] Shortcomings in the achievement of objectives have to do either with the *number* of objectives that have not been achieved (or are not expected to be achieved) or with the *extent* to which one or more objectives have not been achieved (or are not expected to be achieved). Positive unintended results may also be regarded as additional achievements if convincingly documented.

Based on DAC Glossary and IEG evaluation criteria

### NEED FOR GRPP EVALUATIONS TO ASSESS EFFECTIVENESS

10.2    All GRPP evaluations need to include an assessment of the effectiveness of the program in order to demonstrate to stakeholders (a) the degree to which the original objectives are being met, (b) whether the program should adjust or restate its objectives or strategies to reflect changing circumstances, or (c) whether the program needs to put in place additional safeguards or compensatory measures to mitigate any negative unintended results. Depending on the findings of the assessment, the governing body may wish to consider expanding the program or increasing its reach, changing its geographical coverage, devolving some its activities, or even phasing out some or all activities. An assessment of effectiveness is also important to provide accountability to the international community. Given scarce development aid and many alternative uses for constituent taxes and other resources, the evaluation should compare the achievement of the program's objectives not only to the original expectations but also, to the extent possible, to the outcomes from alternative uses of resources.

Draws on IEG's experience with reviewing GRPPs

---

55. As noted in the OECD/DAC *Glossary of Key Terms in Evaluation and Results Based Management*, 2002, "effectiveness" is also used as a broader, aggregate measure — encompassing relevance and efficiency as well — of the overall outcome of a development intervention such as a GRPP. This chapter uses the term "effectiveness" in the narrow sense, which is synonymous with the use of the term "efficacy" in a number of development organizations such as the World Bank.

## Standards and Guidelines

### OBJECTIVES-BASED ASSESSMENT

10.3    The evaluation should first assess the achievement of the stated objectives — objective by objective — and the extent to which each objective has been achieved (or is expected to be achieved). The evaluators should also determine if the program has unstated objectives, since the objectives often differ from the perspective of different partners and other stakeholders, and since objectives are dynamic and change over time. (See also paragraph 9.7 under relevance.)

10.4    The ability to undertake a systematic assessment of the achievement of each objective will depend on the maturity of the program and the existence of a good monitoring framework, including a structured set of qualitative or quantitative input, output, outcome, and impact indicators. When the program is young (less than four years old), it will be more difficult to make a summative assessment of effectiveness. (See also paragraph 6.7.) When the program has not established a good monitoring framework, the evaluators could provide guidance to the secretariat in establishing one. (See also paragraphs 2.22–2.34 on establishing an M&E framework for GRPPs.)

### UNINTENDED OUTCOMES

10.5    The assessment of effectiveness should not be limited to the achievement of expected outputs and outcomes, but should also cover unintended outcomes, whether negative or positive. These would include the unintended results of the program's activities as well as of the partnership itself, such as any harmonization of procedures or effects on aid coordination outside of the partnership itself.

10.6    The assessment of effectiveness should also include how the objectives and strategies of the program have evolved in response to (a) learning from experience or (b) the risks and opportunities arising from a new external environment, technology, or emerging target group. For instance, it may become important for the program to provide compensatory measures if unintended negative outcomes are occurring in relation to the program's safeguard objectives.

### EVIDENCE-BASED CONCLUSIONS

10.7    The assessment of the achievement of objectives, and of other unintended results, should be evidence-based. Evidence-based conclusions distinguish an evaluation report from an expert consultant report, which is based primarily on expert judgments. Evidence-based conclusions and internal consistency among findings based on more than one type of evidence — or triangulation — have the added benefit of helping to ensure independence, regardless of organizational ar-

rangements. The ability to provide evidence-based conclusions depends on the use of measurable indicators, as laid out below.

### THE NEED TO MEASURE INPUTS, THE PROGRESS OF ACTIVITIES, OUTPUTS, OUTCOMES, AND IMPACTS TO THE EXTENT POSSIBLE

10.8 An evaluation should measure inputs, the progress of activities, outputs, outcomes and impacts to the extent possible (or an appropriate rationale should be given as to why not). Findings regarding inputs should be distinguished clearly from those regarding outputs, outcomes, and impacts. Outcomes and impacts should include any multiplier or downstream effects attributable to the GRPP and — as noted above — any unintended effects, whether positive or negative. To the extent possible, each of these should be measured either quantitatively or qualitatively and compared to benchmarks.

10.9 In addition to quantitatively measurable inputs, such as budgets and staffing, the assessment should also consider other causal factors that have an effect on the progress of activities, outputs, and outcomes, such as changes in the location, the legal structure, or the governance processes of the program during the time period of the evaluation.

10.10 For GRPPs, it is also important to measure the program's inputs, progress of activities, outputs, outcomes, and impacts at all levels — global, regional, national, and local — and to find a way to present in summary form the results from the local and national levels and the way in which they affect results at the regional and global levels. A simple aggregation of results may not be ideal if this obscures causal relationships. It is better if the results are presented in a way that highlights the factors that have influenced success or failure in a variety of conditions.[56]

10.11 In addition, outcomes related to the unique contribution of the partnership itself — such as the scale or joint activities made possible by its organizational setup as a GRPP, or its institutional linkages to a host organization — should be measured and assessed. What is the value added of the GRPP relative to what could have been achieved by intervening only at the country or local level, taking into consideration the leadership of the partnership, the roles and responsibilities of the various partners, and the degree of trust developed among the partners?

---

56. Sometimes ratings are used to facilitate aggregation of activity results to the country, regional, or global level. If ratings are used for such purposes, it is essential to distinguish ratings of performance (such as effort and inputs) from ratings of results.

10.12    The M&E system of a GRPP needs to be able to take into ac-
count the evolving nature of its portfolio. It is important that organi-
zations or individuals proposing activities for financing at the country
or local level not only list expected outcomes in their proposals but
also link them to measurable indicators, so that GRPP management
can take steps to incorporate these indicators into the program's M&E
system to facilitate the later assessment of the effectiveness of the in-
terventions.

SPECIAL CONSIDERATIONS IN ASSESSING THE EFFECTIVENESS OF GRPPS

Draws on IEG's
experience with
reviewing
GRPPs

10.13    To assess effectiveness, an evaluation of a GRPP must first
attempt to define the ***boundaries of the program's impact***, which may
be difficult, particularly if these are expected to change over time as
the program grows in scale or reach, or if these vary by activity. De-
fining the boundaries of the potential impact of environmental pro-
gram activities can be particularly relevant, but the same is also true
for programs providing social services, which may have the potential
to serve a large population. For example, an assessment of the effec-
tiveness of pilot health service interventions may need to consider
both the outcome of the actual pilot, with its limited scale, and also
the degree to which it yields useful information on the likelihood of
its success under alternative conditions or at larger scale.

10.14    For many mature programs, the ***large scale*** of the program it-
self presents complications in assessing effectiveness. Choice of a rep-
resentative sample of activities becomes very important. The diversity
of country conditions that need to be captured may be larger, and
finding an appropriate modality for presenting diverse results may be
a challenge — going beyond mere aggregation and capturing the dif-
ferent factors affecting success and failure. The use of ratings, which
may facilitate aggregation, may not be appropriate if the basis for the
ratings is not articulated or understood, or if the raters are diverse.

10.15    That GRPPs typically support activities ***at different levels*** may
create complexities if the objectives of stakeholders at different levels
are different, or even in conflict. For instance, the global/regional
public goods benefits of some environmental actions may be associ-
ated with disproportionate costs relative to benefits for some imple-
menting countries. Thus, it is important to indicate from whose per-
spectives the results are being assessed and to assess trade-offs of
costs and benefits to the various stakeholders.

10.16    Unlike projects, GRPPs are programmatic and typically have
no fixed end-point. Many of the expected results have a longer ***time
frame*** than that of the interventions that contribute to achieving the
results. One needs to consider not only the joint outcomes of
global/country/local interventions, but also the "joint outcomes" or

cumulative effects of different interventions over time. To properly assess this, the program needs to have established a baseline and to have put in place arrangements to gather information at specific times in order to assess long-term results.

10.17    GRPPs have *more stakeholders* and more diversity among stakeholders than country and local-level programs and projects. Hence, there are more perspectives on the achievement of results and objectives that need to be taken into account. Again, it is very important for evaluators to obtain a representative sample of views and survey responses. (Evaluators should schedule interviews not just according to convenience or availability.) And they should always disclose the criteria that they used for selecting interviewees or survey respondents.

10.18    Finally, GRPPs differ from other programs and projects because they have *distinct governance mechanisms and processes* that affect results. It is important to regard the way in which these mechanisms and processes work in practice, as well as any changes in them over time, as a part of the results chain. For instance, any of the following can affect the achievement of results and objectives:

- Interruptions in the continuity of key management positions or of members of the governing body
- Changes in the frequency of governance meetings or in the types of decisions handled by the governing body, as opposed to management
- The processes for allocating resources and choosing activities to support
- Changes in the resource mobilization strategy that affect the scale of the program and, if there is earmarking, the allocation and use of funds
- The influence of host organization representatives or the need to comply with their policies.

ASSESSING EFFECTIVENESS OF DIFFERENT TYPES OF PROGRAMS

10.19    GRPPs support diverse types of activities. (See also paragraph 9.14.) Each type of activity (networking, advocacy, knowledge creation, technical assistance, or investments) presents methodological challenges with respect to the assessment of effectiveness. This is because the different types of activities contribute in different ways to the program's value added and leverage on domestic policy and institutional reform, human resource capacity, and total investments in the sector — as well as to other objectives such as poverty reduction and improvements in welfare.

10.20    For each type of activity, the program should have articulated a results chain or logframe with clear, agreed-upon indicators that allow evaluators to attribute results to the program.[57] (See also paragraphs 2.33, 2.34, 9.5, and 9.6.) The evaluators should aim to capture the distinct contributions of each type of activity toward the achievement of the program's objectives so that the objectives can also be adjusted to increase the program's impacts over the long term. Evaluators also need to understand the way in which the objectives are being achieved — whether through command and control within bureaucracies, through voluntary exchange in markets, through common interest in collective action, or through some combination of these. (See also paragraphs 9.15–9.18.) Unintended outcomes on markets or prices, such as "crowding out a market" or "catalyzing a market," should be noted.

<div style="float:left; width:20%">Draws on IEG's forthcoming review of regional programs</div>

10.21    For regional partnership programs, it is also important to assess the distribution of the benefits and costs of the program among the beneficiary partners. Experience has shown that an inequitable distribution of net benefits can adversely affect the sustainability of the program.

10.22    Where feasible, evaluations should assess final welfare outcomes in relation to a counterfactual in order to isolate the effects of the program on those outcomes. This would include assessing how the outputs of the program have supported enhanced welfare outcomes in the sector and country in which the GRPP is operating. (See also Chapter 15, Impact Evaluation.)

ASSESSING LINKAGES BETWEEN GRPPS AND COUNTRY OR LOCAL-LEVEL ACTIVITIES

<div style="float:left; width:20%">Draws on IEG's experience with reviewing GRPPs</div>

10.23    For GRPPs, it is important to assess the effectiveness of their operational linkages with country or local-level activities, whether or not the latter are supported by donors. For most GRPPs, positive outcomes and impacts at the country or local level are a joint product of both the GRPP and country or local-level activities.

10.24    Two types of linkages need to be assessed: (a) opportunities for direct linkages that are subject to the control of participants at both levels and (b) effects that may operate through markets or the behavior of agents external to the program, which may require a strengthening of safeguard measures or other compensation. The linkages in

---

57. However, much work still needs to be done to develop generic indicators for generic-type activities that are common to GRPPs such as advocacy, improving donor coordination, knowledge generation and dissemination, supporting national-level policy and institutional reforms, and capacity strengthening.

(a) and (b) may result in unambiguous win/win outcomes, or may bring to light trade-offs that need to be taken into account in considering the net benefits to different partners and participants.

10.25   With regard to the first set of linkages — which can be improved through conscious action — linkages in both directions are important. First, country and beneficiary representatives (whether public or private) need to have an effective means of communicating their constraints, requirements, and priorities to GRPP management, thereby potentially increasing the relevance, focus, ownership, and outcomes of the GRPP. The means of communication may be through direct participation on the governing body (important, but not always easy in practice),[58] through other periodic consultation mechanisms like workshops, through GRPP procedures that solicit proposals for assistance (and provide help in shaping them to be successful), and through deliberate exchange and discussion of government planning documents and donor assistance strategies relevant to the country or local beneficiaries. An evaluation of a GRPP should always assess the effectiveness of these various means of communications.

10.26   In addition, it is important to assess the actual outcomes and impacts of the GRPP activities on country or local-level priorities, activities, and deployment of human resources. To the extent possible, the benefits of participating in the GRPP — from the perspective of the beneficiary groups — should be compared with the costs, including increased reporting and compliance requirements and other demands on senior skilled implementers (whether public or private). Such an assessment of the opportunity costs of the participants' time and resources should ideally include participatory methods to directly obtain information on beneficiary group satisfaction, complaints, and suggestions for change.

