

MINISTRY OF FOREIGN AFFAIRS OF DENMARK
Danida

FROM DANIDA'S EVALUATION DEPARTMENT - DEVELOPMENT COOPERATION, MARCH 2008

ASSESSING MULTILATERAL ORGANISATION EFFECTIVENESS



EVALUATION STUDY 2008/3

Assessing Multilateral Organisation Effectiveness

March 2008

Alison Scott, Simon Burall, Nick Highton and Karina Wong
Overseas Development Institute www.odi.org.uk

Disclaimer: The views presented in this paper are those of the authors and do not necessarily represent the views of DANIDA

Acknowledgements: We would like to thank Marta Foresti and Simon Maxwell for their insightful comments during the writing of this paper, and Geoff Handley for responding so quickly to requests for information.

Acronyms

ADF	African Development Fund
AfDB	African Development Bank
ARDE	Annual Review of Development Effectiveness
AROE	Annual Report on Operations Evaluation
AsDB	Asian Development Bank
CIDA	Canadian International Development Agency
COMPAS	Common Performance Assessment System
DAC	Development Assistance Committee (of OECD)
DADA	Duska-Anema Development Associates
DANIDA	Danish International Development Agency
DFID	UK Department for International Development
DRI	Debt Relief International
EBRD	European Bank for Reconstruction and Development
EC	European Commission
EU	European Union
FAO	Food and Agriculture Organisation of the UN
GFATM	Global Fund to fight AIDS, Tuberculosis and Malaria
HIPC	Heavily Indebted Poor Country
HR	Human resources
HQ	Headquarters
IADB	Inter-American Development Bank
ICRC	International Committee of the Red Cross
IDA	International Development Association (of World Bank)
IEE	Independent External Evaluation
IFAD	International Fund for Agricultural Development
IFRC	International Federation of Red Cross Red Crescent Societies
ILO	International Labour Organisation
JIU	Joint Inspection Unit
M&E	Monitoring and Evaluation
MDBs	Multilateral Development Banks
MDES	Multilateral Development Effectiveness Summaries
MDGs	Millennium Development Goals

MEFF	Multilateral Effectiveness Financing Framework
MERA	Multilateral Evaluation Relevance and Assessment system
MfDR	Managing for Development Results
MOPAN	Multilateral Organisations Performance Assessment Network
NGO	Non-governmental Organisation
OCHA	United Nations Office for the Coordination of Humanitarian Affairs
ODA	Official Development Assistance
ODI	Overseas Development Institute
OECD	Organisation for Economic Co-operation and Development
OHCHR	United Nations Office of the High Commissioner for Human Rights
PCR	Project Completion Reports
PMF	Performance Management Framework
PRS	Poverty Reduction Strategy
PRSP	Poverty Reduction Strategy Paper
RBM	Results Based Management
RIDE	Report on IFAD Development Effectiveness
SPA	Strategic Partnership for Africa
UN	United Nations
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNDP	United Nations Development Programme
UNEG	United Nations Evaluation Group
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UNFPA	United Nations Population Fund
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations Children's Fund
UNIDO	United Nations Industrial Development Organisation
UNIFEM	United Nations Development Fund for Women
WB	World Bank
WFP	World Food Programme
WHO	World Health Organisation

Contents

Acknowledgements.....	i
Acronyms.....	ii
Contents.....	iv
Executive Summary.....	1
1 Introduction.....	3
1.1 Background.....	3
1.2 Purpose and Scope of Paper.....	4
2 Identifying Approaches for Assessing Multilateral Effectiveness.....	6
2.1 Defining the Terms.....	6
2.2 Categorising the Assessment Approaches.....	9
2.3 Aspect of Effectiveness Examined.....	11
2.4 Types of Commissioning Agent.....	14
3 Mapping Different Approaches for Assessing Multilateral Effectiveness.....	15
3.1 Evaluations.....	15
3.2 Results Based Management Assessments.....	16
3.3 Assessments of Partnership Behaviour.....	19
3.4 Synthetic Approaches.....	23
4 What can the Different Approaches tell us About Effectiveness?.....	26
4.1 Methodological Issues.....	26
4.2 Assessing the Strengths and Limitations of the Different Approaches...29	
4.3 Gaps.....	31
5 Ways Forward for Multilateral Effectiveness Assessments.....	33
5.1 Can we rely on the Multilaterals to Assess their own Effectiveness?.....	33
5.2 Bilateral Assessments of Multilateral Effectiveness.....	34

