World Bank Independent Evaluation Group

Guidance on Program Evaluation

Providing an Independent Opinion on the Effectiveness of the Program

Independent program evaluation. Programs receiving DGF funding of US$300,000 or more over the life of the program must undertake an external evaluation every three to five years that is independent of the program's management. Some flexibility is allowed with respect to the timing based on program circumstances and the preferences of partners. The DGF encourages programs to follow the IEG/DAC *Sourcebook of Indicative Principles and Standards for Evaluating Global and Regional Partnership Programs* in terms of both process and substance. The programs' governing bodies should commission the evaluation, approve the evaluation terms of reference, select the evaluation team (ideally using competitive methods), and receive the final evaluation report – or establish an oversight subcommittee for these purposes. Both CFPTP and IEG typically provide feedback and comments on the draft terms of reference for upcoming evaluations. More broadly, IEG continues to make progress on two fronts towards improving the quality of evaluations of GRPPs in order to enhance the programs' relevance and effectiveness:

- First, IEG will complete four Global Program Reviews (GPRs) on the following programs in FY10: the *Stop TB Partnership*, the *Global Water Partnership*, the *International Assessment of Agricultural Science and Technology for Development*, and the *Meso-American Biological Corridor*. Each of these reviews assesses the independence and quality of the most recent evaluation of the program; provides a second opinion on the effectiveness of the program; assesses the performance of the Bank as a partner in the program; and draws lessons for the Bank's engagement in GRPPs more generally. Next year, IEG will complete, among others, a Global Program Review of the *Global Fund to Fight AIDS, TB and Malaria*, which is the largest global program in which the Bank is involved. This GPR will have a particular focus on the Bank's engagement with the Global Fund at the country level, given the major role of the Global Fund in financing disease-control investments at the country level and therefore the potential for competition or collaboration between Global Fund and Bank-supported activities at the country level. (IEG's framework for assessing the overall effectiveness of a GRPP and the Bank's role as a partner in the program are summarized below.)

- Second, IEG continues to play a leading role in developing consensus standards for evaluating GRPPs, under the auspices of the OECD/DAC Network on Development Evaluation. Following the publication of the *Sourcebook* in January 2007, IEG has been working on a second, companion document of Good Practice Guidelines and Examples. For this *Guidebook*, IEG has conducted a systematic review of a representative sample of 60 recent evaluations of GRPPs. Each chapter in the Guidebook will provide overall guidance in addressing a particular evaluation criterion (relevance, effectiveness, etc.), sample evaluation questions in relation to 5–6 evaluation issues under each criterion, common approaches and methodologies to addressing each issue, and some good-practice examples taken from the 60 evaluations reviewed. The Guidebook will be reviewed extensively by a diversity of stakeholders before being issued sometime in FY11.

Note: The following framework for assessing the overall effectiveness of a GRPP has been designed to cover the wide range of such programs in which the World Bank is involved, encompassing policy and knowledge networks, technical assistance programs, and investment programs. It is not expected that each GPR should cover every question in this table in detail. At a minimum, each GPR is expected to cover the first four criteria in the following table: (a) relevance, (b) efficacy, (c) efficiency, and (d) governance and management. A GPR may also cover (e) resource mobilization and financial management and (f) sustainability, risk, and strategies for devolution or exit if the latter are important issues for the program at the time of GPR, and if there is sufficient information available on which to base an independent opinion.