10.27   An assessment should also be made of how well the GRPP acts on the information it obtains from beneficiary groups — both up-front information on needs and priorities that might influence strategy or allocation of funds, and periodic feedback that would provide opportunities for improving the outcomes and impacts of the program at the country or local level. GRPP management needs to fashion the program's support to add value to country or local-level activities by contributing new knowledge and technologies, facilitating exchange of good practice among beneficiary groups, and helping to mobilize additional resources or to channel existing resources to more productive activities. They should seek to ensure alignment of the country or local activities they support with country or local-level

---

58. See Chapter 12, Governance and Management, for treatment of the issue of including beneficiary groups in governing bodies.

plans and budgeting priorities and to ensure that complementary country or local-level inputs are available to make GRPP interventions effective. Regional partnership programs — which are often focused on specific cross-border issues and whose success is typically more dependent on country or local-level commitment and capacity — may require special attention to ensure that program priorities and requirements do not hamper the achievement of the rest of the countries' development agendas.

10.28    GRPP management needs to examine periodically the degree to which the GRPP activities and outputs (which may be inputs to country or local-level activities) are relevant to the needs of final beneficiary target groups. Linkages between the GRPP and donor representatives in decentralized country or local-level units are useful but not sufficient; direct dialogue with country or local implementers and beneficiary groups is also needed. At a minimum, this needs to be done through wide dissemination of monitoring reports, annual reports, and evaluations of the GRPPs to all existing and potential beneficiary groups or local implementers. Even better would be active dialogue to solicit the views of beneficiary groups on the responsiveness of the GRPP activities and outputs to their needs.

10.29    Unintended outcomes that operate through effects on trade, commercial markets, or the behavior of agents external to the planned program results may also need to be assessed. Ideally, the potential results would have been identified in the results chain or logframe in the planning stages of the program, thereby facilitating monitoring of such results. Alternatively, participatory methods of evaluation can identify cases where results are perceived to be due to the GRPP, and methods could be devised in the evaluation to test these hypotheses and recommend compensatory adjustments or additional safeguard measures, if needed.

# 11. Efficiency or Cost-Effectiveness

## Principles and Norms

### DEFINITIONS

11.1    *Efficiency* is the extent to which the program has converted or is expected to convert its resources/inputs (such as funds, expertise, time, etc.) economically into results in order to achieve the maximum possible outputs, outcomes, and impacts with the minimum possible inputs. (See also paragraph 13.1.)

Based on DAC Glossary and IEG evaluation criteria

11.2    *Cost-effectiveness* is the extent to which the program has achieved or is expected to achieve its results at a lower cost compared with alternatives.[59] Shortcomings in cost-effectiveness occur when the program is not the least-cost alternative or approach to achieving the same or similar outputs and outcomes.

11.3    An assessment of efficiency relates the results of a program to its costs. Ideally, this would attempt to put a monetary value on the benefits arising from the activities of the program, compare these with the costs of the program, and calculate the internal rate of return that equalizes the present value of the benefits and costs. But in most cases, a monetary quantification of the program's outputs and outcomes is problematic and would be based on potentially controversial assumptions. In these cases, the assessment of efficiency focuses on ratios such as the number of lives saved, the number of children vaccinated, or the number of additional households served with electricity per thousand dollars invested, while also indicating the margins of error in these estimates.

11.4    An assessment of cost-effectiveness takes the benefits arising from the activities of the program as a given and asks whether these could have been produced at a lower cost compared with alternatives. For GRPPs that are providing development assistance to developing countries,[60] the principal alternatives are the traditional means of delivering development assistance (bilateral or multilateral), or other GRPPs operating in the same sector.[61] Ideally, such a comparison of

---

59. *Value-for-money* is a related concept. This assesses the extent to which the program has obtained the maximum benefit from the outputs and outcomes it has produced within the resources available to it.

60. Most GRPPs fall into this category. However, some GRPPs, such as the Prototype Carbon Fund, use trade rather than aid to achieve results.

61. Some may argue that alternative ways of achieving outputs or outcomes without development assistance should also be considered. These might include, for example, community development approaches relying on benefi-

alternatives should assess the costs from both the beneficiary and do-nor perspectives. If this is not possible, the assessment should always state clearly from which perspective the costs are being assessed. (See standards below.)

Elaborates on
UNEG Standard
3.8, paras. 17
and 18

11.5    Development aid is a scarce resource. Therefore, GRPP evalua-tions need to assess the efficiency of the interventions to the extent feasible and to make recommendations for improving the efficient use of resources. Where no efficiency or cost-effectiveness analysis is in-cluded in an evaluation, some rationale for this exclusion should be presented in the objectives or methodology section of the TOR and in the evaluation report. In all cases, evaluators should point out areas of obviously inefficient use of resources.

11.6    It may be difficult, both logically and empirically, to conduct an efficiency or cost-benefit analysis for a GRPP as a whole. However, it is often possible to conduct an analysis for individual activities, which may be compared to sectoral benchmarks and generic cost in-dicators, where available. It may also be possible to compare the costs of delivering similar activities of different GRPPs that are operating in the same sector. In a mature program, an impact evaluation of subsets of activities may also be possible and beneficial. But impact evalua-tions are generally conducted as a separate exercise parallel to and not part of program-level evaluations. (See Chapter 15, Impact Evalua-tion.) If the commissioners of an evaluation choose to include an im-pact evaluation as part of a program-level evaluation, this will require a larger budget, as well as specific impact evaluation skills on the evaluation team.

## Standards and Guidelines

Elaborates on
UNEG Standard
3.8, para. 14

11.7    If assessing efficiency or cost-effectiveness is among the ob-jectives of the evaluation, a range of analytical approaches may be considered, from an elaborate cost-benefit or internal rate of return analysis, to a more limited cost-effectiveness analysis, or to a quick cost comparison. At a minimum, the evaluation should measure and analyze the program's costs in broad categories and categorize and list the program's activities, outputs, outcomes, and other benefits,

---

ciary contributions of labor and other resources to specific activities. While these alternatives may be superior with respect to sustainability, they are unlikely to be able to deliver results at the same scale as the GRPP, which is supported by external development assistance.

even if these cannot be valued in monetary terms. Evaluators should, to the extent possible, address the following broad questions:

- Has the program cost more or less than planned? How did it measure up against its own costing schedule?
- How do actual costs compare with benchmarks from similar programs or activities? Are there obvious cases of inefficiency or wasted resources?
- Do the program benefits outweigh the costs of individual activities? (For regional partnership programs, do the national program benefits outweigh the costs for each country?)
- What is the least-cost way of getting the expected results?
- Were the program's outputs and outcomes achieved in the most cost-effective way?

11.8    Additional relevant questions, based on the scope of the evaluation and the technical and financial resources available to the evaluation team, would include:

*Based on UNEG Standard 3.8, para. 14*

- What would be the implications of scaling the program up or down in terms of costs, cost-effectiveness, or efficiency?
- What would be the costs of replicating the program's activities in a different environment?
- How do costs affect the results and the sustainability of the program?

### FINANCIAL VERSUS OTHER ECONOMIC AND SOCIAL COSTS

11.9    Efficiency and cost-effectiveness analysis in evaluation builds on financial information, but may also involve calculating other economic costs such as labor-in-kind and opportunity costs. Analysis of incremental costs may also be useful. Whether cost comparisons are made only in relation to activities and outputs or also in relation to outcomes and impacts will depend on the purpose of the evaluation and the evaluation questions posed (and also on the maturity of the program). Efficiency and cost-effectiveness analysis should explicitly specify the perspective from which costs are analyzed (such as the perspective of the whole program, selected donors, a country or local-level implementing agency, or individual beneficiaries).

*Based on UNEG Standard 3.8, paras. 15 and 16*

### NEED TO EXPLAIN LIMITATIONS OF ANALYSIS

11.10    The analysis of efficiency and cost-effectiveness should explain any limitations of the analysis, which may include various complexities faced (such as multiple program objectives), poor data, or the limitations on the time and resources experienced by the evaluators.

*Based on UNEG Standard 3.8, para. 16*

Draws on
internal IEG
guidelines

11.11    Where only qualitative assessments are possible, the evaluator should take into account the following factors:

- Implementation progress (delays and redesigning would increase costs)
- Whether the stream of benefits has reached significant levels and is growing at reasonable rates (compared with plans)
- Capacity utilization rates for facilities and services financed
- Adequate operation and maintenance arrangements and financing
- Good-practice standards for services
- Whether the benefits stream is judged to be adequate when compared with the costs.

CONSTRAINTS TO ASSESSING EFFICIENCY OR COST-EFFECTIVENESS

Draws on IEG's
experience with
reviewing
GRPPs

11.12    An IEG review of external evaluations of GRPPs has revealed few cases where efficiency or cost-effectiveness has received the emphasis considered important by the above UNEG and IEG standards. This may be due to the following factors, which governing bodies or commissioners of evaluations should bear in mind when deciding the scope of an evaluation:

- Donor agendas may not consider efficiency or cost-effectiveness of grant aid to be as important as the achievement of objectives (effectiveness) — that is, showing results to constituencies.

- Expectations with regard to efficiency are low in the early years of a GRPP, since the costs of establishing the program and its governance and management arrangements are high relative to activity costs.

- There is inherent complexity. Estimating the value of benefits is always difficult and depends on the perspective adopted (donor, implementer, or beneficiary group). Cost categories are not uniform among programs. Special skills are required. Neither the manager of the evaluation nor the lead evaluator may have the special skills, time, or resources to provide sufficient guidance during the evaluation when problems of measurement are encountered.

- The continuing evolution of a GRPP, with the scale and reach being dependent on the availability of financing, means that the changing economies of scale make the use of benchmarks (or comparison with other programs) difficult.

- The multiplicity of partners and activities makes it particularly difficult to assess results against a counterfactual.

## COST CATEGORIES TO BE CONSIDERED

11.13   At a minimum, GRPP evaluations should record administrative costs relative to activity costs — paying attention to trends over time, taking account of the intended versus the actual breadth and scope of a program's activities, and noting any actual or expected economies of scale.

11.14   Ideally, other cost categories should also be considered, such as the following:

- The transaction costs of convening the partners, such as the travel and subsistence costs for attending meetings of the governing body, not all of which are recorded in the program's expenditure records

- Upfront expenditures spent during preparation, preliminary resource mobilization, and planning of the institutional framework and governance, even though these costs may have been incurred by different founding organizations and donors, and may not have been recorded in the program's expenditure records [62]

- Additional costs incurred by third parties as a consequence of the program's activities, for example, by national authorities during implementation.

## EFFICIENCY AND COST-EFFECTIVENESS FROM THE BENEFICIARY GROUP PERSPECTIVE

11.15   GRPP evaluations should analyze efficiency and cost-effectiveness of the GRPP from the perspective of the beneficiary groups. Is the menu of program benefits responsive to the needs of the beneficiary group? Is the group getting fair access to the benefits? Are there benefits of aid harmonization or improved aid coordination associated with the GRPP? Given the benefits received, are the costs of participating (such as preparing proposals, reporting, time spent in GRPP meetings) worthwhile? Does receiving the development assistance through the GRPP increase the transaction costs for the beneficiary groups over what would be the case (or what is the case) from development assistance delivered through traditional bilateral or multilateral programs? In what ways could transaction costs to the beneficiary groups be reduced further?

Draws on IEG's experience with reviewing GRPPs

---

62. Global programs often incur heavy up-front expenditures to convene an effective working partnership and program platform.

Draws on IEG's experience with reviewing GRPPs

11.16    GRPP evaluations should also analyze efficiency and cost-effectiveness from the perspective of donors and partners. Is the GRPP delivering the expected outputs and outcomes in a timely manner? Is reporting adequate to satisfy donors' and partners' need for visibility and accountability to stakeholders? How do the benefits and costs of delivering the development assistance through the GRPP compare with those of traditional development assistance in which donors and partners take part? Has there been a reduction of over-lapping work among donor agencies and partners (such as through joint supervision, monitoring, or evaluation)? Similarly, to what de-gree is the GRPP contributing to increased process harmonization of efforts between donors within the country? Is this having any effect on donor costs? While it may be difficult to actually measure benefit streams, and while benefits and costs for different donors and part-ners will differ, surveys of donors and partners can help record and quantify perceived benefits and provide the basis for such an aggre-gate assessment.

## COMPARING ALTERNATIVES

Draws on IEG's experience with reviewing GRPPs

11.17    GRPP evaluations should also compare the progress of activi-ties, outputs, and outcomes with alternative ways of delivering the same activities or achieving similar results more cost-effectively, in-cluding through another similar program or through a lower-cost means. For example, if the underlying intent of the program is to forge greater understanding or linkages between two international organizations in a specific thematic area, could this have been achieved more cost-effectively through improved knowledge man-agement and dissemination, or a staff-level working group?

