

Partnering for Evaluations: Seven guiding principles for evaluating partnership programs An IEG note for the DAC

By Anna Aghumian, Evaluation officer, IEG

November 10, 2014

Development partners need to collaborate better to improve the authorizing environment and governance of partnership program evaluations.

Overall, when evaluating GRPPs and MDTFs, there is more shared understanding around the evaluation criteria than around the standards that should guide the evaluation process, follow up and use. Partners need to make more effort to agree upon and reinforce such standards at all stages of evaluation in each and every partnership program: starting from synchronizing their evaluation plans with other partners, taking steps to ensure independence of the evaluation, to making sure the recommendations are evidence-based and actionable.

Introduction

1. Global and regional partnership programs (GRPPs)¹ and other multi-donor trust funds (MDTFs)² designed to deliver development assistance at global, regional and national levels proliferate over the last 15 years. The World Bank alone is involved in about 200 partnership programs and MDTFs. Evaluation is one of the main tools development partners, donors, and stakeholders increasingly use to assess the effectiveness of these partnership programs. Independent, high quality evaluations can be a powerful tool for the stakeholders to hold the program accountable for results and to learn what works and what doesn't. Yet, increasing number of evaluations raises the questions about their credibility and quality, and the extent to which they serve the purpose.

2. Achieving shared understanding of what are the key principles and criteria for evaluating these programs, what constitutes high quality evaluation and how to plan such evaluations is important. Adopting agreed common standards for evaluating these programs would improve the quality, utilization and impact of the evaluations. It is even more important to put such shared understanding in action by having all the development partners to encourage and to cooperate in utilizing those standards systematically.

¹ GRPPs are programmatic partnerships in which a) the partners dedicate resources towards achieving agreed objectives over time, b) conduct activities that are global, regional, or multi-country in scope, 3) partners establish a new organization with shared governance and management unit to deliver these activities.

² MDTFs, like GRPPs are (a) programmatic in nature and (b) conducting activities that are global, regional, or simply multi-country in scope, but (c) do not have a governing body and the program manager reports only to his/her line manager, and ultimately to the board of the host organization.

3. The DAC Network on Development Evaluation and the Evaluation Cooperation Group have taken some role in promoting quality evaluations, disseminating good evaluation practices and encouraging better collaboration among partners in development evaluations in the past.

4. This note aims to reinvigorate those efforts in two ways: by highlighting challenges peculiar to evaluating partnership programs and by offering guidance on selected aspects of evaluating partnership programs.

The note draws on two sources of evidence (attachment 1):

- IEG reviews of 24 GRPPs conducted between 2006 and 2014. These reviews applied an evaluation framework based on OECD/DAC and UNEG’s evaluation standards that has been adapted to evaluating GRPPs.
- 20 recent external evaluations of (mostly) partnership programs carried out in 2011-2014. Where possible the evaluative evidence is extended to MDTFs.³

The note offers seven guiding principles for evaluating partnership programs. These principles cover the areas that are especially weak in partnership program evaluations:

1. To improve the authorizing environment, make sure there is a mutually agreed evaluation policy.
2. To improve credibility, ensure evaluation independence.
3. Invest time in planning the evaluation.
4. To have quality evaluation, choose criteria that fit the purpose.
5. To guarantee credibility and ownership, make sure evaluation is transparent and key stakeholders are consulted.
6. Make sure that recommendations are agreed on and followed up.
7. Plan dissemination in advance.

What makes partnership programs different when it comes to evaluating them? Partnership programs present four main challenges for evaluation:

- They are large and have broad and more complex objectives.
- They often don’t have a fixed end date and therefore their objectives and results frameworks continually evolve.
- They are programmatic – that is, they operate at multiple levels and often support more than one type of activity; therefore establishing monitoring frameworks can be challenging.
- Partners may have different development objectives and different accountability requirements.

³ While so far all the evaluative evidence is based on evaluating GRPPs, many of these criteria and questions are also valid for evaluating multi-donor programmatic trust funds. The note will extend the findings to MDTFs whenever applicable.

5. There are accepted standards of evaluating development assistance that program evaluations apply.⁴ Generally, these standards designed to evaluate development interventions work well when evaluating partnership programs. However, the distinct features of partnership programs create challenges the commissioners of evaluations and the evaluators have to be mindful about. Some key challenges are not in the principles and criteria per se but rather in the expectations and behavior of the key partners that may be reflecting the constraints of their own institutions. The solution often is not in developing new standards but paying more attention to the authorizing environment and governance of evaluations of GRPPs and MDTFs and having a stronger cooperation around those issues.

1. To improve the authorizing environment, make sure there is a mutually agreed evaluation policy

6. Program evaluations, in general, have two objectives:

- to provide accountability and learning to funders and other stakeholders about current progress or past accomplishments and
- to inform key decisions about the program's future.