Evaluation Criteria and Questions
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.
1. Supply-side relevance — the existence of an international consensus that global/regional collective action is required. To what extent does the program reflect an international consensus on the need for action, on the definition of the problem being addressed, on priorities, and on strategies for action? Is the original consensus that led to the creation of the program still present? Is the program still needed to address specific global/regional public concerns? Take into account the origin of the program in answering these questions: • Is the program formally responsible for implementing an international convention? • Did the program arise out of an international conference? • Is the program facilitating the implementation of formal standards and approaches? • Is the program primarily donor-driven? Did donors establish the program with little consultation with developing countries? • Is the program primarily Bank-driven? Did the World Bank found the program and then seek other partners?
2. Demand-side relevance — alignment with beneficiary needs, priorities, and strategies. To what extent are the objectives consistent with the needs, priorities, and strategies of beneficiary countries as articulated in the countries' own PRSPs, and in donors' strategies such as the World Bank CASs, and the UN Development Assistance Frameworks? To what extent has the voice of developing and transition countries been expressed in the international consensus underlying the program?
3. Vertical relevance — consistency with the subsidiarity principle. To what extent are the activities of the program being carried out at the most appropriate level — global, regional, national, or local — in terms of efficiency and responsiveness to the needs of beneficiaries? To what extent are the activities of the program competing with or substituting for activities that individual donors or countries could do more efficiently by themselves? Pay particular attention to those programs that, on the face of it, are primarily supporting the provision of national or local public goods.
4. Horizontal relevance — the absence of alternative sources of supply. What is the comparative advantage, value added, or core competency of the program relative to other GRPPs with similar or complementary objectives? To what extent is the program providing additional funding, advocacy, or technical capacity that is otherwise unavailable to meet the program's objectives? To what extent are the good and services being provided by the program in the nature of public goods? Are there alternative ways of providing these goods and services, such as by the private sector under regular market conditions?

Evaluation Criteria and Questions

5. Relevance of the design of the program

To what extent are the strategies and priority activities of the program appropriate for achieving its objectives?

What are the major activities of the program:

- Policy and knowledge networking?
- Financing country and local-level technical assistance?
- Financing investments to deliver national, regional, or global public goods? (See Annex Table 4.)

Has the program articulated an expected results chain or logical framework, along with assumptions that relate the progress of activities with the achievement of the objectives? Does the results chain identify the extent to which the achievement of the objectives depends on the effective functioning of bureaucracies, markets, or collectivities? If so, to what extent are these assumptions valid?

For programs providing global or regional public goods, is the design of the program consistent with the way in which the individual efforts of the partners contribute to the collective outcome for the program as a whole — whether “best shot”, “summation”, or “weakest link?”

Efficacy: The extent to which the program has achieved, or is expected to achieve, its objectives, taking into account their relative importance.

6. Achievement of objectives

To what extent have the stated objectives of the program been achieved, or has satisfactory progress been made towards achieving these objectives?

To what extent are there implicit objectives that are well understood and agreed upon by the partners and to which the program should also be held accountable?

To what extent are there any positive, unintended outcomes of the program that have been convincingly document?

To what extent have these assessments by the program or the evaluation been evidence-based?

7. Progress of activities, outputs, and outcomes.

To what extent has the program or the evaluation measured the progress of activities, outputs, and outcomes?

How did the program or the evaluation aggregate its outputs and outcomes at all levels — global, regional, national, and local — to provide an overall summary of its results?

To what extent have factors such as changes in the location of the program, its legal structure, or governance processes affected the outputs and outcomes of the program?

To what extent have there been outcomes that can be uniquely attributed to the partnership itself — such as the scale of or joint activities made possible by its organizational setup as a GRPP, or its institutional linkages to a host organization?

8. Linkages to country or local-level activities.

To what extent has the program established effective operational linkages with country-level activities, taking into account that:

- The desired nature of these linkages varies according to the objectives, design, and implementation of each program?
- Positive outcomes at the country or local level are generally a joint product of both global/regional and county-level activities?

Evaluation Criteria and Questions

Efficiency or cost-effectiveness:

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.

Cost-effectiveness — the extent to which the program has achieved or is expected to achieve its results at a lower cost compared with alternatives.

9. Efficiency

To what extent is it possible to place a monetary value on the benefits arising from the activities of the program?

To what extent has the program or the evaluation conducted impact evaluations of representative program activities?

To what extent has the program or the evaluation analyzed the program's costs in broad categories (such as overhead vs. activity costs), and categorized the program's activities and associated benefits, even if these cannot be valued in monetary terms?

10. Cost-effectiveness

To what extent is the program measuring up against its own business plans:

- Has the program cost more or less than planned? How did it measure up against its own costing schedule?
- Have there been any obvious cases of inefficiency or wasted resources?

To what extent is the program delivering its activities cost-effectively in comparison with alternatives:

- How do actual costs compare with benchmarks from similar programs or activities?
- Are the overhead costs of governing and managing the program reasonable and appropriate in relation to the objectives and activities of the program?