11.18    Commissioners of evaluations may wish to propose specific areas where a detailed analysis of cost-effective alternatives would be beneficial. One such area might be examining alternative ways of providing opportunities for stakeholders to participate in program governance (either in general or in key strategic decisions). (See also Chapter 12, Governance and Management.) While legitimate partici-pation of beneficiary groups from developing countries is a key norm for GRPPs, there may be a variety of ways this can be achieved. Evaluators could examine the trade-off between representation on the governing body and alternative means of participation, including the use of new technologies such as videoconferencing or moderated e-discussions. Another such area might be how best to deliver a sub-set of activities or services. For instance, evaluators could conduct a comparative analysis of the efficiency of centralized programs versus partially or fully decentralized programs.

# 12.  Governance and Management

## Principles and Norms

### DEFINITIONS

12.1    *Governance* concerns the structures, functions, processes, and organizational traditions that have been put in place within the context of a program's authorizing environment "to ensure that the [program] is run in such a way that it achieves its objectives in an effective and transparent manner."[63] It is the "framework of accountability to users, stakeholders and the wider community, within which organizations take decisions, and lead and control their functions, to achieve their objectives."[64] Good governance adds value by improving the performance of the program through more efficient management, more strategic and equitable resource allocation and service provision, and other such efficiency improvements that lend themselves to improved development outcomes and impacts. It also ensures the ethical and effective implementation of its core functions.

Elaborates on IEG's Phase 2 report of the World Bank's involvement in global programs

12.2    *Management* concerns the day-to-day operation of the program within the context of the strategies, policies, processes, and procedures that have been established by the governing body. Whereas governance is concerned with "doing the right thing," management is concerned with "doing things right."[65]

12.3    The boundary between governance and management is not hard and fast. In particular, both the maturity and the size of the program will influence the dividing line and the degree of separation between the program's governance and management structures. Less-mature programs may take time to establish formal governance mechanisms. Smaller programs with limited staffing and financial resources may tend to blend responsibilities between those who govern and those who manage, and to call on governing body members to be more involved in specific day-to-day management decisions. The extent of governance should be proportionate to the size of the program in order not to result in an over-governed and under-performing program.

---

63. Institute of Chartered Secretaries and Administrators International, no date, *Principles of Corporate Governance for Charities*, p. 2.

64. United Kingdom Audit Commission, October 2003, *Corporate Governance: Improvement and Trust in Local Public Services*, p. 4.

65. This distinction is attributed to Robert Tricker: "The role of management is to run the enterprise and that of the board is to see that it is being run well and in the right direction." Robert I. Tricker, 1998, *Pocket Director*, p. 8.

Elaborates on
and applies
OECD *Principles
of Corporate
Governance* to
GRPPs

12.4    The governing bodies of GRPPs typically exercise six core
functions:[66]

- **Strategic direction.** Exercising effective leadership that opti-
  mizes the use of the financial, human, social, and technological
  resources of the program. Establishing a vision or a mission
  for the program, reviewing and approving strategic docu-
  ments, and establishing operational policies and guidelines.
  Continually monitoring the effectiveness of the program's
  governance arrangements and making changes as needed.

- **Management oversight.** Monitoring managerial performance
  and program implementation, appointing key personnel, ap-
  proving annual budgets and business plans, and overseeing
  major capital expenditures. Promoting high performance and
  efficient processes by establishing an appropriate balance be-
  tween control by the governing body and entrepreneurship by
  the management unit. Monitoring compliance with all appli-
  cable laws and regulations, and with the regulations and pro-
  cedures of the host organization, as the case may be.[67]

- **Stakeholder participation.** Establishing policies for inclusion
  of stakeholders in programmatic activities. Ensuring adequate
  consultation, communication, transparency, and disclosure in
  relation to program stakeholders that are not represented on
  the governing bodies of the program.

- **Risk management.** Establishing a policy for managing risks
  and monitoring the implementation of the policy. Ensuring

---

66. These core functions, and the criteria for assessing the performance of
governing bodies in the standards section below, are adapted from the
OECD *Principles of Corporate Governance* (2004). Although there exist other
similar statements of such principles at the national level, the OECD *Princi-
ples* are the only set of corporate governance principles on which there is
clear international consensus. Many governance functions for the for-profit
private sector, as laid out in the OECD Principles, translate directly into
equivalent functions for GRPPs (as well as for other public sector organiza-
tions, NGOs, and foundations). The key differences for GRPPs are the ab-
sence of tradable shares, the need to establish legitimacy on a basis other
than shareholder rights, and the greater need for transparency in the use of
public sector resources in achieving public policy goals.

67. In this Sourcebook, the terms "oversight" and "supervision" are used for
two distinctly different activities. *Oversight* refers to the monitoring of the
program management unit by the governing body, while *supervision* refers
to the monitoring of individual program activities by the staff (or in some
cases contractors) of the program management unit.

that the volume of financial resources is commensurate with the program's needs and that the sources of finance are adequately diversified to mitigate financial shocks.

- **Conflict management.** Monitoring and managing the potential conflicts of interest of members of the governing body and staff of the management unit. Monitoring and managing conflicting interests among program partners and participants, especially those that arise during the process of program implementation.[68]

- **Audit and evaluation.** Ensuring the integrity of the program's accounting and financial reporting systems, including independent audits. Setting evaluation policy, commissioning evaluations in a timely way, and overseeing management uptake and implementation of accepted recommendations. Ensuring that evaluations lead to learning and programmatic enhancement.

12.5    In the case of programs that are housed in other organizations, the host organization may be responsible for performing some of these functions in collaboration or consultation with the governing body.

FUNCTIONS OF MANAGEMENT

12.6    Management functions vary by program size and type, partnership arrangement, legal arrangement, etc. While the proceeding list is not exhaustive, seven general functions of GRPP management are as follows:

- **Program implementation.** Managing financial and human resources. Reviewing proposals for inclusion in the portfolio of activities and allocating financial resources among activities. Supervising the implementation of activities. Contracting with implementing or executing agencies to implement individual activities. Ensuring that these agencies are self-monitoring and reporting their progress in a timely way.

- **Regulatory compliance.** Ensuring compliance with all applicable laws and regulations at the international, national, and institutional levels, including the regulations and procedures of the host organization, as the case may be. Being aware of

---

68. This is particularly important for regional partnership programs that are explicitly involved in mitigating conflicts among countries in relation to trade or resource use.

and adhering to these requirements and standards on a day-to-day basis.

- **Reviewing and reporting.** Taking stock of the overall performance of the portfolio in relation to the program's objectives and strategies. Reporting progress to the governing body, including any adverse effects of the program's activities. Serving the needs of the governing body by preparing strategies, policy statements, etc.

- **Administrative efficiency.** Maintaining a lean administrative cost structure (while recognizing that administrative costs tend to be higher during the launch period of a GRPP). Proposing ways to maintain high performance while reducing costs to increase operational effectiveness.

- **Stakeholder communication.** Implementing board-approved policies for stakeholder inclusion in programmatic activities. Finding ways to increase the effectiveness of stakeholder participation in all aspects of the program.

- **Learning.** Distilling and discerning lessons from the implementation of activities across the portfolio. Transmitting these lessons to both governing partners and beneficiaries in order to inform policy making and to enhance implementation of activities.

- **Performance assessment.** Reviewing the performance of operational staff on a regular basis, as well as the performance of consultants at the end of their assignments.

## NEED FOR GRPP EVALUATIONS TO ASSESS GOVERNANCE AND MANAGEMENT

Draws on IEG's experience with reviewing GRPPs

12.7     All GRPP evaluations should include an assessment of the legitimacy and effectiveness of the governance of the program, because the formal programmatic partnership represented by these governance structures is the raison d'être of a GRPP. The partners have established the partnership in order to achieve something collectively that the individual partners could not achieve at all, or as efficiently, by acting alone.

12.8     It is neither practical nor appropriate for evaluations to assess all aspects of management. Therefore, the TOR should clearly specify which aspects of management have been selected for assessment. The assessment should focus on those aspects that most directly affect program performance, and avoid the type of "micro-management" or "micro-evaluation" that is outside the purview of both a program's

governing body and an evaluation team.[69] For instance, an evaluation could undertake a broad assessment of the adequacy of the management of financial and human resources in light of the objectives of the program.[70] It could also review the effectiveness of various key processes such as preparing strategies, allocating financial resources, and reviewing proposals for inclusion in the portfolio. (See also Chapter 13, Resource Mobilization and Financial Management.)

12.9     The evaluation could also assess aspects of the performance of the host organization and/or the program's partners, where the host organization is performing some governance or management functions on behalf of the program, and where the partners have made specific commitments to the program (such as pledges to provide funding). However, the inclusion of these in the TOR should be cleared with the host organization and the partners, respectively.

12.10    The importance of assessing the governance of the program, as well as some aspects of management, implies the need for a governance expert on the evaluation team. It is also important that evaluators have access to the minutes (at least in summary form) of the governing, executive, and advisory bodies as the case may be, and be allowed to attend the meetings of such bodies as an observer.

## Standards and Guidelines

### SUGGESTED CRITERIA FOR ASSESSING GOVERNANCE AND MANAGEMENT

12.11    GRPPs employ a diverse array of governance models associated with the history and culture of each program. Therefore, it is not practical to base the assessment of governance and management on a particular governance model. Rather, it is suggested that the assessment should be based on compliance with seven generally accepted principles of good governance: legitimacy, accountability, responsibility, fairness, transparency, efficiency, and probity.

Elaborates on and applies the OECD *Principles of Corporate Governance* to GRPPs

---

69. The assessment of management needs to avoid the assessment of the individual performance of managers. UNEG Norm 11, para. 11.5, states that "evaluators are not expected to evaluate the personal performance of individuals and must balance an evaluation of management functions with due consideration for this principle." In addition, UNEG Standard 3.16, para. 38, states that "evaluations should not substitute, or be used for, decision making in individual human resources matters."

70. If a more detailed assessment of human resource management is included in the TOR, the following questions are particularly relevant to GRPPs: How are international and domestic staff salaried, and is this an efficient structure? If GRPP employees are staff of an international organization, do the associated benefits justify the costs? Is the GRPP pulling essential country counterparts away from domestic priority tasks in the concerned sector?

12.12    The assessment of governance and management should also build upon and add to the previous assessments of relevance, effectiveness, and efficiency. For instance, legitimacy is closely related to the relevance of the program, and efficient governance is related to the efficiency or cost-effectiveness of the program. Responsibility and fairness are closely related to participation and inclusion (discussed in Chapter 4), and transparent governance is related to transparency and disclosure (discussed in Chapter 5). The focus in the present chapter, however, is on the structures and processes of governance and management. To what extent are these well articulated and working well to bring about legitimate and effective governance and management of the program?

12.13    **Legitimacy.** This refers to the way in which governmental and managerial authority is exercised in relation to those with a legitimate interest in the program — including shareholders, other stakeholders, implementers, beneficiaries, and the community at large.[71] This is closely related to the relevance of the program (discussed in Chapter 9). The concern here is the extent to which the governance and management structures permit and facilitate the effective participation and voice of the different categories of stakeholders in the major governance and management decisions, taking into account their respective roles and relative importance. Because GRPPs are international public sector programs with a "duty of care" to identify and respond to the needs and demands of developing countries, and because most are involved in channeling development assistance to developing countries, it is particularly important that the voices of developing countries and technical experts can be effectively expressed and heard. For instance, to what extent is the most up-to-date scientific and technical advice being sought to inform policy making and operational effectiveness?

---

71. As discussed in the overview to this Sourcebook, the term *donor* is used in the generic sense as referring to those who make financial or in-kind contributions to the program that are reflected in the audited financial statements of the program. Therefore, the term includes not only "official donors" but also developing countries that contribute annual membership dues, seconded staff, or office space, provided that these are formally recognized, as they should be, in the financial statements of the program. Donors can also be beneficiaries. But the term *donor* does not extend to beneficiary countries or groups that are providing counterpart contributions that are not formally recognized in the financial statements of the program. *Shareholders* are here defined as the subset of donors that are involved in the governance of the program. Therefore, shareholders do not include individual (particularly anonymous) donors who choose not to be so involved, or who are not entitled to be involved if their contribution does not meet the minimum requirement, say, for membership on the governing body.

12.14 **Accountability**. This concerns the extent to which accountability is defined, accepted, and exercised along the chain of command and control, starting with the annual general meeting of the members or parties at the top and going down to the executive board, the chief executive officer, task team leaders, implementers, and in some cases, to the beneficiaries of the program. For instance, to what extent is the assignment and exercise of responsibilities between governance and management appropriate relative to good practice? There may also be mutual accountability at various steps in the reporting chain. Accountability is enhanced when the roles and responsibilities are clearly articulated in a program charter, memorandum of understanding, or partnership agreement, and when these agreements work out such issues as to whom and for what purposes the members of the governing body are accountable — to the program or to their constituency. Stakeholder participation in the formulation of these agreements and their public disclosure also strengthens the accountability of program governance.