5.3 Improving the Bilateral Governance Role.....	37
5.4 Meeting the Needs of Partner Countries	38
5.5 Implication of Effectiveness Assessments for the Global Aid Architecture	39
6 Conclusion	40
Annex 1 - References	43
Annex 2 – Terms of Reference.....	46
Annex 3 – Comparing the Different Approaches to Assessing Multilateral Effectiveness.....	48
Annex 4 – The Importance of Additional Partnership Behaviour Criteria for Partners	54

Executive Summary

Key Points

- Individual approaches for assessing multilateral effectiveness only give a very partial picture of effectiveness; all are useful and have both strengths and limitations. Donors should continue to explore approaches that synthesise information from a range of sources.
- Common standards for assessing effectiveness will need to be developed through international networks such as the Multilateral Organisations Performance Assessment Network (MOPAN) group
- The current proliferation of bilateral assessments of effectiveness is inefficient. Ideally, bilaterals should use their role on the governing boards of multilaterals to improve multilateral reporting on effectiveness so that bilateral assessments become redundant. However, in the short run, bilateral assessments are likely to continue.
- Bilaterals should clarify the rationale for conducting separate assessments of multilateral effectiveness for their internal decision-making processes and conduct these assessments collectively with other donors through networks such as MOPAN rather than separately.
- In the context of the Paris Declaration, and the emphasis on ownership and mutual accountability, donors must consider ways to include partner countries more directly in current assessments and how better to meet their information needs.
- Assessments of multilateral effectiveness provide information on only one part of the international aid system. Donors should consider ways of opening up broader discussions about the effectiveness of the system as whole.

The issue of the cost-effectiveness of bilateral aid budgets has become increasingly important over the past ten years or so. A key element in the debate has been the extent to which bilateral agencies should fund the activities of multilateral agencies, and which out of more than 230 that are currently operating to fund. In order to aid these decisions, donors have developed different methodologies for assessing the effectiveness of organisations within the system. In this paper we focus on approaches to assess multilateral effectiveness. We note that the term effectiveness has a number of different meanings and we provide some key definitions.

We develop a two dimensional framework to categorise the different assessment approaches. The first dimension focuses on the *aspect* of effectiveness that is assessed; the second dimension looks at the type of *commissioning agency*. Within this framework we identify the strengths and weaknesses of each approach as well as

more generally. We conclude that no single approach is able to provide a complete view of multilateral effectiveness, and that all provide important information, although there are gaps. There is a need to synthesise information on effectiveness from a range of sources.

We highlight a number of problems associated with the proliferation of assessments by bilateral donors. We conclude that, ideally, bilaterals should use their positions on the governing bodies of multilateral agencies to advocate improvements in multilateral reporting on effectiveness so that separate bilateral assessments would become unnecessary. We note that the International Fund for Agricultural Development's (IFAD) Report on IFAD's Development Effectiveness (RIDE) may be one model to build on.

In the short run though, and given the need for improvements in multilateral reporting, bilateral agencies are likely to continue to need separate assessments of multilateral effectiveness as an input to their internal decision-making. However, they should be more explicit in what those factors are and how they use the information. If bilaterals are to carry out their own assessments, they should do this collectively through international groupings such as the MOPAN group. In addition, such groupings should expand their remit to include the development of an international consensus on criteria and standards for the assessment of effectiveness.

Bilateral donors are not the only actors in the system making decisions about how, when and to what extent to engage with multilateral organisations. Partner governments also make decisions of this nature when they seek additional financing. Donors should consider how best to increase the involvement of partner governments and wider stakeholder groups in assessment methodologies as part of a move towards meeting their information needs more directly. These must focus on the effectiveness of both multilateral and bilateral agencies if partner governments are to make decisions about their engagement with the system as a whole.