7. Most evaluations can simultaneously meet both accountability and learning objectives, since partners generally time evaluations in accordance with the program's phases or funding decisions. But, not always. A general problem in partnership programs is that a decision to evaluate a program may originate from the program stakeholders, host organization, or donors anytime. If, for example, different donors require separate reports, or if one donor's funding cycle requires an evaluation that's out of sync with the program's phasing, the result can be duplicative. The World Bank- hosted Energy Sector Management Assistance Program (ESMAP) program, for example, has undergone two reviews: one comprehensive external evaluation agreed by ESMAP's Consultative Group of Donors, and a second commissioned by one of its major donors, Australia at almost the same time and with very similar evaluation objectives.⁵

8. Preventing such overlaps can start with an evaluation policy setting out expectations about why, when, and how to use evaluation. The evaluation policy is essentially for establishing an authorizing environment for conducting an evaluation. It will ensure that the evaluation is timely and useful for the partnership program, will reconcile the different accountability requirements the donors and host agency may have and clarify the roles and responsibilities of all the participants. Such a policy should at least address three questions:

- When and why will evaluations be conducted – at fixed intervals, or at stages of program activities completed?

⁴ These include [OECD/DAC, 1991, Principles for Evaluation of Development Assistance](#); [OECD-DAC, 2010, Quality Standards for Development Evaluation](#); [UNEG's standards for Evaluation](#); [IEG/DAC Global Evaluations Sourcebook](#); Evaluation Cooperation Group, 2012, [ECG Big Book of Good Practice Standards good practice standards](#).

⁵ MDTFs sometimes fall into another extreme. Asia Sustainable and Alternative Energy Program (ASTAE), a smaller scale program closely associated with ESMAP that supports energy access operations in East Asia region has not have any external evaluation since its inception in 1992.

- Will all evaluations be commissioned and funded by donors as a group? Or will individual donors or the host organization have the prerogative to conduct their own evaluations?
- If the program carries out self-evaluations, will they be validated and accepted by all the stakeholders?

Box.1. Reconciling accountability and learning objectives of an independent evaluation

The stakeholders of the Roll Back Malaria program have agreed to synchronize their evaluation objectives.

The evaluation fulfilled a requirement of the World Bank, one of the lead donors to the program, to conduct an independent, external evaluation every 3 to 5 years. Stakeholders agreed that, in addition to fulfilling this requirement, the evaluation presents an opportunity to look forward and to inform the development of the second Global Malaria Action Plan.

Thus, the program's recent evaluation was meant to i) understand the contributions of the program toward Global Malaria Action Plan (GMAP) targets and ii) inform decisions about how to address gaps in the Plan's final years, and iii) develop the second GMAP and position the program in post-2015.

Source: Boston University Center for Global Health and Development, External Evaluation of the Roll Back Malaria Partnership 2009-2013, December 2013

9. Many large partnership programs and funds have developed formal evaluation policies (see for example, [CGIAR](#), [GEF](#), and [GAVI](#) evaluation policies). GRPPs with governing bodies should follow these examples. A few, such as GEF and CGIAR, have also established independent evaluation offices.

10. While not every program needs to develop detailed formal evaluation policies, partners and donors do need to establish commonly agreed evaluation rules for the programs they sponsor. For instance:

- For multi-donor programs without governing bodies hosted in larger organizations host's evaluation policies may apply. The World Bank, for instance, requires an independent evaluation of all GRPPs it funds through its own grant making facility-DGF. It also requires carrying out an independent evaluation for all programmatic trust-funds housed in the Bank that have disbursed more than \$5million over their lifetime. The evaluations should be conducted in accordance with the principles and standards laid out in the [IEG/DAC Sourcebook for Evaluating GRPPs and OECD/DAC criteria](#).⁶
- Evaluation policies of MDTFs can also be set in the funding agreements. Donors of the Health Research and Innovation Trust Fund Program, DFID and Norway, and the World

⁶ However, the application of this rule is uneven at the World Bank. Therefore, other partners also need to make sure there is an agreed evaluation policy.

Bank, where the program is located,⁷ have simply put program’s evaluation policy in the trust fund agreement. The latter mandates periodic, donor initiated independent external evaluations to be undertaken in 2012, 2016, and final evaluation in 2022.

2. **To improve credibility, ensure evaluation independence**

11. Generally if an evaluation is for accountability purposes it is important that those who commission and receive the draft report do not excessively influence the evaluation at any stage. There is ample guidance on how to commission independent evaluation but the practice is diverse.⁸

12. How much independence is “enough independent”? The answer varies considerably across the multilateral development institutions and bilateral donors. Varying standards of independence among partners has a direct impact on GRPP evaluations that are either hosted or funded by these partners. This is a source of confusion itself. As a starting point partners need to develop a shared understanding about what are the acceptable standards for independent evaluation.

13. As for partnership programs, the spectrum is wide and depending on the type of a program, there are steps that can help ensure that the assessment is unbiased and independent from any influence.