How does the program compare with traditional development assistance programs:

- For beneficiary countries, has receiving the development assistance through the GRPP increased the transactions costs compared with traditional development assistance programs?
- For donors, has delivering the development assistance through the GRPP reduced donor costs by harmonizing efforts among donors or by reducing overlapping work (such as through joint supervision, monitoring and evaluation)?

Governance and management:

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.

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. Whereas governance is concerned with "doing the right thing," management is concerned with "doing things right."

11. Compliance with generally accepted principles of good governance.

To what extent are the governance and management structures and processes well articulated and working well to bring about legitimate and effective governance and management?

To what extent do governance and management practices comply with the following seven principles:

- Legitimacy — 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?
- Accountability — 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?

Evaluation Criteria and Questions
• Responsibility — 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? • Fairness — the extent to which partners and participants, similarly situated, have equal opportunity to influence the program and to receive benefits from the program? • Transparency — the extent to which a program's decision making, reporting, and evaluation processes are open and freely available to the general public? • Efficiency — the extent to which the governance and management structures enhance efficiency or cost-effectiveness in the allocation and use of the program's resources? • Probity — 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?
12. Partnerships and participation To what extent has the program identified a complete list of stakeholders, or "stakeholder map", including the agreed-upon or perceived roles and responsibilities of the categories of stakeholders identified? To what extent is this a routine programmatic function, updated regularly, and transparently available? Has the program adopted primarily a shareholder model of governance (in which membership on the governing body is limited to financial and other contributors), or a stakeholder model (in which membership also includes non-contributors)? To what extent, if any, is the program's legitimacy being sacrificed in order to achieve greater efficiency, or vice-versa?
13. Programs located in host organizations To what extent is the location of the program in the Bank or other partner organization adversely affecting the governance, management, or other aspects of the program, such as compliance with the principles of transparency and fairness? For which functions is the program manager accountable to the host organization and the governing body of the program, respectively? Are conflicts of interest being managed appropriately? To what extent does the host organization play such a dominant role in the program, thereby reducing the incentives of other partners to participate effectively, or reducing the ability of the host organization to look at the weaknesses of the program objectively?
Resource mobilization and financial management: **Resource mobilization** — the processes by which resources are solicited by a program and provided by donors and partners. **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. In cases where funds accumulate over time, this would also include the management of the cash and investment portfolio.

Evaluation Criteria and Questions
14. Resource mobilization To what extent has the program succeeded in raising financial resources commensurate with its objectives? And from what sources — the Bank, bilateral donors, foundations, etc.? To what extent has the program succeeded in diversifying its funding beyond a small number of donors? To what extent are the sources of funding for the program (including donor restrictions on the use of resources) affecting, positively or negatively: • The strategic focus of the program? • The outputs and outcomes of the program? • The governance and management of the program? • The sustainability of the program?
15. Financial management Are there any issues that have emerged during the course of the review in relation to: • The quality of financial management and accounting? • The methods, criteria, and processes for allocating funds among different activities of the program? • Financial management during the early stages of the program?
Sustainability, risk, and strategy for devolution or exit: **Sustainability** — When 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 applied to a **program** itself, the extent to which the organization or program is likely to continue its operational activities over time. **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.
16. Sustainability of the benefits of the program's activities What is the risk, at the time of evaluation, that the development outcomes (or expected outcomes) of the program will not be maintained (or realized)? This depends on (a) the likelihood that some changes may occur that are detrimental to maintaining or realizing the expected outcomes, and (b) the affect on the expected outcomes if some or all of these changes actually materialize?
17. Sustainability of the program This will depend on a number of factors, such as the continued legitimacy of the program, its financial stability, its continuity of effective management, and its ability to withstand changing market or other conditions. To what extent is there still a sufficient convergence or accommodation of interests among the major partners to sustain the program financially? To what extent has the program developed institutional capacity such as performance-based management, personnel policies, learning programs, and knowledge management that help to sustain a program? In what areas could the program improve in order to enhance its sustainability, such as better marketing of the program's achievements in order to sustain its reputation?