Finally, the international context is changing. The system is becoming more complex and there are concerns that it is becoming less efficient as a result. Assessments of multilateral effectiveness only provide information about individual elements of the system and not about the effectiveness of the system as a whole. Donors must therefore consider ways in which the effectiveness debate can contribute to dialogue with other actors about improving the system as a whole.

1 Introduction

1.1 Background

Over the last decade, the international community has increasingly focused on issues of performance, effectiveness and results when assessing progress towards poverty reduction in the developing world. This focus has been accompanied by efforts amongst bilateral and multilateral agencies to measure, monitor and report on results. Major factors in promoting this agenda have been:

- The espousal during the 1990s, of Results Based Management (RBM) by Western governments as part of the ‘modernising government’ agenda, and the incorporation of these techniques into bilateral aid agencies;
- The international adoption of an agreed framework for measuring progress towards development results, based on the Millennium Development Goals (MDGs); and
- Increasing domestic pressures on bilateral governments to account for increases in their overseas development assistance (ODA) and support to the multilaterals.¹

As financiers of the multilaterals, bilateral governments have increasingly pressurised the multilaterals to improve their performance, and to demonstrate this through better monitoring, evaluation and reporting. In response, the Multilateral Development Banks (MDBs) have developed results frameworks for concessional lending, and many United Nations (UN) agencies have developed results-oriented planning and reporting.²

At the same time, bilateral agencies have increasingly sought their own independent assessments of multilateral effectiveness. Initially, most of these were done internally and consisted of surveys of their own staffs’ perceptions of multilateral performance. With time, they became more systematic and rigorous, and expanded in scope to include a review of multilateral RBM systems. As these studies proliferated, the need for cooperation amongst bilaterals arose in order to pool findings and reduce transactions costs. The MOPAN was formed in 2002 and now has 11 bilateral donor members³. It has conducted joint annual surveys of multilateral partnership behaviour since 2003.

¹ Development Assistance Committee (DAC) projections are for ODA to increase to around \$140bn by 2010 (see <http://www.oecd.org/dataoecd/13/26/38341348.pdf>)

² The World Bank introduced results measurement frameworks in International Development Association (IDA) rounds 13 and 14, the African Development Bank (AfDB) followed suit in African Development Fund (ADF) X round, United Nations Development Programme (UNDP) introduced a results oriented Multi-Year Funding Framework

³ Austria, Canada, Denmark, Finland, France, Ireland, The Netherlands, Norway, Sweden, Switzerland, The United Kingdom

Other cooperative ventures also emerged amongst the multilaterals and more generally. In 2003, the MDBs formed a Working Group on Managing for Development Results (MfDR). It developed the Common Performance Assessment System (COMPAS) in 2005, which produces an annual report on the performance of the five MDBs. Following the adoption of international commitments towards harmonisation and alignment in Monterrey 2002, Rome 2003 and Paris 2005, the Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) has spearheaded a number of country surveys to monitor the rollout of the harmonisation and alignment agenda. A few non-governmental organisations (NGO), such as Debt Relief International (DRI), have also developed their own surveys of multilateral performance.

Not surprisingly, bilateral donors are now entering a period of reflection about the value of these assessments, reviewing the relative merits of the different approaches and seeking improved ways to pool and synthesise results.

1.2 Purpose and Scope of Paper

This paper has been commissioned as a contribution to the reflections of bilateral donors about the value of the various approaches to assessing multilateral effectiveness. We begin section 2 by defining some of the key terms and then propose a scheme for categorising the different assessment approaches. Section 3 briefly describes a number of the assessment approaches according to this categorisation. In section 4, we explore some key methodological issues. We then consider the strengths and weaknesses of the different types of approach before identifying a number of gaps that are not covered by any of the existing approaches. In section 5 we discuss ways forward for future assessments of multilateral effectiveness, including:

- The relative merits of reliance on assessments by multilaterals as opposed to bilateral agencies;
- The best way of fulfilling the multiple purposes that bilaterals have for effectiveness assessments;
- How the needs of partner countries can be met by effectiveness assessments; and
- The need for effectiveness assessments to inform discussion about reform of the aid system as a whole.