12.15 **Responsibility**. This concerns the extent to which the program accepts and exercises responsibility to stakeholders who are not directly involved in the governance of the program and who are not part of the direct chain of accountability in the implementation of the program. As international public sector organizations, GRPPs should be ahead of the curve when it comes to "corporate social responsibility." For instance, they should be adhering in their operations to accepted global norms regarding human rights, poverty reduction, environmental sustainability, and gender inclusion. They should be obligated to report responsibly on their adherence to these norms, to adhere to social and environmental safeguards, to disclose potential or realized adverse effects, and to propose mitigation plans.

12.16 **Fairness**. This concerns the extent to which partners and participants, similarly situated, have equal opportunity to influence the program and to receive benefits from the program. To what extent does access to information, consultation, or decisions of the governing body and management favor the interests of some partners and participants over others, at both the governance and management levels? Fairness can be impeded not only by structures and processes, but also by language, technical, and legal barriers.

12.17 **Transparency**. This concerns the extent to which the program's decision-making, reporting, and evaluation processes are open and freely available to the general public. To what extent does the program have a policy on transparency and disclosure that covers governance and management, decision making, accountabilities, staffing, contracting, dissemination, financial accounting, auditing, and

M&E?[72] To what extent do these policies meet or achieve good-practice standards such as publicly disclosing the minutes of all board meetings (at least in summary form)? To what extent are they being applied? (See also Chapter 5, Transparency and Disclosure).

12.18   **Efficiency.** This is closely related to the efficiency or cost-effectiveness of the program as a whole (discussed in Chapter 11). The concern here is the extent to which the governance and management structures enhance efficiency or cost-effectiveness in the allocation and use of the program's resources. Theory suggests that traditional shareholder models of governance (in which membership on the governing body is limited to financial and other contributors) may be more efficient but at some cost to legitimacy, while stakeholder models (in which membership also includes non-contributors) may be more legitimate but sometimes at the expense of efficiency, if the number of participants becomes large and the costs of organizing diverse interests to pursue a common goal becomes high relative to the expected benefits. For both types of programs, evaluators need to recognize the tensions that exist between legitimacy and efficiency, and ascertain if one principle is being sacrificed at the expense of the other, since effective governance requires both. In reality, a certain degree of convergence of practice appears to be taking place between programs that had previously followed shareholder and stakeholder models, respectively.

12.19   **Probity**. This refers to the adherence by all persons in leadership positions to high standards of ethics and professional conduct over and above compliance with the rules and regulations governing the operation of the program. Members of the governing, executive, and advisory bodies, as well as members of the management team, must exercise personal and professional integrity, including the avoidance of conflicts of interest. Evaluators sometimes discover evidence of wrong-doing, fraud, or misconduct. When they do, they should report their findings confidentially to the appropriate investigating authority, and in severe cases, even discontinue the evaluation. (See also paragraph 8.15 on wrong-doing, fraud, and misconduct.)

---

72. On issues of transparency, the principles for governance and management diverge to some extent. Governance processes need to be open to ensure accountability and responsibility to shareholders and stakeholders. However, certain management processes, particularly the management of human resources, need to be confidential in order to protect the privacy of individuals. Thus, management should have some discretion in determining the appropriate disclosure of information in relation to the day-to-day management of the program. (See also footnote 35 on page 27.)

## SUGGESTED STRATEGY FOR ASSESSING GOVERNANCE AND MANAGEMENT

12.20    The assessment of governance and management should begin with an analysis of the role and performance of the program's governing bodies and their relationship to the host organization, if applicable. This would include:

Draws on IEG's experience with reviewing GRPPs

- A full description of the governing bodies and management units (including executive and advisory bodies), their representation and mandates, and their evolution in relation to the maturing of the program. To what extent are their roles and responsibilities clear, as well as the mechanisms to modify and amend these over time?

- If the program is housed in another (host) organization, a full description of the legal and administrative relationships between the program and the host organization, including the mechanisms in place to resolve disputes between the two parties. To what extent are these clearly articulated? What are the benefits and costs to the program of being located in the host organization?

- The extent to which the assignment of functions and decision making to different bodies — and the host organization — has been appropriate in relation to the goals of efficiency, timeliness, application of needed expertise, representation, and inclusion.

- The performance of each governing body and the program management unit relative to its terms of reference, expected duties, and commitments. This would include the performance of the host organization and partners if and as specified in the evaluation TOR.

## SPECIAL CONSIDERATIONS IN ASSESSING GOVERNANCE AND MANAGEMENT

12.21    How governance is practiced and who actually influences the program's direction is rarely understood from a cursory examination of a program's charter, organizational charts, and terms of reference documents. History, culture, personalities, the quality of relationships, and path-dependence [73] can all influence practice and effectiveness.

Draws on IEG's experience with reviewing GRPPs

---

73. *Path-dependence* is the dependence of institutional choices and economic outcomes on the path of previous choices and outcomes, rather than simply on current conditions. In path-dependent processes, institutions are self-reinforcing, history has an enduring influence, and choices are made on the basis of transitory conditions that persist long after these conditions change. Thus, understanding path-dependent processes — such as the standard typewriter keyboard — requires looking at history, rather than simply at current conditions of technology, preferences, and other factors that influence outcomes.

Evaluators should review the "rules of the game" (implicit and explicit) to ensure that all partners and participants, similarly situated, can participate equitably. For instance, if some members of the governing body are permanent and others are rotating, does decision making effectively reside mainly within the purview of the permanent members?

12.22    Evaluators should assess how effectively the roles and responsibilities of the various partners, participants, and host organizations are articulated at each level. To what extent are these clear, appropriate (in terms of influence over decision making), and being followed? For regional partnership programs in particular, this includes the extent to which the program has clearly delineated the roles and responsibilities for program implementation between the regional and national levels, and then followed through.

12.23    Evaluators should assess the appropriateness of the specific mix of partners at the governance level and participants at the implementation level. Does the program have the right partners and participants to achieve its objectives?[74] Are the respective mandates of the international organization partners sufficiently convergent to effectively address the global/regional challenges in question? Are there any significant actors in the sector at the global/regional levels who are missing from the partnership, and why? Have country-level representation and voice been adequate in relation to the issues being addressed and the objectives of the program? To what extent have the necessary stakeholders been involved?

12.24    Evaluators should assess to what extent the program is seeking the most up-to-date scientific and technical advice from the point of view of policy making and operational effectiveness. For this explicit purpose, many GRPPs have established scientific and technical advisory bodies in order to seek advice from experts who are not entitled to representation on the formal governing body. Such advisory bodies can enhance the program's professional reputation and help weigh the risks of alternative strategies. Evaluators need to assess the performance of these bodies, where they exist, in achieving their mandates and in contributing to the legitimacy and effectiveness of the program.

12.25    An assessment of the "partnership mix" should also comment on the participation (or lack thereof) of NGOs and the commercial private sector. Since GRPPs are international public sector initiatives

74. This is a particularly important question for regional partnership programs. For both global and regional programs, the way in which the individual efforts add up to the collective outcome for the program — whether "best shot," "summation," or "weakest link" — can also help determine who needs to be involved in order for the program to achieve its objectives. (See paragraph 9.17.)

that are aiming to promote a public interest in a particular area of development, partnerships with the commercial private sector and NGOs may pose legitimacy issues at the governance level and conflict of interest or favoritism issues at both the governance and implementation levels. Evaluators should assess whether the program has established and is effectively applying a policy that addresses such concerns. Do the benefits from such partnerships outweigh the reputational and other risks to the program?

### PROGRAMS LOCATED IN HOST ORGANIZATIONS

12.26    The majority of GRPP secretariats are located in existing international organizations or bilateral agencies, and the managers of such programs typically report both to the program's governing body and to their managers within the host organization — a classic "two-masters" problem. Evaluators should ascertain to what extent this arrangement is adversely affecting the governance and management of the program, since there has frequently been a lack of precision concerning for what functions the program manager is accountable to each "master," and how conflicts between the two are to be resolved.[75]

Draws on IEG's experience with reviewing GRPPs

12.27    The host organization often exercises a major influence over the strategic direction of the program as well as bearing a disproportionate share of responsibility for oversight, consultation, risk management, and evaluation. If so, evaluators should ascertain if such a dominant role of the host organization in the governance and management of the program is leading to organizational capture and adversely affecting the program's performance (or other criteria such as transparency and fairness).[76] For instance, is this reducing the incentives of other partners to participate effectively in the program, or reducing the ability of the host organization to look at the weaknesses of the program objectively?

75. Who determines the performance of the program manager is a particularly complex issue. In some case, managers' performance evaluations are completed as if they were employees of the host organization. In other cases, feedback is obtained from members of the governing body. See also Michael Davis and Andrew Stark, eds., 2001, *Conflict of Interest in the Professions*, for more on the "two-masters" problem.

76. *Organizational capture* means that the host organization takes over and runs the program as if it were one of its own. Therefore, the respective roles and responsibilities of the host organization and the program need to be clearly specified and understood. And the relationship between the host organization and the GRPP must be properly managed in order to ensure appropriate accountability all the way down to the country level, where the lead country representative of the host organization may be to some extent accountable for what the program is doing at that level.

# 13.   Resource Mobilization and Financial Management

## Principles and Norms

### DEFINITIONS

13.1   *Resources* are the inputs that are used in the activities of a program. Broadly speaking, the term encompasses natural, physical, financial, human, and social resources, but the vast majority of the resources that make up the inputs to GRPPs are financial resources. In-kind resources such as the provision of office space, seconded staff, or partner participation at board meetings are a second level of resources.

13.2   *Resource mobilization* is the process by which resources are solicited by the program and provided by donors and partners. This is particularly important for GRPPs, since GRPPs are typically externally financed programs with little or no capacity to earn income from their own resources. Most are public sector programs, which typically provide goods and services (including financial resources) to beneficiaries on a grant or in-kind basis.

13.3   The process of mobilizing resources begins with the formulation of a *resource mobilization strategy*, which may include separate strategies for mobilizing financial and in-kind resources. Carrying out a financial resource mobilization strategy includes the following steps: identifying potential sources of funds, actively soliciting pledges, following up on pledges to obtain funds, depositing these funds, and recording the transactions and any restrictions on their use. The process is generally governed by legal agreements at various stages.

13.4   Resource mobilization strategies and processes may be constrained by parameters or rules established by the partners at the inception of the program and recorded in the charter or initiating legal documents. For example, these may require donors to contribute a minimum amount per year in order to have a seat on the governing body. They may specify that funds cannot be accepted from private sector sources, or only under certain conditions. Or they may require separate accounts for different expected uses of funds, which would affect the recording of the deposits.

*13.5   Financial management* refers to all the processes that govern the recording and use of funds, including allocation processes, crediting and debiting of accounts, controls that restrict use, and accounting and periodic financial reporting systems. In this Sourcebook, financial

Elaborates on definition of inputs in DAC Glossary

Draws on IEG's experience with reviewing GRPPs

management also includes the processes which ensure that funds are used for the purposes intended — a fiduciary standard that is expected by the vast majority of donors.[77] In cases where funds received accumulate over time, it would also include the management of the cash and investment portfolio.

NEED FOR GRPP EVALUATIONS TO COVER RESOURCE MOBILIZATION AND FINANCIAL MANAGEMENT

Draws on IEG's experience with reviewing GRPPs

13.6    For GRPPs, it is important to review resource mobilization and financial management from both a static and a dynamic perspective. From a *static* perspective, the financial resources at any point in time are the major input that determines results, and analyzing their sources and uses is an essential part of tracking progress and attributing results to the program. From a *dynamic* perspective, the processes of formulating the resource mobilization strategy, managing the peculiarities of responding to diverse donor funding cycles, and committing and allocating funds need to be examined in their own right, because these affect the ability of the program to achieve its objectives on its current scale — as well as the potential to achieve its objectives on a larger scale or in new ways. Accountability for the final use of funds in a strict legal sense, however, is normally done through the formal audit process. (See also paragraph 1.7.)

13.7    At a minimum, all GRPP evaluations should describe the sources and uses of public and private funds for the program and assess how the patterns of financing have affected the scope, reach, and results the program achieved. They should also analyze the allocation processes and any effects that donor restrictions (such as tying or earmarking funds to particular activities) have had on the achievement of the program's objectives. Also, the evaluation should include — in any assessment of the strategy of the program — the degree to which the program's resource mobilization strategy and execution is adequate to meet the needs of the program and to achieve its desired scale. This assessment may be linked to the assessment of governance, since the involvement of new donors may affect the dynamics of the governing body. Finally, it may also be important to assess the degree to which the financial management system and finan-

---

77. Most legal agreements involving official, multilateral, private, or foundation donors will contain a phrase calling for assurances that the funds are used as intended. Only selected individual donors (particularly anonymous individual donors) typically provide contributions to the program as a whole without a legal agreement that sets expectations on reporting or fiduciary assurances.

cial reporting are meeting the expectations of donors, since this can have a significant effect on mobilizing resources.[78]

## Standards and Guidelines

13.8     To assess the effectiveness of the program's resource mobilization and financial management system, evaluators should consider:

Draws on IEG's experience with reviewing GRPPs

- **The link between governance and financing**. For example, are there financial requirements, such as minimum annual contributions, that condition membership in the governing body? Does this effectively exclude some potential stakeholders (such as beneficiary countries) from participating in governance? Does the participation of some donors on the governing bodies discourage other donors from contributing? Should different roles for different types of donors (such as the private sector and individuals) be considered?