- In large GRPPs, like GEF and CGIAR that have an internal independent evaluation function, the evaluation is commissioned by their governing body and independent evaluation units report to it;
- In GRPPs without internal evaluation capacity, an independent evaluation can be commissioned by the partnership’s governing body or a steering committee the governing body appoints to commission and oversee the evaluation.
- Many multi-donor trust funds without governing bodies are hosted in other organizations and support the implementation of host’s programs (such as the World Bank or World Health Organization). In these programs, the donors can commission the evaluation themselves and fund it out of their own resources to maintain the independence of evaluation process.
- A program can also opt for commissioning evaluation to an external panel or an advisory group formed for that purpose only. For programs conducting self-evaluations a review and validation of the results by an external advisory panel or peer review can help ensure both the quality and impartiality.

⁷ The Health Results Innovations Trust Fund (HRITF) is a multi-donor trust fund established in 2009 managed by the World Bank where donors (Norway and UK) committed \$575 million through 2022.

⁸ For guidance, see [IEG/DAC Sourcebook](#); [ECG, Template for Assessing the Independence of Evaluation Organizations](#); [United Nations Evaluation Group, Norms for Evaluation in the UN System, 2005](#).

Box 2. Commissioning independent evaluations of partnership programs without governing bodies

Health Research and Innovation Trust Fund Program (HRITF) recent evaluation was commissioned by the Evaluation Department of the Norwegian Agency for Development Cooperation and carried out by an external evaluation firm. All decisions concerning the TOR, the inception report, draft report and final report were subject to approval by the donor's evaluation department.

Source: Norad, Evaluation of the Health Results Innovation Trust Fund, April 2012

3. Invest time in planning the evaluation

14. A common challenge in partnership program evaluation is wide scope and coverage that can make them costly and time-consuming. It is indeed challenging to define where to draw the line in GRPPs and MDTFs when planning the evaluation. These programs are multi-year programs with activities at different levels. They support more than one type of activity (knowledge, technical assistance, and investments) and they have many stakeholders, donors and host agencies that often have different expectations and accountability requirements. For those who authorize and initiate an evaluation it is important to set clear expectations about the purpose of the evaluation.

15. The scope and coverage of evaluation derive from the purpose of evaluation and its objectives. Three key questions should guide those who commission and design a partnership evaluation before setting its scope:

- *What is the purpose of the evaluation?* If it is for accountability purposes it would generally require independence and broader coverage that would include the assessment at both the program and the activity level; while one focused on learning, could focused on specific issues (e.g. governance and management effectiveness, or assessing the results of a pilot phase)
- *Are resources sufficient* to conduct an assessment to serve the purpose of the evaluation? IEG's review of external evaluations indicates that on average, program level evaluations cost ranged between 1 to 3% of the program's annual expenditures.⁹ This is of course indicative and highly dependent on the size of the program and what the evaluation aims to achieve. Is there a sufficient evaluation capability?
- *Are enough data available to assess outcomes and impacts?* For programs with weak M&E, evaluators will need to re-construct programs' results chain to make a meaningful assessment. This work adds to the cost and time of the evaluation.

⁹ Independent Evaluation Group. 2011. [The World Bank's Involvement in Global and Regional Partnership Programs: An Independent Assessment.](#)

16. Another key issue in the evaluation planning is whether the authorizing body should itself draft the detailed evaluation plan, or delegate this work to contracted firm/experts. If the program and its governing body lack capacity to design the evaluation, they may tap the expertise of host agency evaluation department or hire an expert. A common practice is for the governing body to prepare broad terms of reference and the contracted evaluators prepare a detailed inception report, thus leaving the detailed design to the evaluators. The inception report, which is especially useful for evaluating large and complex partnerships, will have detailed evaluation design, state the objectives, key questions and methods, and data sources. Most of the external evaluations reviewed for these note prepared inception report to set evaluation objectives, key questions, methods and level of effort and activity timeline, etc.

4. **To have quality evaluation, choose criteria that fit the purpose**

17. The criteria (along with their respective detailed evaluation questions) to evaluate partnership programs can be adapted from standard criteria established by OECD/DAC, UNEG, IEG or a partner’s evaluation department. Most international development agencies and bilateral donors frame their evaluation questions around those criteria, at times adding few “customized” criteria. The latter is mostly to assess the results against donor’s own strategic priorities. [AusAid’s evaluation framework](#), for instance, assesses gender equity, analysis and learning in addition to those standards. [DFID’s evaluation policy](#) suggests looking at coherence, coordination and coverage especially in humanitarian context.¹⁰

18. A partnership evaluation generally uses two broad sets of criteria: a) **the development outputs, outcomes and results of the program itself;** b) **the organizational effectiveness of the program or how well the partnership is functioning?** The first set of criteria is about *results*, while the second set is about *process*.