This paper provides a selective rather than comprehensive review of the approaches to assess multilateral effectiveness. In part this is because it only examines those approaches for which there is publicly available information on methodology and results. We examine approaches undertaken by two bilateral agencies, by multilateral agencies individually, by international groupings of bilateral and/ or multilateral agencies, and one by an NGO. We focus mainly on the implications for bilateral donors while carrying out this analysis, though, in the context of the Paris Declaration and the focus on country ownership, we also explore the implications for partner governments.

This paper was produced by reviewing the available literature and has not included any consultations with any of the agencies implementing the approaches. It is also important to note that the MOPAN group is currently discussing changing their methodology and the final shape of this new approach has yet to be agreed.

2 Identifying Approaches for Assessing Multilateral Effectiveness

2.1 *Defining the Terms*

Development agencies have developed a number of different ways of thinking about the term effectiveness as they try to assess their contribution to development goals. We focus below on the main ways that the term ‘effectiveness’ is used. All of the different uses of the term take as their ultimate goal how well any intended results are achieved. As is clear below, these terms are interrelated and can be used in different ways in different contexts.

2.1.1 *Effectiveness*

At its simplest, ‘effectiveness’ is the likelihood of achieving the intended objectives of an activity, policy or other intervention. The DAC defines effectiveness as, “*the extent to which a development intervention has attained, or is expected to attain, its relevant objectives*” (OECD/DAC, 2001, p5). The concept of ‘results’ is integral to the discussion of effectiveness, but the term is very general and can be applied to a variety of effects at country level, and within the aid organisation.

At country level, ‘results’ can refer to outputs, outcomes, and impacts.⁴ The term ‘*output*’ normally refers to the products and services directly produced by a development project or programme (e.g. roads constructed, teachers trained). ‘*Outcomes*’ are the results achieved beyond immediate project outputs (e.g. markets expanded, service delivery times reduced, literacy rates increased); they can be influenced by other factors external to the project and may extend beyond direct project beneficiaries. Outcomes vary in their timeframes and usually refer to short to medium-term results – also referred to as ‘intermediate outcomes’ to distinguish from longer-term ones. The longer the timeframe, the wider the populations that may be affected and the more difficult it is to trace the causal connections between donor activity and results at this level. ‘*Impacts*’ usually refer to long-term development results and sustainability at a more general societal level. These are the most difficult of all for the measurement of effectiveness and require special evaluation methodologies (see below 2.3.1). For this reason, most discussions of multilateral effectiveness at country level are confined to outputs and intermediate outcomes.⁵

The discussion of organisational effectiveness also includes results that are internal to the organisation, but which can have an effect on country level results, such as project design quality, speed of disbursement etc. (see 2.1.6).

⁴ ‘Results’ are defined by the DAC as ‘the output, outcome or impact (intended or unintended, positive and/or negative) of a development intervention.’ (OECD/DAC, 2002)

⁵ Some RBM experts confine the discussion to outcomes only, while others – such as the present study - include outputs and outcomes.

2.1.2 *Development effectiveness*

Development effectiveness is the likelihood of achieving country-level development objectives. Since the UN Millennium Summit in 2000, the broad international consensus is that these objectives should be expressed in terms of the eight MDGs. However, these MDGs refer to mid- to long-term development results, which arise from the activities of a wide range of actors including the partner government, NGOs, businesses, and donor agencies. It is therefore difficult to attribute these results to the actions of specific agencies. The discussion of *multilateral* development effectiveness therefore normally refers to the outputs and intermediate outcomes that can be more closely linked to the activities of the organisations concerned.