- **The role of the governing body in mobilizing resources**. Is the governing body appropriately exercising its role in (a) guiding the formulation of a resource mobilization strategy responsive to strategic directions; (b) setting policy rules regarding acceptance of tied or earmarked funds, private sector funds, or different financial instruments such as promissory notes; and (c) staying open to the possibility of new donors, including private donors, foundations, and, if applicable, "emerging official donors" (that is, former developing countries that have graduated from development assistance)?

- **The prospects for beneficiary country or local partners to make financial contributions to the program now or in the future, particularly in regional partnership programs.** Does the resource mobilization strategy address this issue? Has a timeline been established for the country partners to take over more responsibility for financing and implementation of program activities at both the national and regional levels? (See also paragraphs 14.14 and 14.16–14.19 on strategy for devolution or exit.)

---

78. Usually the evaluation will not assess financial controls in detail. If any concerns are expressed by donors or management, the evaluators may recommend an audit or more detailed assessment by financial specialists. (See paragraph 1.7.)

- **The quality of financial management and accounting.** Have financial management systems met all standards of trustees and contributing donors? Are financial reporting and auditing arrangements satisfactory? Do the recorded categories of expenditures facilitate adequate monitoring and attribution of costs to activities and results?

- **The methods, criteria, and processes for allocating funds**. Are the processes and criteria that have been established for allocating financial resources to activities being applied? To what extent have these evolved over time in response to new priorities or objectives? How effective and efficient are these processes? (See also Chapter 12, Governance and Management.)

RESOURCE MOBILIZATION AND FINANCIAL MANAGEMENT IN THE EARLY STAGES OF A PROGRAM

13.9     For the newer GRPPs, the evaluation should include an analysis of the performance of the program in mobilizing and deploying initial donor resources in its first phase while moving to a more sustainable model of financing over time. This may include: (a) the manner in which partners were chosen and funds channeled and allocated; (b) whether co-financing and/or counterpart funding was sought; and (c) decisions on the organizational structures and staffing[79] as related to donor relations and reporting.

DONOR RESTRICTIONS ON USE OF RESOURCES

Draws on IEG's experience with reviewing GRPPs

13.10    A GRPP evaluation should compare the costs and benefits of such constraints imposed by donors. On the one hand, the need to accommodate donors' preferences, expressed through tied funding arrangements or earmarking, can constrain program-wide prioritization processes and result in an inefficient allocation of resources. On the other hand, channeling the additional funds through the program rather than to uncoordinated parallel activities may have important benefits, such as expanding the scale or scope of the program, adopting a new, special focus for the program, or better aid coordination.

---

79. While evaluations should broadly assess the degree to which the management of financial resources is meeting the fiduciary expectations of donors, assessing the management of human resources may fall outside the TOR.

# 14.  Sustainability, Risk, and Strategies for Devolution or Exit

## Principles and Norms

14.1    *Sustainability* is the continuation of benefits from a development intervention after major development assistance has been completed. It is also the probability of continued long-term benefits and the resilience to risk of the net benefit flows over time.

Based on DAC Glossary

14.2    *Sustainability*, when **applied to organizations or programs**, refers to the likelihood that the organization or program will be able to continue its operational activities over time. This may depend on a number of factors, such as the continued relevance and legitimacy of the program, its financial stability, its continuity of effective management, and its ability to withstand changing market or other conditions.

14.3    *Risk to development outcome* is the risk, at the time of evaluation, that the expected outcomes will not be realized or maintained. This has two dimensions: (a) the likelihood that some changes may occur that are detrimental to the ultimate achievement of the expected outcomes, and (b) the affect on the expected outcomes if some or all of these changes actually materialize. Risks may be internal to the program or arise from external factors at the country or local level (such as prices) or at the global level (such as technological change). The actual effect of these risks on the ultimate outcomes will depend on both the severity and nature of the changes that occur and on the adaptability (or lack thereof) of the design of the program and its activities. Ideally, the potential risks to the expected outcomes should have been identified at the inception of the program and pertinent indicators for their monitoring included in the M&E framework.

Based on IEG evaluation criteria

14.4    *A strategy for devolution or exit* [80] of the program refers to a proactive strategy to change the design of the program, to devolve

---

80. *Exit strategy* is the term used by grant makers in the grant-making and foundation literature to refer to the "weaning" of a program from grant support, which then allows the grant maker to spread its support more widely to new programs. Some grant-makers even require an exit strategy as a condition for the initial provision of grant funds. In this Sourcebook, the concept of *exit strategy* is defined from the perspective of the governing body or management unit of the program, and refers specifically to the program as a whole phasing out its operations. It does not refer to the program "exiting" or ending its support for activities in specific countries that have no further need of the program.

some of its implementation responsibilities, to reduce dependency on external funding,[81] or to phase out the program on the grounds that it has achieved its objectives or that its current design is no longer the best way to sustain the results which the program has achieved.[82] Other possible strategies include *transforming the program* into an informal network of country or local implementers or *spinning off the program* and establishing a *new legal entity* that is no longer hosted by one of the partner organizations. These possibilities pose questions similar to those for sustainability, but from a different perspective: Does the program need to be sustained? Is the continuation of the program the best way of sustaining the results achieved? Should the design of the program be modified as a result of changed circumstances (either positive or negative)? What other alternatives should be considered to sustain the program's results more cost-effectively?

## NEED FOR GRPP EVALUATIONS TO ASSESS SUSTAINABILITY

<div style="float:left">Draws on IEG's experience with reviewing GRPPs</div>

14.5    The TOR should clearly specify to what extent the evaluation should assess (a) the sustainability of the benefits arising from the activities of the program and/or (b) the sustainability of the program itself, since it is not appropriate for all evaluations to do so. Among the various features of GRPPs that could be taken into account in making this decision (Table 7), it is more appropriate for evaluations to assess the sustainability of the benefits of *mature* programs compared with young programs that have not yet had the opportunity to complete many activities or achieve many outcomes. (See also paragraph 6.7.) For mature programs, the evaluation should also attempt to determine why the benefits of the activities are or are not sustainable. For younger programs, it is probably more appropriate for the evaluation to focus on the extent to which the program is effectively planning for the sustainability of country or local-level activities after GRPP support ceases.

14.6    It is also more appropriate for evaluations to assess the sustainability of *mature partnerships* that are still relevant and legitimate and generating benefits that are worth sustaining. Among other things, the partnership could be at risk of diminished legitimacy (if interests of key stakeholders diverge) or of declining financial resources (if competing programs emerge in the sector).[83]

---

81. This generally means through increased cost sharing with beneficiary groups or taking advantage of revenue-earning opportunities. But it could also involve diversifying sources of donor funding.

82. The governing body could also adopt an exit strategy on the grounds that the program has failed to achieve its objectives and is not likely to do so, even with a change in design.

83. For instance, a new GRPP may emerge that is aimed at achieving similar goals as an existing GRPP, but using different approaches or mechanisms,

Table 7. Features of GRPPs to Consider in Deciding Whether to Include Sustainability in the Scope of an Evaluation

| GRPP Feature | Implications for Assessing Sustainability |
|---|---|
| Unlike projects, GRPPs do not have predetermined end-points. Programs mature and evolve, revising both their objectives and their approaches to achieving their objectives over time. | Depending on the findings with respect to relevance, effectiveness, and efficiency, the evaluation of a mature program should assess whether the program should continue to grow, modify its objectives or strategies, or consider alternative strategies such as devolution or exit. |
| GRPPs are diverse in size, sectoral focus, and type of activities supported (advocacy, knowledge dissemination, technical assistance, or investments). | While sustainability is difficult to assess for activities related to advocacy and knowledge dissemination, evaluations of technical assistance or investments should normally include an assessment of sustainability. |
| GRPPs comprise multiple stakeholders, whose interests do not always coincide. | If there are signs of diverging relevance from the perspective of the different stakeholders, assessing the sustainability of the program is essential. Among other things, this would take into account the likelihood of continued political support by different stakeholders and the possible need to consider a strategy for devolution or exit. |
| GRPPs take several years to set up. Sunk costs are relatively high, especially at the initial stages. | It may take many years for the program to reach sufficient maturity to meaningfully assess whether it is reaching its potential and to appropriately consider issues of sustainability, devolution, or exit. |
| GRPPs are typically grant-financed, with little capacity to generate revenue from their own resources. The availability of adequate financing has implications for the sustainability of both the program's outcomes and the partnership itself. | For programs producing public goods, financing in the early stages depends crucially on donor contributions, and the evaluation should assess the sustainability of their financial support. For more mature programs producing public goods at the country or local level, it is appropriate to assess whether country or local governments could contribute to the costs, and to what extent this would be sustainable. For programs producing some private goods, an assessment of the feasibility of user charges as a means of ensuring financial sustainability may be appropriate. |
| GRPPs operate at multiple levels — global, regional, national, and local. Most governance and management functions (including the location of secretariats) are undertaken at global or regional levels. | For mature programs, if there is a strategy for devolution, it may be necessary to examine alternative arrangements to sustain the benefits arising from the program. It may be easier to devolve the implementation responsibilities for country or local-level activities than the coordination of the knowledge management and capacity-strengthening functions typically provided by the global/regional secretariats of GRPPs. |

while targeting essentially the same beneficiaries. As some members of the existing partnership move to form such a competing program, an existing GRPP may be at risk of becoming irrelevant or of losing much of its funding base.

14.7    Some GRPPs function primarily as advocacy or knowledge networks. Although their goal is for stakeholders to **apply** the knowledge that the program has generated and disseminated where applicable, and although benefits may result at the regional, country, or community levels, the GRPP in question **may not have directly supported** any of the interventions at these levels. In such cases, it may not be feasible or cost-effective to assess the sustainability of the outcomes of such advocacy or knowledge activities. Attribution may be very difficult, if not impossible to demonstrate, especially if the M&E framework is not adequately robust. The governing bodies of these types of GRPPs should carefully weigh the pros and cons of including an assessment of sustainability in the scope of the evaluation.

### NEED FOR GRPP EVALUATIONS TO ASSESS STRATEGIES FOR DEVOLUTION OR EXIT

Draws on IEG's experience with reviewing GRPPs

14.8    A strategy for devolution or exit may or may not figure in a program's strategic documents, depending on its maturity or the requirements of its donors. The evaluation TOR could call for an assessment of the appropriateness of this strategy if it exists, or for assessing the relative merits of a range of alternative strategies if one does not presently exist. There have not yet been many assessments of existing, potential, or implemented exit strategies of GRPPs to date, because GRPPs are a fairly new phenomenon.[84] But in view of the recent growth of GRPPs, it is important that more program-level evaluations do so.

14.9    GRPPs typically have large sunk costs, which are relevant up to a point. Nonetheless, in the case of a mature program, the evaluation should pose the question of whether the program should continue in its present form or at all. GRPPs are typically grant funded from scarce development resources. Regardless of their origin (developed and developing country budgets, private foundations, or the private sector), these scarce resources should be applied to development interventions that are most effectively and efficiently designed and implemented. Evaluations should help guard against the perpetuation of programs that no longer meet the criteria of relevance, effectiveness, efficiency, and sustainability.

Based on IEG's forthcoming review of regional programs

14.10    It is particularly important to include the issues of sustainability, devolution, and exit in evaluations of regional programs with a contiguous geographical dimension to them, such as body of water, a

---

84. IEG reviewed the experience of several GRPPs that exited from the World Bank in its Phase 2 Report (2004), and the Bank's Development Grant Facility has reviewed the experience of 61 GRPPs that exited from grant support through June 2004. But these represent exits from the perspective of the Bank or the Development Grant Facility, not from the perspective of the governing body or management unit of the program.

river system, or a transport or power system. Such regional programs typically exist for the specific purpose of resolving collective action dilemmas regarding the use of the common natural resource. These programs need to plan for the sustainability of both national-level activities and regional coordination arrangements when external donor support ceases. Member countries have generally been more willing to assume responsibility for financing the continuation of the national-level activities than the regional coordination arrangements, except where the latter costs can be covered by self-generating resources (such as an electric power grid). So the financing of regional coordination arrangements has continued to be borne largely by external donor sources. To what extent can this be sustained, or should alternative financing mechanisms be more vigorously explored?