4a. How well the program itself is working: relevance, effectiveness (achievement of outputs and outcomes defined in the results framework), efficiency or value for money, sustainability and impact.¹¹

19. *Relevance:* Most GRPPs have highly relevant goals and objectives at the time when they are established. Both the 24 IEG reviews and 20 external evaluations reveal three common relevance problems that many multi-donor programs with global and regional scope face:

- Weak evidence of beneficiary demand;
- Fading relevance of the program in changing global, regional or country context;
- Weak relevance of design: overly ambitious objectives do not match with the program’s resources and scope, or key activities are not linked strategically to the program’s objectives.

¹⁰ These additional criteria for evaluating humanitarian assistance were first developed in the DAC Guidance on Evaluating Humanitarian Assistance in Complex Emergencies, OECD, 1999.

¹¹ These criteria are indicative aimed to help plan the evaluation more logically and should be used selectively based on the objectives of the evaluation.

Box 3. One size doesn't fit all: Lighting Africa pilot program

The Mid-term Evaluation of IFC-World Bank Lighting Africa Program concluded that Lighting Africa overall had been an innovative program that had made important contributions to the growth of the market for quality portable off-grid lighting.

The program's relevance and effectiveness varied across the two pilot countries. The program was better fit in Kenya than in Ghana. In Ghana, relatively higher electrification rate, smaller market size, low government buy-in, and the presence of larger initiative with competing objectives and overlapping beneficiary groups hindered the pilot.

The effectiveness of program's individual components also varied. The program has been more effective in market intelligence and quality assurance components than in access to finance and consumer awareness raising.

While pilot was successful at country level, the review found that the replication and scale up of Kenya/Ghana style pilots across Africa region can be costly and a leaner, more regional model needs to be developed.

Source: Dalberg, *Mid-term Evaluation of IFC/World Bank Lighting Africa Project: Final Evaluation Report, November 2011*

20. Vague objectives and weak alignment of activities and outcomes are more common to partnership programs at early stage of development. Mature programs, by contrast, may be more likely to struggling to keep their relevance and to demonstrate their comparative advantages in a changing global context. Often weaknesses in program's design are clue to understanding why the program fails and is unable to achieve expected results. Flaws in program's design can be equally problematic in both mature programs and the new ones. Therefore, the assessment of program's design relevance should be an important part of GRPP evaluation.

21. *Effectiveness or efficacy*: Evaluations should be designed with more or less realistic expectations how much they can learn about the achievement of results. The quality of assessment of program's effectiveness highly depends on the availability and quality of data. Therefore the assessment of GRPP's monitoring and reporting system should be a key starting point.

22. In addition, partners need to be mindful of methodological challenges that are inherent to GRPPs. Generally, partnership programs support three types of activities – networking, knowledge creation and dissemination, country level technical assistance and investments. Assessing the effectiveness of each type of activity presents methodological challenges because they contribute in different ways and some are distant from the intended beneficiaries:

- Knowledge, advocacy, and standard-setting networks – can be effective and have an impact only if the knowledge products it generated are used by others. How to measure the extent the intended beneficiaries truly benefited from the program? ¹²
- Providing country-level technical assistance that aims to build capacity that may or may not lead to public and private sector investments. How to measure whether the products and services the program provided led to investments and whether the investment can be attributed to that specific intervention?
- Financing country level investments are often time supply driven and not aligned with country priorities.

23. Relying on [OECD-DAC definition of effectiveness](#), IEG reviews assess effectiveness or efficacy by establishing the program’s theory of change and systematically reviewing the progress of the program’s activities (outputs) in relations to plans and the extent to which these outputs are contributing to the achievement of the program’s objectives (outcomes) in accordance with its theory of change.

24. Some key findings from the IEG reviews in assessing efficacy of GRPPs are 1) overall, there is little systematic evidence of the achievement of objectives at the outcome level; 2) investment programs are more likely to demonstrate their achievement of outputs and outcomes than technical assistance programs and to even lesser extent knowledge networks; 2) programs that have narrower and more focused objectives have generally been more successful in achieving their outputs and outcomes.

25. *Efficiency* is another criterion that is applied in many ways across GRPPs, although not always thoroughly. Assessing efficiency is especially difficult for knowledge networks and norm- setting programs. While conducting an efficiency analysis for the partnership program as a whole can be difficult because of the different levels of activities, it is possible to conduct such analysis for individual activities and to compare these with sectoral benchmarks, or the cost of similar activities carried out by other programs. Some external evaluations conduct interviews and surveys of stakeholders to gauge the perceptions of program stakeholders on efficiency and effectiveness. (see for example Water and Sanitation Program’s 2013 external evaluation)

Box 4. Assessing Efficiency of Cities Alliance

The evaluation of Cities Alliance(CA) (2012) technical assistance program has focused its efficiency analysis on two questions:

- Are programs overhead costs reasonable and appropriate in relation to its objectives and activities?
- For beneficiary countries, has receiving the development assistance through Cities Alliance caused additional or reduced transactions costs? If so, why?