Evaluation Criteria and Questions
18. Prospects for continuation and strategies for devolution or exit To what extent should the program 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, in the light of the previous evaluation findings with respect to relevance, efficacy, efficiency, and sustainability: • Reinventing the program with the same governance? • Phasing out the program? • Continuing country or local-level activities with or without devolution of implementation? • Seeking alternative financing arrangements, such as revenue-generation, or self-financing to reduce dependency on external sources? • "Spinning off" from the host organization?

Assessing the Bank's Performance as a Partner in the Program

Evaluation Questions
1. Comparative advantage at the global/regional level. To what extent is the Bank playing up to its comparative advantages at the global/regional level — its global mandate and reach and convening power? To what extent is the Bank's presence as a partner in the program catalyzing other resources and partners for the program?
2. Comparative advantage at the country level. To what extent is the Bank contributing multi-sector capacity, analytical expertise, and country-level presence and knowledge to the program? To what extent has the Bank's country operations established linkages to the GRPP, where appropriate, to enhance the effectiveness of both?
3. Oversight. To what extent is the Bank exercising effective and independent oversight of its involvement in the program, as appropriate, whether the program is housed in the Bank or externally managed? To what extent is the Bank's oversight independent of the management of the program? To what extent does the Bank's representative on the governing body have a clear terms of reference?
4. Risks and risk management. To what extent have the risks associated with the program been identified and are being effectively managed? For example, IEG identified the following risks in its global review: • Bank bears a disproportionate share of responsibility for governing and managing in-house programs? • Confusion at the country level between global program activities, Bank activities, and Borrower activities? • Representation of NGOs and the commercial private sector on program governing bodies? • Unclear role and application of Bank's safeguards? • Trust-funded consultants and seconded staff representing the Bank on some program governing bodies?
5. Disengagement strategy. To what extent is the Bank engaged at the appropriate level in relation to the Bank's new strategic framework: • Watching brief? • Research and knowledge exchange? • Policy or advocacy network? • Operational platform? To what extent is the Bank facilitating an effective, flexible, and transparent disengagement strategy for the program, in relation to the Bank's objectives for its involvement in the program: • The program declares "mission accomplished" and closes? • The program continues and the Bank withdraws from all aspects of its participation? • The program continues and the Bank remains engaged, but the degree of the Bank's engagement in some or all aspects (such as financing) declines over time?

Assessing the Bank's Performance as a Partner in the Program

Evaluation Questions
6. Comparative advantage at the global/regional level. To what extent is the Bank playing up to its comparative advantages at the global/regional level — its global mandate and reach and convening power? To what extent is the Bank's presence as a partner in the program catalyzing other resources and partners for the program?
7. Comparative advantage at the country level. To what extent is the Bank contributing multi-sector capacity, analytical expertise, and country-level presence and knowledge to the program? To what extent has the Bank's country operations established linkages to the GRPP, where appropriate, to enhance the effectiveness of both?
8. Oversight. To what extent is the Bank exercising effective and independent oversight of its involvement in the program, as appropriate, whether the program is housed in the Bank or externally managed? To what extent is the Bank's oversight independent of the management of the program? To what extent does the Bank's representative on the governing body have a clear terms of reference?
9. Risks and risk management. To what extent have the risks associated with the program been identified and are being effectively managed? For example, IEG identified the following risks in its global review: • Bank bears a disproportionate share of responsibility for governing and managing in-house programs? • Confusion at the country level between global program activities, Bank activities, and Borrower activities? • Representation of NGOs and the commercial private sector on program governing bodies? • Unclear role and application of Bank's safeguards? • Trust-funded consultants and seconded staff representing the Bank on some program governing bodies?
10. Disengagement strategy. To what extent is the Bank engaged at the appropriate level in relation to the Bank's new strategic framework: • Watching brief? • Research and knowledge exchange? • Policy or advocacy network? • Operational platform? To what extent is the Bank facilitating an effective, flexible, and transparent disengagement strategy for the program, in relation to the Bank's objectives for its involvement in the program: • The program declares "mission accomplished" and closes? • The program continues and the Bank withdraws from all aspects of its participation? • The program continues and the Bank remains engaged, but the degree of the Bank's engagement in some or all aspects (such as financing) declines over time?