## Standards and Guidelines

### ASSESSING SUSTAINABILITY OF THE BENEFITS OF GRPP ACTIVITIES

14.11    Sustainability answers the following questions: At the time of evaluation, to what extent are the benefits arising from GRPP activities likely to be sustained beyond the planned life of the activities supported by the GRPP? The answer is likely to depend on the extent to which the program, in its early stages, built in measures to strengthen local capacity and ownership. In addition, the evaluation should assess the resilience of the future stream of benefits to changes in conditions external to the influence of the program. How sensitive are the benefits to future changes in the local operating environment? How well can the mechanisms put in place by the activities continue to generate the benefits, while weathering shocks and changing circumstances in the political, economic, environmental, or social arenas?[85]

Draws on internal IEG guidelines

14.12    Some factors to take into account in assessing the sustainability of the benefits arising from the activities of a GRPP include the following:

- Financial resilience (including policies on resource mobilization, cost recovery, operation and maintenance, and budgeting for contingencies)
- Government demand and ownership, if relevant (by both central government agencies and implementing agencies)

---

85. Ideally, these questions should be addressed in activity completion reports at the completion of individual activities, and as part of the monitoring system of the program. The evaluators would normally only attempt to validate these results for a sample of completed activities.

- Other stakeholder ownership (which may be influenced by lo-cal participation, beneficiary incentives, civil society/NGO advocacy, and private sector linkages)
- Institutional support (including a supportive legal and regula-tory framework; organizational and management effectiveness in implementing entities, whether public or private; and sup-port for capacity strengthening)
- Social support (including safeguard policies and the availabil-ity of complementary services from other agencies or NGOs in case of an interruption of GRPP services)
- Ability to adapt to exogenous influences (such as changing technologies, competing global development priorities, new sources of donor funding or expertise, regional political and security situations, and natural disasters).

14.13    Specific indicators for assessing the sustainability of the bene-fits arising from the activities of a GRPP will depend on the maturity of the program. For most GRPPs, it may be more meaningful to focus on the sustainability of key expected outcomes, rather than on net benefit flows arising from specific GRPP activities, because the latter are not easily measurable.

### ASSESSING THE SUSTAINABILITY OF THE PROGRAM

Draws on IEG's experience with reviewing GRPPs

14.14    An assessment of the sustainability of the GRPP itself could address the following questions: Assuming that the program is still judged to be relevant (see Chapter 9), to what extent is the partner-ship sustainable? To what extent are the range and depth of political commitment, support, and financing for the program and its objec-tives sustainable? Given the multiple and wide range of members and stakeholders, to what extent is there still sufficient convergence or ac-commodation of interests to sustain the program? Has the program developed institutional capacity in the following areas necessary for sustainability: knowledge management that helps the program stay attuned to external conditions and markets; learning programs to up-date skills and knowledge to meet changing program requirements; personnel policies that attract and retain staff; and performance-based management that helps ensure self-correction and steady progress toward program objectives?

14.15    In what areas could the program improve in order to enhance the likelihood of sustainability — such as better marketing of program achievements to uphold its reputation, improved knowledge man-agement and dissemination of program outputs to enable their wider application, changes in the governance and management arrange-ments, and exploration of alternative resource mobilization strategies. The evaluation should shed light on these important issues and le-

gitimize the continuation of a relevant partnership program, or rec-
ommend steps to address the concerns of its constituents.

## ASSESSING PROSPECTS FOR CONTINUATION AND STRATEGIES FOR DEVOLUTION OR EXIT

14.16    When a significant portion of the benefits of a GRPP are most
effectively delivered at the *global or regional level* (such as advocacy
or research and development) and attribution of the beneficial out-
comes to the GRPP is clear, the case for continuing the global or re-
gional program may be strong. But many GRPPs evolve over time and
take on new challenges (which may be relevant but only remotely re-
lated to the original mission statements). In assessing the prospects for
continuation, the evaluation needs to take into account this evolution
of objectives and explicitly assess the degree to which the "reinvention
of the program" has been justified by continuing relevance and by de-
mand from beneficiaries. The vested interests of those involved in the
program's governance or management to continue the program are
also a factor that should be taken into account in such an assessment.

Draws on IEG's experience with reviewing GRPPs

14.17    When asked to assess the prospects for continuation of the pro-
gram, the evaluation team should clarify with the commissioners of
the evaluation to what extent the TOR incorporates an assumption of
the desirability of continuing the program as it is currently designed,
or to what extent the TOR regards different organizational and finan-
cial arrangements or strategies for devolution or exit as under consid-
eration.

14.18    Table 8 illustrates when the need for such assessments may arise
and the questions that could be addressed under different scenarios.
The merits of different strategic options should be assessed in the light
of the previous evaluation findings with respect to relevance, effective-
ness, efficiency, and sustainability. For instance, the evaluation might
find that the GRPP has achieved most of its relevant objectives in its ex-
isting form, or that the GRPP is no longer the most legitimate or effi-
cient means of sustaining the benefits arising from its activities. In these
cases, the evaluation could assess the relative merits of the following
range of possibilities: (a) reinventing itself, such as changing its objec-
tives or increasing its reach; (b) phasing out the program; (c) modifying
its implementation arrangements, such as devolving responsibility for
implementation to the regional, country, or local levels; (d) seeking al-
ternative sources of grant-financing or revenue generation; or (e) taking
on new organizational forms, such as spinning off from the host organi-
zation and establishing an independent legal entity.

14.19    Even if the need for such an assessment has not been expressed
in the TOR, the evaluator could, after the evaluation findings have
been shared and conclusions reached, recommend that such an as-

sessment take place. If the TOR for the evaluation does not explicitly refer to issues regarding the continuation of the program or to a strategy for devolution or exit, then such questions may be best dealt with by the evaluation team as secondary questions that derive from the implications of its findings with respect to relevance, results, and sustainability.

Table 8. Indicative Questions for Assessing Strategies for Devolution or Exit under Different Scenarios

| Possible Scenarios | Possible Strategies | | | | |
| | Reinvent with same governance and funding | Consider phasing out the program | Continue country or local-level activities with or without devolution of implementation | Seek alternative financing arrangements, such as revenue-generation, or self-financing to reduce dependency on external sources | "Spin off" from host organization |
| | A | B | C | D | E |
| The objectives that led to the establishment of the GRPP have all been accomplished, or the objectives are no longer judged to be relevant. | Should the program set new objectives? (Check for comparative advantage and competition from other programs.) | Should the program be phased out? (Consider costs and benefits.) | Assuming that there are activities at the country or local level supported or induced by the GRPP that could continue to yield positive benefits even if the program were phased out, are there country or local-level institutions that can sustain them? | Should different financial arrangements be considered, including mobilization of resources at the country or local level? (Only relevant if A or C is also true, and then only relevant in the longer term.) | If A or B is selected, are there central/regional support activities that need to be provided through some other means? Would an informal association be sufficient? Or would a new legal entity be beneficial? |
| The objectives of the GRPP are still relevant and there is more to be accomplished; the strategy is working and outcomes are sustainable. | Not an immediate issue | Not an immediate issue | Could more be accomplished with devolution of implementation responsibilities? | Are there ways for the program to generate revenue or self-finance (such as introducing charges) that would not adversely affect results or sustainability? | Are there any benefits to be reaped by spinning off central functions from the host organization? |
| The objectives are still relevant and there is more to be accomplished, but the strategy is not working and sustainability is in doubt. | Should the program modify its strategy? | Can the program modify its strategy? If not, should it be phased out? | Would a strategy with more devolution of implementation responsibility improve results and sustainability? | Would different financing arrangements modify incentives in such a way that would improve the results of the strategy and sustainability? | Would spinning off from a host organization modify incentives in such a way that would improve the results of the strategy and sustainability? |

# 15.   Impact Evaluation

## Principles and Norms

### DEFINITION

15.1    ***Impact evaluation*** is one of a range of evaluations that may be applied to GRPPs at any given time, but usually after the program has evolved to a steady state. It is commonly defined as the systematic assessment of the effects — positive or negative, intended or unintended — of one or more development interventions on the final welfare outcomes of the affected individuals, households, and communities, and the extent to which these outcomes can be attributed to the development intervention(s). In its most rigorous form, an impact evaluation compares the welfare outcomes of the intervention(s) during the period being evaluated with an explicit ***counterfactual*** — the hypothetical situation that would have prevailed in the absence of the intervention(s). Different approaches to impact evaluation include quantitative impact evaluation, participatory impact evaluation, and theory-based (program logic) approaches. Good impact evaluations will combine all three approaches. (See standards below.)

Draws on materials presented at the International Workshop on Impact Evaluation for Development, November 15, 2006

### NEED FOR IMPACT EVALUATION

15.2    In spite of the increased focus on achieving final development outcomes such as those in the Millennium Development Goals, credible impact evaluation studies, which provide scientific evidence of causal links between ongoing development interventions and final welfare outcomes, are fewer than would be desirable to help guide priority setting in development aid. This is as true for GRPPs as for other forms of development assistance. While most GRPPs undertake periodic evaluations, these are usually formative evaluations for improving a specific aspect of the program's performance, or summative evaluations of program outcomes, rather than rigorous impact evaluations. However, impact evaluations can effectively complement or contribute to these formative and summative evaluations in providing accountability and in confirming that development funds have been spent wisely on effective interventions. While a rigorous impact evaluation of a GRPP at the program level would be extremely difficult, because of the diversity of its components and the resultantly complex causality and aggregation problems, impact evaluations of selected activities are feasible and are encouraged.

### ADVANCE PLANNING FOR CONDUCTING IMPACT EVALUATIONS

15.3    As emphasized in Chapter 2 (paragraph 2.23), impact evaluation should be planned in advance, for several reasons. First, like all evaluations, impact evaluation needs to be based on accurate data,

which come from a mature and tested monitoring system. Second, to be the most credible, it is necessary to compare the welfare outcomes arising from the program with a counterfactual — what would have occurred in the absence of the program. A good technique for establishing the counterfactual consists of identifying one group receiving the intervention(s) and a similar control group not receiving the intervention(s), and then initiating early baseline data collection relating to *both groups* before the intervention(s) begin. This ensures that adequate information is available for the subsequent comparison of the situation with and without the intervention(s), once these have been in place long enough to have had an impact. This *"double-difference"* approach requires the early establishment of a research design and the early collection of baseline data, ideally even before the potential beneficiaries learn of the intervention and develop expectations that may affect their behavior.

15.4     If baseline data have not been collected, an impact evaluation can still be conducted by comparing the welfare outcomes of the group receiving the interventions with those of a control group, while attempting to control for other influences through statistical methods. This *"single difference"* approach also requires careful design and planning, and is also dependent on an established and tested monitoring system.

## Standards and Guidelines

### PLANNING FOR A PARTICULAR IMPACT EVALUATION

Draws on materials presented at the International Workshop on Impact Evaluation for Development, November 15, 2006

15.5     Impact evaluation needs to be planned carefully and employed selectively as one of several types of useful evaluations that can serve different purposes at different stages of a program. Impact evaluation would normally be considered more feasible and relevant after the program has reached a steady state in terms of financing, scope, and coverage. Because impacts are not usually manifest until after the passage of some time, the scope of an impact evaluation may cover only a subset of activities which have reached a certain stage of gestation or which were completed during a previous period of the program's life.[86] If at any stage, a future quantitative impact evaluation of any intervention(s) is considered likely, the program should make provision for the collection of baseline data to provide the basis for comparison with the counterfactual.

---

86. The Consultative Group on International Agricultural Research (CGIAR) is the major GRPP that has conducted impact evaluations on its productivity-enhancing agricultural research. Some of the health programs, such as the Special Programme on Research and Training in Tropical Diseases (TDR), have also evaluated the impacts of their research on diseases of the poor such as onchocerciasis, leprosy, and malaria.

15.6    Impact evaluations of a specific intervention can be conducted as soon as it is judged likely that welfare outcomes have been realized. However, since impact evaluations are generally more costly than other forms of evaluation, the governing body should carefully consider the costs and benefits of conducting an impact evaluation of a given set of interventions at a given time. As a general guideline, for accountability and for assurance of continued relevance, the governing body should consider financing a comprehensive impact evaluation after 10 years of a program's life.[87]

CONDUCT OF IMPACT EVALUATIONS [88]

15.7    Impact evaluations will usually be conducted for subsets of activities where impact is judged to be more measurable than for the program as a whole, or where there is a pressing need for an assessment of impact to influence design adjustments or decisions on replicability and scaling up. Impact evaluations will normally be conducted parallel to and not as part of a program-level evaluation. If the results of the impact evaluation are to be used in subsequent program-level evaluations, it is important that the sample for the impact evaluation be chosen in order to be representative. For example, the sample might include people from one village where the conditions seem favorable for high impact and from another where conditions are less favorable. Selecting an appropriate comparison group and avoiding selection bias are two of the major challenges in impact evaluation.[89]

15.8    Good impact evaluations use a combination of quantitative impact evaluation, participatory impact evaluation, and theory-based (program logic) approaches. Qualitative participatory analysis helps to add context to and provide confirmation of findings derived from the other approaches. A theory-based approach helps to track the influences at different points in the results chain and to enhance understanding of when or why the program works well or not. Quantitative methods give an authoritative and credible indication of the relative impact of the program, compared with the counterfactual situation.