¹² UNEG, [Handbook for Conducting Evaluations of Normative Work in the UN System](#) 2013 offers guidelines how to assessing knowledge networks and standard setting programs; another useful source is ECG’s “[Evaluating Technical Assistance: Taking Stock of the Practices of International Financial Institutions](#)”, 2012.

The first question was assessed through comparing the annual administrative overhead costs of Cities Alliance to overhead costs in previous years and overhead costs in the World Bank and in other global partnership programs. It also supplemented the data with interviews with different CA members and partners and the program's secretariat. To assess the second question the evaluation focused on the transaction costs of the Cities Alliance's grant process for intended recipients.

Source: COWI, Independent evaluation of the Cities Alliance, Final Report, April 2012

26. A basic efficiency question is whether the costs of governing and managing the program are appropriate in relation to the objectives, strategy and activities, a comparison of those costs with programs of similar scale can already predict whether the program is in the right direction. Other important efficiency issues to consider are the effect of earmarking of funds on how the program allocates its funds, whether the program is increasing or decreasing transaction costs for both donors and beneficiary countries.

27. *Sustainability*: Two aspects of sustainability are important for GRPPs: 1) the sustainability of the program and, 2) the sustainability of the benefits generated by the program. The latter depends not only on the sustainability of the program itself, but also on the complementary activities of its donor partners and on the strengthening of institutional and human resource capacity in beneficiary countries to sustain the benefits.

Box 5. Weak efficiency can point to serious partnership issues

The Mid-term review of the Global Water Partnership (GWP) (2011), an advocacy and network program, found three sources of inefficiency:

- Overhead costs for governance and the secretariat are too high, comprising 30 percent of the budget.
- Program's current structure and its funding priorities are incompatible. Program's funding is too limited to achieve an acceptable level of efficiency and effectiveness.

Program's lack of a resource mobilization strategy threatens its sustainability.

Source: Ramboll Natura, Global Water Partnership Strategy 2009 to 2013 Mid-Term Review Final Review, October 2011

28. Generally, assessing sustainability of benefits of technical assistance and investment programs is easier than that of knowledge networks and standard setting programs.

29. Sustainability is more often assessed in mature programs when the program already completed many activities and achieved some of its results. For newer programs the evaluations could assess the steps the program is taking to make the benefits of its activities to be sustained in the future after the activities have been completed, such as strengthening institutional and human resource capacity of beneficiary countries and enhancing cooperation with other partners.

For programs that are set up to pilot innovative approaches a critical issue to look at could be whether the pilots are replicable and could be scaled up.

30. Most of 20 evaluations touched upon the issue of sustainability of the program, but fewer assessed the sustainability of program benefits. Most common problems affecting program sustainability were:

- the weak resource mobilization strategy;
- weaknesses in program’s governance and management;
- diminishing relevance of the program;
- failure to demonstrate results;
- high costs of the pilot preventing replication and scale up.

Box 6. Assessing sustainability of knowledge networks

The evaluation of Global Development Network (GDN), a knowledge network program, assessed sustainability from two perspectives: 1) Program’s financial sustainability and 2) the sustainability of benefits of the network program. The questions were assessed through a desk review of program’s current resource mobilization strategy and existing partnerships, and surveys of donors, staff and GDN’s regional network partners.

- The evaluation found that while the program has broadened its donor base, its financial sustainability is still weak because donors prefer to earmark their support. Survey indicated that donors are more willing to fund specific activities of GDN rather than contribute to GDN’s core operations. This could lead to distortion of program’s strategic direction.

The upcoming closure of GDNet, which is key in GDN’s ability to sustain the “network of networks” and to maximize the reach and impact of the research, would put the sustainability of program’s benefits in danger. The program needs to integrate some of the key functions of GDNet, such as knowledge capture, processing and sharing to sustain the program’s benefits.

Source: Chris Garforth, Nicholas Ozor, Richard Usher and Alison Bell, Global Development Network independent evaluation 2013, Final Report, February 10, 2014

4b. How well the partnership is functioning? A typical GRPP has a formal governing body along with a management unit. The way a program is run can be one of the key drivers of program’s success or failure. IEG’s 2011 review, for instance, found a strong qualitative correlation between effectiveness of the programs’ governance and the achievement of their objectives.

31. Some of the common questions the evaluators ask are:

- how well the governing body is overseeing the program management;
- are partners getting what they want out of it;
- are voice and power equitably distributed;
- is the financial burden-sharing appropriate;
- how effective is the partnership with the host organization;
- exit strategy.

32. Issues related to governance effectiveness are process issues, for which objective benchmarks are difficult to design. Structured interview combined with desk reviews of program's key documents is a key method for gathering views. IEG's recent review of [World Bank Group's Partnership with the Global Environment Facility](#) is a good example of gauging different partners' views on the effectiveness of the partnership between the two organizations. The study designed and conducted a set of structured interviews with GEF and Bank Group staff, members of the Advisory Panel, and with staff in other GEF Agencies involved in GEF operations. The review conducted electronic surveys of GEF Focal Points, GEF Program Managers, Task Team Leaders of World Bank environment projects, and World Bank Country Economists and Task Team Leaders of Country Partnership Strategies.