2.1.3 *Aid effectiveness*

Aid effectiveness is the likelihood of aid (usually defined as financial flows) contributing to development. The term raises questions such as what country level results can be associated with a particular aid volume and different aid delivery modalities, and how allocation across countries, or sectors, for example, affects these results. There is a long literature, primarily using econometric techniques, that attempts to link aid to long-term development impacts.⁶

2.1.4 *The Paris Declaration on Aid Effectiveness*

The international consensus on the importance of *aid quality* (as opposed to deliverables) is best captured by the Paris Declaration on Aid Effectiveness, which synthesises earlier thinking in this area (OECD/DAC, 2005). The Declaration is based on the premise that donor behaviour at the country-level can enhance the effectiveness of aid by improving their partner relationships and aid delivery processes. The Paris Declaration is built around five mutually reinforcing dimensions: ownership; alignment; harmonisation; managing for results; and mutual accountability (see Figure 1). The changes in behaviour implied by these dimensions are expected to improve the development effectiveness of the country-level system as a whole.

2.1.5 *Organisational effectiveness*

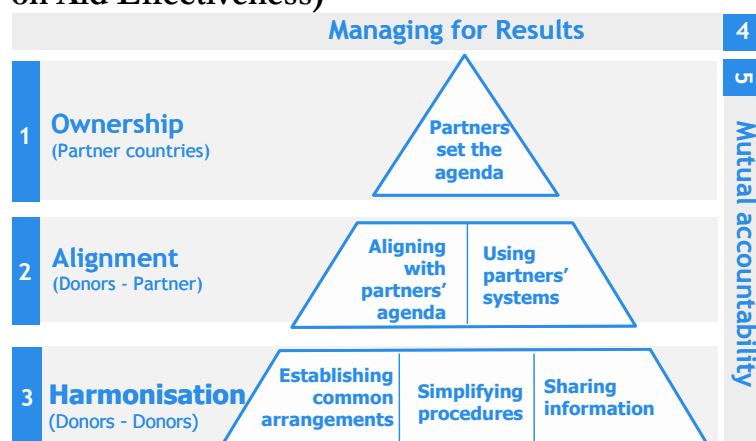
The term ‘organisational effectiveness’ is usually deployed to contrast with ‘development effectiveness’. It focuses on the direct results of an agency’s interventions, for which it can be held accountable, in contrast with development outcomes, which are the effect of many agencies’ interventions. The term organisational effectiveness usually focuses on the internal systems that are geared towards producing development outputs and outcomes.⁷ In this sense, it is synonymous with results based management and managing for development results (see below). There are two main questions:

- How well do organisations manage themselves internally in order to maximise their development outcomes? and
- How well do organisations behave in relation to the Paris Declaration?

⁶ See for example Collier and Dollar, 2001

⁷ In some studies, organisational effectiveness includes project outputs, but not outcomes.

Figure 1: The Five Pillars of the Paris Declaration (OECD Working Party on Aid Effectiveness)



2.1.6 Results based management

RBM originated in the private sector, but was soon adopted by public sector organisations in western governments. RBM is also known as Performance Based Management. RBM is a management tool that applies a results focus to all internal business processes (e.g. planning, operations, budget, human resources, knowledge management), and all levels of the organisation (corporate management, divisions and departments, work units, individuals). As deployed by development agencies, RBM can be defined as a management strategy that aims to link internal performance to the achievement of development outcomes (OECD/DAC, 2000). RBM emphasises a learning culture and stresses the measurement of results through monitoring and evaluation.

2.1.7 Managing for Development Results

This refers to how well organisations are managing themselves internally through the application of RBM techniques in order to maximise their contribution to development outcomes. It is essentially an RBM approach, but focuses on the difference between managing *for* as opposed to *by* results. This is because, in contrast to private sector firms, it is much more difficult to attribute results directly to an individual agency's interventions. Early thinking on MfDR was developed in 2002 and led to the formation of the MDB MfDR Working Group. As part of the OECD/DAC Working Group on Aid Effectiveness, a Joint Venture on MfDR was formed between the DAC and the MDBs, which has set out some core principles, promoted MfDR at developing country level and produced a best practice Sourcebook on MfDR.⁸ This is the fourth dimension of the Paris Declaration. MfDR may also be referred to as 'organisational effectiveness' or 'internal performance.'