87. Several GRPPs that have existed for more than 15 years have never had an impact evaluation.

88. Impact evaluation is the subject of an ongoing working group of the OECD/DAC, and more detailed guidelines are expected.

89. *Selection bias* is the distortion that arises in a statistical analysis due to the methodology that was used to collect the samples. For instance, the beneficiaries of a certain intervention may be selected (or self-selected) on the basis of certain characteristics. If these are observed, then it is important to select a comparison group with the same characteristics. If these are unobserved, then only a randomized approach can in principle eliminate the selection bias.

# EVALUATION CHECKLISTS

## 16.  Terms of Reference

### Principles and Norms

#### NEED FOR TERMS OF REFERENCE TO ADDRESS ALL STAKEHOLDER CONCERNS

16.1    The evaluation TOR should address issues of concern to each group of stakeholders. The TOR or the evaluation team should specify how the views and expertise of groups affected by the program would form an integral part of the evaluation.

*Based on DAC Principle VI, paras. 23 and 24*

### Standards and Guidelines

#### PURPOSE AND CONTENT OF THE TERMS OF REFERENCE

16.2     The TOR should provide the purpose and describe the context, process, and product(s) of the evaluation. A clear justification should be provided for undertaking the evaluation at a particular time. The design of the evaluation should be described as precisely as possible.

*Based on UNEG Standards 3.2 and 3.3*

16.3    The GRPP being evaluated should be clearly described, including what it aims to achieve, the means chosen to address the problem(s) identified, the implementation modalities, the financial parameters, and a measure of scope and coverage in terms of beneficiaries.

*Draws on UNEG Standard 3.4*

#### CHECKLIST FOR COMPLETENESS

16.4    The TOR should include the following elements:

- Context for the evaluation, including a stakeholder map
- Purpose, objectives, and scope of the evaluation
- Evaluation criteria (such as relevance, effectiveness, efficiency, and sustainability)
- Key evaluation questions
- Methodology — the chosen approach for data collection and analysis and for participation of stakeholders
- Work plan, including organization, budget, a possible inception report review phase, any criteria for composition of the evaluation team, and details of access to support services or facilities if applicable
- Products and reporting, including the process for reviewing the draft evaluation report before it is finalized

*Based on UNEG Standard 3.2, para. 2*

- Planned dissemination, disclosure and use of evaluation re-sults, and any restrictions related to confidentiality
- Any responsibilities involving follow-up after publication of the final evaluation report.

## MEANING AND CONTENT OF VARIOUS COMPONENTS

Based on UNEG Standard 3.5, paras. 6–8

16.5    The *objectives* of the evaluation should follow from the pur-pose of the evaluation. These should be clear and agreed upon by all stakeholders involved. The *scope* establishes the boundaries of the evaluation, tailoring the objectives and evaluation criteria to the given situation. The scope should also include the explicit *coverage* of the evaluation — the time period, stage of implementation, geographical area, and the dimensions of stakeholder participation being examined — and acknowledge any limits of the evaluation. Evaluations are also oriented by *evaluation questions*, which add more detail to the objec-tives and contribute to defining the scope. The most commonly ap-plied *evaluation criteria* are relevance, effectiveness, efficiency, im-pact, and sustainability. Sometimes, value-for-money and target group satisfaction are assessed as well. Not all criteria are applicable to every evaluation.

## REVISIONS OF THE TERMS OF REFERENCE

Draws on IEG's experience with reviewing GRPPs

16.6    The TOR is a key reference at two stages of the evaluation: in selecting the evaluation team and during implementation. The TOR may be revised between these two stages:

- During the selection process, the TOR provides the essential information for evaluators who are presenting their qualifica-tions and proposals. The TOR defines the factors that delimit the evaluation exercise, providing key inputs to the potential bidders to determine if they can organize a team and process to meet expectations for credibility and quality. Such delimit-ing factors include time-frame, budget, any specifications with respect to the team composition, scope and methodology (if specified), evaluation criteria and questions (if specified).

- During implementation, the TOR determines the deliverables for the contract, as well as any specifications on process or mandatory stages of review that the evaluation must pass through before being finalized.

16.7    Frequently, the TOR provides for the preparation of an incep-tion report, or the commissioners of the evaluation may invite the evaluation team that has been selected to prepare one. The prepara-tion and review of an inception report provide an opportunity to fur-ther specify methodological and organizational aspects of the evalua-tion, including any provisions for needed meetings, interviews, site

visit travel, new data collection, etc. The inception report may also provide an opportunity for the team to point out any limitations they perceive which might affect the credibility and quality of the evaluation if the TOR is followed to the letter — such as inadequate budget, tight time frame, lack of consensus on program or evaluation objectives, lack of an M&E framework, poor data, or lack of provision (budget and/or time) for building participation and consultation of stakeholders into the design of the evaluation.

16.8    In some cases, a preliminary evaluability exercise may be needed. (See paragraph 2.7.) In other cases, it may be necessary to expand the scope of the evaluation to include design and discussion of a logframe or larger M&E framework to provide a foundation for a credible present or future evaluation. Or it may be desirable to increase the level of participation and consultation. The commissioners of the evaluation may also decide that the budget and contract need to be revised, or the deliverables phased differently. In all cases, the TOR should be formally revised and again approved by the commissioner of the evaluation. The evaluation team is then held to the revised TOR, which is published in the final evaluation report.

# 17. Final Reports and Other Evaluation Products

## Principles and Norms

### COVERAGE OF QUALITY EVALUATION REPORTS

17.1    GRPP evaluation reports must include a profile of the GRPP and the key issues or questions addressed, and explain the methodology followed and criteria used (including any limitations or exceptions). They must present in a clear, complete, and balanced way the evidence-based findings; dissident views; and consequent conclusions, recommendations, and lessons. They must have an executive summary that encapsulates the essence of the information contained in the report and facilitates dissemination and distillation of lessons.

Applies DAC Principle X, para. 39, and UNEG Norm 8, para. 8.2, to GRPPs

### PRESENTATION OF FINDINGS AND RECOMMENDATIONS

17.2    Evaluation reports must distinguish between findings and recommendations. Relevant information to support findings should be included in a way that does not compromise sources. To have an effect on decision making, evaluation findings must be presented in a clear and concise way. They should fully reflect the different views and interests of the many parties involved in development cooperation. Easy accessibility is crucial for usefulness.

Based on DAC Principle IV, paras. 18, 20, and 21

### OTHER EVALUATION PRODUCTS

17.3    Evaluation results may be disseminated in several ways apart from the evaluation report itself: annual reports providing a synthesis of findings; abstracts/summaries providing a synopsis of findings; electronic extracts posted on Web sites; and workshops. Ways should be found to present findings in an accessible form as needed for some stakeholder groups, including evaluation products in local languages.

Elaborates on DAC Principle X, para. 41, and GEF Policy, section 5.2, para. 83

## Standards and Guidelines

### SUMMARY STANDARDS FOR EVALUATION REPORTS

17.4    A reader of an evaluation report must be able to understand:

- The purpose of the evaluation
- Exactly what was evaluated
- How the evaluation was designed, conducted, and reviewed, including the degree of stakeholder participation
- Methodology, evaluation questions, evidence found, and conclusions drawn

Based on UNEG Standard 3.16, para. 37

- Recommendations
- Distillation of lessons.

| | |
|---|---|
| DAC Standards 10.1–10.3 | 17.5    "The evaluation report answers all the questions and information needs detailed in the scope of the evaluation. Where this is not possible, reasons and explanations are provided. The analysis is structured with a logical flow. Data and information are presented, analyzed and interpreted systematically. Findings and conclusions are clearly identified and flow logically from the analysis of the data and information. Underlying assumptions are made explicit. Conclusions are substantiated by findings and analysis. Recommendations and lessons learned follow logically from the conclusions." |

OVERVIEW OF RECOMMENDED CONTENTS

| | |
|---|---|
| Based on UNEG Standards 4.1–4.11 | 17.6    The evaluation report should provide a clear and complete description of the following: |

- Reference information on opening pages
- The evaluation process and the TOR (in a preface or annex)
- The purpose and context of the evaluation
- The evaluation objectives and the scope of the evaluation
- The subject being evaluated, namely the GRPP or the relevant subset of its activities, and the context in which it operates
- The logframe, the expected results chain, and the intended impacts of the program, its implementation strategy, and key assumptions
- The role and contributions of the partner organizations, governing bodies, and other stakeholders in GRPP governance and management
- The evaluation methodology applied, including any limitations to the methodology
- The data collection instruments (usually in the annexes)
- The evaluation criteria the evaluators used
- The performance standards or benchmarks used in the evaluation, if any
- The level of stakeholder participation in the evaluation and the rationale for selecting that particular level
- The extent to which the evaluation design included ethical safeguards, where appropriate.

The following paragraphs provide more details on each of these.

REFERENCE INFORMATION ON OPENING PAGES

| | |
|---|---|
| Based on UNEG Standard 4.1, para. 1 | 17.7    The title page and opening pages should provide key basic information, such as the name of the GRPP evaluated; the date; the table of contents, including annexes; the name(s) and organization(s) of the |

evaluators; and the name and address of the organization(s) that commissioned the evaluation.[90]

## PREFACE

17.8    A preface (or annex) should provide key information on the process of the evaluation, the coverage of which should be in accordance with the policy on evaluation and disclosure approved by the governing body. The preface would cover selected topics from the following:

Draws on IEG's experience with reviewing GRPPs

- Who commissioned the evaluation (essential)
- Funding source for the evaluation (essential)
- Who approved the TOR and any peer reviewers, if applicable (recommended)
- Rationale for the level of participation chosen for the evaluation (can be helpful for transparency)
- How the evaluation team was selected (whether competitive or not) and the criteria applied (recommended)
- Who managed the evaluation and to whom the team reported (essential)
- Any conflicts of interest and how they were dealt with (essential)
- Any other organizational information relevant to the evaluation and the degree of independence of the process (recommended)
- The budget (or staff weeks estimated to be required) for the evaluation (recommended, with the agreement of the governing body and any donors involved)
- Actual resources expended for the evaluation (recommended for accountability and transparency, if feasible, and if remuneration information can be kept confidential)
- A description of any changes in the TOR during the evaluation process and the reasons (can be helpful for transparency)
- Information on the process of reviewing the findings, conclusions, and/or final report (can be helpful for transparency)
- Information on planned dissemination of the final report and any other related evaluation products or workshops (recommended).

17.9    The final TOR should always be included in the final report, either in a preface or annex.

17.10    The responses of the commissioners of the evaluation, the governing body, and program management should be proactively

---

90. DAC also emphasizes in Principle III that giving the actual names of the authors increases transparency.

disseminated to the key stakeholders and disclosed to the public. These may be in either a preface or annex of the final evaluation report if they emerge in a timely fashion, but may also be disclosed later through a means other than the final evaluation report, if necessary.

## EXECUTIVE SUMMARY

Elaborates on UNEG Standard 4.2, paras. 2 and 3

17.11    The executive summary should provide a synopsis of the substantive elements of the evaluation report. To facilitate higher readership, the Executive Summary should be brief and should "stand alone." The level of information should provide the uninitiated reader with a clear understanding of what was found and recommended and what has been learned from the evaluation. The executive summary should include:

- The commissioner of the evaluation and the members of evaluation team
- A brief description of the program being evaluated, including financial parameters and main activities
- The origin, context, and present situation of the program
- The purpose of the evaluation, the intended audience of the evaluation report, and the expected use of the evaluation report
- The objectives of the evaluation and key evaluation questions
- A short description of the methodology, including the rationale for the choice of methodology, data sources used, data collection and analysis methods used, and major limitations
- The most important findings and conclusions
- Main recommendations and lessons learned.

## DESCRIPTION OF THE PROGRAM AND CONTEXT

Based on DAC Standards 3.1, 3.2 and 3.3

17.12    The evaluation report provides a description of the context relevant to the GRPP, the development interventions it supports, and their influence on the outcomes and impacts, for example:

- The circumstances surrounding the origin of the program and its maturity [91]
- The objectives of the GRPP, its coverage and scale (in financial terms), its stakeholders, and the range of activities supported
- References to the relevant program policy documents, objectives, and strategies
- Description of the institutional environment and stakeholder participation relevant to the GRPP and its activities

---

91. This treatment should include the raison d'être of the program, namely, why global or regional collective action was deemed necessary or useful, and what additional features the partnership brings to the program.

- Description of the socio-political context within which the GRPP operates and the evaluated activities take place
- Description of the organizational arrangements established for implementation of the development intervention, including the roles of donors and partners
- Expected outcomes and impacts affecting specific target groups.