Box 7. Assessing the effectiveness of governance CGIAR's research programs

CGIAR's Independent Evaluation Arrangement reviewed the governance and management arrangements of 16 CGIAR Research Programs (2014) using the seven criteria, drawn from the IEG/DAC Sourcebook for Evaluating GRPPs: *legitimacy, accountability, fairness, transparency, efficiency, effectiveness, and independence*.

Drawing on desk reviews, interviews and surveys of different stockholders in those programs the evaluation found that:

- CGIAR Research Programs' (CRP) structures are duplicative and need to be streamlined. CRPs lack a single, balanced and independent body that can provide oversight and make effective decisions.
- CRP management structures need to be simplified, so that accountability for the performance of a CRP can clearly rest with CRP management.
- The dominant role of centers in CRP governance, which may be attributable to the level of center resources committed to CRPs, negatively affects the legitimacy of decision making, raises issues of conflicts of interest, and leads to insufficient participation of key stakeholders, including external partners, women and individuals from target regions.

Source: CGIAR Independent Evaluation Arrangement (IEA), Review of CGIAR Research Programs Governance and Management: Final report, March 2014.

33. IEG has suggested in 2007 a set of [principles of good governance](#) for assessing the governance and management of GRPPs. These principles were based on the [OECD](#)

[Principles of Corporate Governance](#). The main incentive for developing such criteria was to promote standards of good governance that all GRPPs as international public sector organization, should adhere.

34. More than half of the external evaluations touched upon issues related to governance and management: some of them explicitly referring to the principles of good governance. Issues more often discussed were around accountability and reporting mechanisms, transparency and risk management. Some of the common findings were:

- lack of clarity in roles and responsibilities between the governing body and the management;
- tensions in complying with host organization’s rules;
- inadequate attention to providing strategic guidance, to resource mobilization, and risk management;
- effectiveness of decision making.

5. **To guarantee credibility and ownership, make sure evaluation is transparent and key stakeholders are consulted**

35. It’s not enough for an evaluation to be properly carried out – if it is to have an impact, its findings need to be received, discussed, and acted upon.

5a. Transparency: Credible evaluation requires that the evaluation be as transparent as possible. As with any evaluation, evidence must be gathered in a systematic way and all data and evidence, (except personnel and salary matters, and the names of individuals quoted) should be made available. Many multilateral institutions, funds and bilateral partners, made transparency central to their evaluation policy. This effort, however, needs to be consistently extended to all the partnership programs they support. Transparency and disclosure can also facilitate consensus-building and ownership of the findings, conclusions, and recommendations among stakeholders.

5b. Consultation: To help ensure that all partners consider the evaluation credible and legitimate, all partners and stakeholders should be given equal opportunity to present their views at key stages of the evaluation process—planning, design, conduct, and follow up. Structured interviews are a good technique. The evaluators should also ask partners to suggest additional persons to be interviewed.

36. The evaluation team may consider organizing a workshop, round table, ‘learning session’, or similar forum to which all stakeholders are invited. If possible, this should be held on ‘neutral’ ground and at an early stage of the evaluation. The proceedings should be recorded and made available. GEF’s recent comprehensive evaluation (OPS 5), for instance, was extensively discussed by the program’s stakeholders in two meetings prior to finalizing it and the proceedings are posted on the program’s [website](#).

Box 8. Using stakeholder input to shape an evaluation

Before launching the “Five year evaluation of the Global Fund to Fight AIDs, Tuberculosis, and Malaria, the Evaluation Reference Group commissioned an independent assessment of stakeholder opinion “[360° Stakeholder Assessment](#)” on the organization’s reputation, performance, strengths and weaknesses. This was an online survey of over 900 respondents across the world, representing all major stakeholder groups and regions. The results helped to shape the focus and methodology of the evaluation and are posted on the program’s website.

37. The evaluation team may consider organizing a workshop, round table, ‘learning session’, or similar forum to which all stakeholders are invited. If possible, this should be held on ‘neutral’ ground and at an early stage of the evaluation. The proceedings should be recorded and made available. GEF’s recent comprehensive evaluation (OPS 5), for instance, was extensively discussed by the program’s stakeholders in two meetings prior to finalizing it and the proceedings are posted on the program’s [website](#).

6. **Ensure that recommendations are agreed on and followed up**

38. Evaluation findings should not come as a surprise to the program management and stakeholders. Since all partners will share accountability for the findings, the evaluators should reveal their preliminary findings and invite feedback. This can be done by circulating a preliminary draft or making a presentation at a steering committee meeting or other event.