## EVALUATION CRITERIA AND QUESTIONS

17.13   The criteria used, such as relevance, effectiveness, efficiency, and sustainability, are mentioned, as are any other pertinent benchmarks. The questions asked, as well as any revision to the original questions, are documented in the report so that readers can assess whether the evaluation team has sufficiently assessed them.

Elaborates on DAC Standard 2.4

## EXPLANATION OF METHODOLOGY USED

17.14   "The evaluation report describes and explains the evaluation method and process and discusses validity and reliability. It acknowledges any constraints encountered and their effect on the evaluation, including their effect on the independence of the evaluation. It details the methods and techniques used for data and information collection and processing. The choices are justified and limitations and shortcomings are explained."

DAC Standard 4.1

17.15   The description of the methodology should include:

- Data sources
- Description of data collection methods and analysis (including level of precision required for quantitative methods, value scales, or coding used for qualitative analysis)
- Description of sampling (area and population represented, rationale for selection, mechanics of selection, numbers selected out of potential subjects, limitations to sample)
- Reference indicators and benchmarks, where applicable
- Any deviations from the evaluation plan
- Key limitations.

Based on DAC Standard 4.4 and UNEG Standard 4.9

## INFORMATION SOURCES AND GATHERING PROCEDURES

17.16   "The evaluation report describes the sources of information used (documentation, respondents, literature, etc.) in sufficient detail, so that the adequacy of the information can be assessed. Complete lists of interviewees and documents consulted are included, to the extent that this does not conflict with the privacy and confidentiality of participants."

DAC Standards 5.1 and 5.2

| UNEG Standard 4.12, para. 22 | 17.17 "Data [do] not need to be presented in full; only data that support a finding needs to be given, and full data can be put in an annex." |

### DESCRIPTION OF PARTICIPATION AND CONSULTATION OF STAKEHOLDERS

| DAC Standard 4.3 and UNEG Standard 4.10, para. 17 | 17.18 "The evaluation report indicates the stakeholders consulted and the criteria for their selection and describes stakeholders' participation. If less than the full range of stakeholders was consulted or invited to participate, the methods and reasons for selection of particular stakeholders are described." |

### INTERVENTION LOGIC AS RELATED TO FINDINGS

| Based on DAC Standard 2.2 and UNEG Standard 4.6, para. 10 | 17.19 The evaluation report should briefly describe and assess the intervention logic and distinguish between findings at the different stages of the results chain: inputs, activities, outputs, reach, outcomes, and impacts. The report should also provide a brief overall assessment of the intervention logic. Any value judgments should be presented transparently. |

### FINDINGS AND CONCLUSIONS

| Based on DAC Standard 9.1 and UNEG Standards 4.6, 4.12, and 4.14 | 17.20 The evaluation findings should be relevant to the GRPP and to the purpose of the evaluation. They should cover all the evaluation objectives, showing a clear line of evidence to support the conclusions. The evaluators should explain the evaluation criteria that were used. Measurement of inputs, the progress of activities, outputs, and outcomes, and impacts should be presented to the extent possible, with reference to appropriate benchmarks (or an appropriate rationale given as to why these were not measured). Findings regarding inputs and activities should be distinguished clearly from outputs, outcomes, and impacts. Outcomes and impacts should include any unintended effects, whether beneficial or harmful. Additionally, any multiplier or downstream effects of GRPP activities should be included. Any discrepancies between the planned and actual implementation of the GRPP activities should be explained with reference to factors, including external factors, which were especially constraining or enabling. |

| UNEG Standard 4.15, para. 29 | 17.21 "Conclusions must focus on issues of significance to the program as determined by the evaluation objectives and the key evaluation questions. Simple conclusions that are already well known and obvious are not useful, and should be avoided." |

### RECOMMENDATIONS AND LESSONS LEARNED

| Based on DAC Standard 9.3 and UNEG Standard 4.16 | 17.22 Recommendations and lessons learned should be relevant and targeted to the intended users. Recommendations should be the logical implications of the findings and conclusions and be firmly based |

on evidence and analysis. They should be realistic: the priorities, responsibilities for action, and provisional time-frame for action should be clear to the extent possible.

17.23    A good evaluation report should correctly identify lessons that stem logically from the findings, present an analysis of how these can be applied to different contexts and/or different sectors, and take into account evidential limitations such as generalizing from single point observations. But not all evaluations generate lessons. Lessons should only be drawn if they represent a contribution to general knowledge.

Based on UNEG Standard 4.17, paras. 33 and 34

### ANNEXES

17.24    Additional supplementary information to the evaluation that should be included in annexes includes:

Based on UNEG Standard 4.18

- A list of abbreviations, if not included in the early pages
- The final TOR for the evaluation (and earlier versions if appropriate, if not in the preface)
- Program logical framework
- List of persons interviewed (if confidentiality allows) and sites visited
- Data collection instruments (copies of questionnaires, surveys, etc.)
- Documents consulted and references.

# REFERENCES

African Evaluation Association. 2002. *Draft African Evaluation Guidelines*. http://66.201.108.198/afrea/content/index.cfm?navID=5&itemID=204

American Evaluation Association. 2004. *Guiding Principles for Evaluators*. http://www.eval.org/Publications/GuidingPrinciples.asp

Asian Development Bank. 2006. *Impact Evaluation: Methodological and Operational Issues*. Manila.

Barrett, Scott. 2006. "Making International Cooperation Pay: Financing as a Strategic Initiative." In Inge Kaul and Pedro Conceição, eds., *The New Public Finance: Responding to Global Challenges*. New York: Oxford University Press.

Bovaird, Tony, and Elke Loffler. 2001. "Emerging Trends in Public Management and Governance." *Bristol Business School Teaching and Research Review* 5, Winter.

Canadian Evaluation Society. No date. *Guidelines for Ethical Conduct*. http://www.evaluationcanada.ca/site.cgi?s=1

Council on Foundations. 2003. *Evaluation Approaches and Methods*. http://www.cof.org/Learn/content.cfm?ItemNumber=1379

Davis, Michael, and Andrew Stark, eds. 2001. *Conflict of Interest in the Professions*. Oxford: Oxford University Press.

Evaluation Cooperation Group. No date. *Good Practice Standards*.

Evaluation Cooperation Group. 2004. *Template for Assessing the Independence of Evaluation Organizations*. http://www.ecgnet.org/docs/ecg.doc

Global Corporate Governance Forum. 2005. *Developing Corporate Governance Codes of Best Practice,* Toolkit #2. http://www.gcgf.org/ifcext/cgf.nsf/AttachmentsByTitle/Toolkit2-read.pdf/$FILE/Toolkit2-read.pdf

Global Environment Facility Evaluation Office. 2006. *The GEF Monitoring and Evaluation Policy*. http://www.gefweb.org/MonitoringandEvaluation/MEPoliciesProcedures/documents/Policies_and_Guidelines-Tools_and_Guidelines-New_ME_Policy-020306.pdf

High-Level Forum. 2005. *Paris Declaration on Aid Effectiveness: Ownership, Harmonization, Alignment, Results and Mutual Accountability*, March 2, 2005. http://www1.worldbank.org/harmonization/Paris/FINALPARISDECLARATION.pdf

Independent Evaluation Group – World Bank. Forthcoming. *Regional Development Programs: An Independent Evaluation of World Bank Support*. Washington, DC: World Bank.

Independent Evaluation Group – World Bank. 2006. "Guidelines for Global Program Reviews."
http://www.worldbank.org/ieg/grpp/docs/gprguidelines.pdf

Independent Evaluation Group – World Bank, in collaboration with the Development Assistance Committee Secretariat. 2006. *Impact Evaluation: An Overview and Some Issues for Discussion*. Washington, DC.

Independent Evaluation Group – World Bank, in collaboration with the Poverty Analysis, Monitoring, and Impact Evaluation Thematic Group (PREM Network). 2006. *Conducting Quality Impact Evaluations under Budget, Time, and Data Constraints*. Washington, DC.

Institute of Chartered Secretaries and Administrators International. No date. *Principles of Corporate Governance for Charities*. 16 Park Crescent, London W1B 1AH, United Kingdom.

International Task Force on Global Public Goods. 2006. *Meeting Global Challenges: International Cooperation in the National Interest.*
http://www.gpgtaskforce.org/bazment.aspx?page_id=174

Joint Committee on Standards for Educational Evaluation. 1994. *The Program Evaluation Standards*, 2nd edition. Newbury Park, CA: Sage Publications.
http://www.wmich.edu/evalctr/jc/

Kaul, Inge, Isabelle Grunberg, and Marc Stern, eds. 1999. *Global Public Goods: International Cooperation in the 21st Century*. New York: Oxford University Press.

Kaul, Inge, Pedro Conceição, Katell Le Goulven, and Ronald U. Mendoza, eds. 2003. *Providing Global Public Goods: Managing Globalization*. New York: Oxford University Press.

Kaul, Inge, and Pedro Conceição, eds. 2006. *The New Public Finance: Responding to Global Challenges*. New York: Oxford University Press.

Kusek, Jody Zall, and Ray C. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington DC: The World Bank.

Netherlands Ministry of Finance. 2000. "Government Governance: Corporate Governance in the Public Sector, Why and How?" Paper presented at the 9th FEE Public Sector Conference, November 2-4.

Organisation for Economic Cooperation and Development. 2004. *Principles of Corporate Governance* (rev. ed).
http://www.oecd.org/document/49/0,2340,en_2649_37439_315308 65_1_1_1_37439,00.html

OECD, Development Assistance Committee. 2006a. *DAC Evaluation Quality Standards*. DAC Evaluation Network.

OECD, Development Assistance Committee. 2006b. *Guidance for Managing Joint Evaluations*. DAC Evaluation Series.
http://www.oecd.org/dataoecd/29/28/37512030.pdf

OECD, Development Assistance Committee. 2005. *Joint Evaluations: Recent Experiences, Lessons Learned and Options for the Future*. DAC Evaluation Network Working Paper.

OECD, Development Assistance Committee. 2002. *Glossary of Key Terms in Evaluation and Results Based Management*. Paris: Organization for Economic Cooperation and Development.

OECD, Development Assistance Committee. 1998. *Review of DAC Principles for Evaluation of Development Assistance*. Paris: Organization for Economic Cooperation and Development.

OECD, Development Assistance Committee. 1991. *Principles for Evaluation of Development Assistance*. Paris: Organization for Economic Cooperation and Development.

Operations Evaluation Department.[92] 2005. *Influential Evaluations: Detailed Case Studies*. http://www.worldbank.org/oed/influential_evaluations/

Operations Evaluation Department. 2004. *Addressing the Challenges of Globalization: An Independent Evaluation of the World Bank's Approach to Global Programs*, Phase 2 Report. Washington DC: World Bank. http://www.worldbank.org/ieg/grpp/browse_all.html#completed

Operations Evaluation Department. 2002. *The World Bank's Approach to Global Programs: An Independent Evaluation*, Phase 1 Report. Washington DC: World Bank. http://www.worldbank.org/ieg/grpp/browse_all.html#completed

Reinicke, Wolfgang, and Francis Deng. 2000. *Critical Choices: The United Nations, Networks, and the Future of Global Governance*. Ottawa: International Development Research Centre for the UN Vision Project on Global Public Policy.

Rozman, Rudi. 2000. "The Organizational Function of Governance: Development, Problems, and Possible Changes." *Management* 5(2), pp. 94-110.

Schivo-Campo, Salvatore. 1999. "'Performance' in the Public Sector," *Asian Journal of Political Science* 7(2).

Tricker, Robert I. 1998. *Pocket Director*. London: Economist Books.

United Kingdom Audit Commission. October 2003. *Corporate Governance: Improvement and Trust in Local Public Services*. http://www.audit-commission.gov.uk/reports/NATIONAL-REPORT.asp?CategoryID=&ProdID=7374209E-8060-4218-B4B2-2CCF2FD490C5

United Nations Evaluation Group. 2005a. *Norms for Evaluation in the UN System*. Adopted by the United Nations Evaluation Group (UNEG) at its Annual Meeting in Rome, April 2005. http://www.uneval.org/docs/ACFFC9F.pdf

92. OED formally changed its name to the Independent Evaluation Group of the World Bank in December 2005.

United Nations Evaluation Group. 2005b. *Standards for Evaluation in the UN System*. New York: United Nations. http://www.uneval.org/docs/ACFFCA1.pdf

White, Howard, Shampa Sinha, and Ann Flanagan. 2006. "A Review of the State of Impact Evaluation." Paper presented at the International Workshop on Impact Evaluation for Development, Paris, November 15, 2006.

World Bank Development Grant Facility. 2003. Technical Note on Independent Evaluation: Principles, Guidelines and Good Practice. Washington DC: World Bank. http://siteresources.worldbank.org/INTDGF/Resources/Evaluation&LearningNote.pdf

World Bank. 2004. *The LogFrame Handbook: A Logical Framework Approach to Project Cycle Management*. Washington, DC: World Bank. http://www1.worldbank.org/education/adultoutreach/designing.logframe.asp