39. All the feedback — whether or not adopted in the final report — should be placed on the record at the partnership’s website. In addition, the formal responses to the evaluation may be annexed to the final evaluation report if they emerge in a timely fashion, or they may be disclosed later through other means.

40. As with all evaluations, the lessons and recommendations should derive directly from findings. In partnership programs recommendations may be targeted to particular groups of partners – for example, donors, implementing agencies, and governance bodies. Joint evaluation of [Joint Programmes on Gender Equality in the United Nations System](#) (2013), for instance, directed separate recommendations to the UN Agencies, host governments and citizens, and donors, and UNDG.

41. It is usually desirable to limit the number of formal recommendations to the highest-priority issues that program stakeholders can readily monitor. Such limitation can be challenging in comprehensive evaluations of large and complex programs. One option can be to craft sub-recommendations under each major recommendation, as was done in the Global Environment Facility’s most recent [Overall Performance Study \(OPS5\)](#), as shown in the box below.

Box 9. Using a hierarchy of recommendations in the GEF's comprehensive evaluation

Recommendation 2: The business model of the GEF needs major overhaul in the GEF -6 period

- The results-based management (RBM) framework for GEF -6 should include a limited number of outcome indicators that can be measured through existing or easily generated data. The Independent Evaluation Office should assess the evaluability of this framework before it is finalized by the Council.
- The tracking tools should be simplified, and where global public knowledge databases are receiving the generated data, this should be implemented and funded adequately. The burden of the tracking tools on multifocal area projects should be reduced.
- The GEF should shift co-financing considerations to programming and to the CEO endorsement and GEF Agency approval stages, to encourage partners on the ground to continue to find appropriate solutions that lead to high levels of co-financing, solid financing of baselines, and increased global environmental benefits.

Source: Global Environment Facility Independent Evaluation Office, OPS5 Fifth Overall Performance Study of the GEF, Final Report: At the Crossroads for Higher Impact, 2014

7. Plan dissemination in advance

42. The final evaluation report should be discussed by the partnership program's governing body or steering committee, and should be placed on the partnership's website, along with any management response or action plan.

43. Three-quarters of the 20 recent external evaluations are posted on the respective program's website. This is an improvement since 2011, when fewer than half of programs reviewed by IEG had made their external evaluations publicly available.¹³

44. Of the 15 programs whose evaluation is posted, only six also display their management response and actions to be taken. [Global Water Partnership](#) is a good practice example of sharing the external review of the program and the concrete steps program management intends to take as a response to the evaluation.

Moving forward

45. Despite challenges, GRPP are evaluated more often and more systematically than MDTFs. This is because GRPPs are of higher profile and the demand to hold them accountable is higher. Their designated governing bodies play critical role in that.

¹³ Independent Evaluation Group. 2011. *The World Bank's Involvement in Global and Regional Partnership Programs: An Independent Assessment*.

46. Their evaluations have had notable impacts on the programs that IEG has reviewed¹⁴.
- Half the programs modified their governance arrangements in some way, such as amending their charters or establishing an executive committee to make their governance process more agile and efficient.
 - Almost half the programs revised their strategies
 - About one-third improved their M&E systems and results reporting.
47. Overall, in external evaluations there is more shared understanding around the evaluation criteria on how to assess development effectiveness of the programs. With slight variations those criteria and related questions are anchored in the OECD-DAC standards.
48. Practice varies when it comes to applying good practice standards during evaluation planning and follow-up. There the expectations and behavior of development partners is so diverse that can result in anything between the duplication of evaluation efforts, sub-standard evaluations and a high quality independent assessment. What could be the value of applying similar evaluation frameworks by different partners who commission parallel evaluations of the same program almost at the same time?
49. The authorizing environment and governance of an evaluation matter. Development partners need to be consistent and collaborate better to support good practice standards for credible and quality evaluations at all stages of evaluation in each and every partnership program— starting from synchronizing their evaluation plans with other partners, taking steps to ensure independence of the evaluation, to making sure the recommendations are evidence- based and actionable.
50. The evaluations of MDTFs either external or self-evaluations, are more limited (at least for those located at the World Bank) suggesting that both donors and the hosts need to pay more attention to elevating the monitoring and evaluation standards for these types of programs as well. Such program seems to be below the radar of the key partners, including the hosts in terms of transparency and accountability for results.

¹⁴ This is based only on IEG’s Global Program reviews.

Attachment 1

References

External Evaluations reviewed:

1. Baastel, *External Evaluation of ESMAP 2007-2011, Final Evaluation Report*, June 19, 2012
2. Boston University Center for Global Health and Development, *External Evaluation of the Roll Back Malaria (RBM) Partnership 2009-2013*, December 31, 2013
3. Cambridge Economic Policy Associates Ltd., *Public Private Infrastructure Advisory Facility 2012-2013 Evaluation, Final Report*, June 21, 2013
4. CGIAR Independent Evaluation Arrangement (IEA), *Review of CGIAR Research Programs Governance and Management: Final report*, March 2014.
5. Chris Garforth, Nicholas Ozor, Richard Usher and Alison Bell, *Global Development Network independent evaluation 2013, Final Report*, February 10, 2014
6. COWI, *Independent evaluation of the Cities Alliance, Final Report*, April 2012
7. Dalberg, *Mid-term Evaluation of IFC/World Bank Lighting Africa Project, Final Evaluation Report*, November 1, 2011
8. DARA, *Retrospective evaluation of the GFDRR program in a sample of disaster-prone countries*, April 2014
9. Global Environment Facility Independent Evaluation Office, *OPS5 Fifth Overall Performance Study of the GEF, Final Report: At the Crossroads for Higher Impact*, 2014
10. ICF International, *Independent Evaluation of the Climate Investment Funds, Volume 1, Evaluation Report*, June 13, 2014
11. IOD Parc, *Joint Evaluation of Joint Programmes on Gender Equality in the United Nations System Final Synthesis Report*, November 2013
12. Macro International, *The Five-Year Evaluation of the Global Fund to Fight AIDS, Tuberculosis, and Malaria Synthesis of Study Areas 1, 2 and 3*, March 2009
13. Nexus Associates, *Better Work Global – Stage II Evaluation Summary*, April 2012
14. Norad, *Evaluation of the Health Results Innovation Trust Fund*, April 2012
15. Rajasingham, Sanjivi, Gwen, Catherine, and Klas Ringskog, *Independent Evaluation of the Infrastructure Partnership programs and the Water and Sanitation Initiative Global program*, March 2012.
16. Ramboll Natura, *Global Water Partnership Strategy 2009 to 2013 Mid-Term Review Final Review*, October 2011
17. The State and Peace-Building Fund, *Mid-Term Review of the State-and Peace-Building Fund Executive Summary*, October 2012
18. Toroa Strategy, *Independent Evaluation Report of the Global Public-Private Alliance for Responsible Fisheries (ALLFISH) 2009 to 2011*, May 2012
19. Universalia, *CGAP Phase IV Mid-Term Evaluation, Revised Report*, April 2012
20. Universalia, *External Mid-Term Evaluation of Water and Sanitation Program (WSP) for FY11-FY12 Final Report*, June 28, 2013

[IEG Global Program reviews:](#)

1. Global program review: Together for Energy: How Partnership Programs Support Energy Access: A review of Global Partnership on Output-Based Aid (GPOBA), Energy Sector Management Assistance Program (ESMAP), Lighting Africa, and Asia Sustainable and Alternative Energy Program (ASTAE), (2014 forthcoming)
2. Global Program Review: The World Bank's Partnership with the GAVI Alliance, August 1, 2014
3. Global Program Review: The World Bank Group's Partnership with the Global Environment Facility December 6, 2013
4. Global Program Review: The World Bank Group's Partnership with the Global Global Program Review: The Global Facility for Disaster Reduction and Recovery (GFDRR), October 10, 2012
5. Global Program Review: The Global Fund to Fight AIDs, Tuberculosis and Malaria: Volume I and II: Main Report and Appendixes, February 8, 2012
6. Global Program Review: Multi-Donor Trust Fund for the Extractive Industries Transparency Initiative, February 1, 2011
7. Global Program Review: Global Water Partnership, August 1, 2010
8. Global Program Review: The Global Development Network, May 28, 2009
9. Global Program Review: The Consultative Group to Assist the Poor (CGAP), December 1, 2008
10. Global Program Review: Cities Alliance, June 28, 2007
11. Global Program Review: The Forest Carbon Partnership Facility (FCPF), October 10, 2012
12. Global Program Review: Marrakech Action Plan for Statistics (MAPS), Partnership in Statistics for Development in the 21st Century (PARIS21), Trust Fund for Statistical Capacity Building (TFSCB), June 30, 2011
13. Regional Program Review: The Mesoamerican Biological Corridor: A Regional Program Review, May 24, 2011
14. Global Program Review: international Assessment of Agricultural Knowledge, Science, and Technology for Development (IAASTD), June 1, 2010
15. Global Program Review: The Stop TB Partnership, November 19, 2009
16. Global Program Review: Global Invasive Species Program, September 3, 2009
17. Global Program Review: Global Forum for Health Research, June 23, 2009
18. Global Program Review: International Land Coalition (ILC), June 1, 2008
19. Global Program Review: Association for the Development of Education in Africa (ADEA), March 1, 2008
20. Global Program Review: Population and Reproductive Health Capacity Building Program (PRHCBP), March 1, 2008
21. Global Program Review: The Critical Ecosystem Partnership Fund (CEPF), December 1, 2007
22. Global Program Review: The Development Gateway Foundation, June 26, 2007
23. Global Program Review: Medicines for Malaria Venture, June 26, 2007
24. Global Program Review: The Provention Consortium, June 28, 2006