⁸ The MfDR principles are: focusing the dialogue on results at all phases of the development process; aligning programming, monitoring, and evaluation with results; keeping measurement and reporting simple; managing *for*, not *by*, results; using results information for learning and decision-making (OECD/DAC, 2007a).

2.1.8 *Efficiency*

The term *effectiveness* is often contrasted with that of *efficiency*. Efficiency focuses on the input-output relationship, as opposed to outputs and outcomes. High efficiency would be exemplified by the delivery of a large number of outputs (schools built, teachers trained) for given inputs. However, these outputs may not necessarily produce the desired outcomes, such as increased primary school enrolments or literacy rates because of, for example, faulty design (schools built in the wrong places, teachers given inappropriate training). Assessing efficiency can be an aspect of both RBM and MfDR approaches.

2.1.9 *Relevance*

The term relevance has three aspects:

- Relevance to the *partner country* priorities (as expressed in Poverty Reduction Strategy Papers (PRSPs), for example);
- Relevance to *donor policies and strategies* (e.g. gender, environment, human rights); and
- Relevance to the *mandate* of the multilateral.

These three aspects may conflict. For example, a donor may give higher priority to some aspects (e.g. human rights) than a partner country, and a multilateral may be restricted by its mandate to addressing aspects considered less important by the bilateral donor.

2.1.8 *Performance*

This is usually considered to be a broader term than effectiveness and includes efficiency and behaviour. However, in practice, these terms are often used interchangeably.

2.2 *Categorising the Assessment Approaches*

We propose a two dimensional framework for categorising the different assessment approaches.⁹ The first dimension is based on the *aspect* of effectiveness examined; the second dimension highlights the *commissioning agency* that undertakes the assessment. These two dimensions are important because they imply differences in the objectives, the types of information gathered, the methodologies used, and the conclusions that may be reached about multilateral effectiveness.

In terms of the *aspect* of effectiveness examined, there are the following:

- *Evaluations* assess the development outcomes and impacts that are achieved at country level as a result of the multilateral's activities, policies or processes;

⁹ There are other ways of categorising multilateral effectiveness assessments. Meier looks at bilateral approaches only and categorises them according to their degree of institutionalisation (Meier, 2007a); Obser, looking at both bilateral and multilateral approaches, categorises them by the purpose of the assessment approach (Obser, 2007).

- *Results based management assessments* focus on the effectiveness of the multilateral's internal business processes, which are aimed at achieving development outcomes;
- *Assessments of partnership behaviour* examine how multilaterals conduct their business with other development partners, especially at country level. Increasingly such assessments examine how well organisations are performing against the principles of the Paris Declaration;
- *Synthetic Approaches* cover several of these aspects in order to make a more comprehensive assessment of effectiveness.

In terms of the *commissioning agency*, there are:

- *Multilaterals* themselves, including their internal monitoring and evaluation (M&E) systems (such as portfolio reviews) and their independent evaluations;
- *Bilateral* assessments, carried out by agencies separately;
- *International collaborations*, amongst bilaterals or by bilateral and multilaterals together, including joint bilateral surveys (MOPAN), or bilateral/multilateral surveys (OECD/DAC and the Strategic Partnership for Africa (SPA)); and
- *NGOs*, such as Debt Relief International.

The approaches identified with this categorisation are presented in Table 1 below. A number of other bilateral approaches are not listed in this table, such as the Canadian Multilateral Evaluation Relevance Assessment (MERA) and the Dutch Multilateral Monitoring Survey because neither their methodologies nor their results are publicly available.¹⁰

¹⁰ For more information on these approaches see DADA International, 2006 and Meier, 2007a.

Table 1: Framework for analysing approaches to assessing multilateral effectiveness

Commissioning Agency	Evaluations	Results Based Management	Partnership Behaviour	Synthetic
Multilateral	External evaluations Internal independent evaluations UN Joint Inspection Unit (JIU)	(COMPAS (MDB) Internal M&E (portfolio reviews, sector reviews etc)	Client surveys	RIDE (IFAD)
Bilateral		Multilateral Effectiveness Framework (MEFF) (UK Department for International Development (DFID)) RBM (Danish International Development Agency (DANIDA))	Partner stakeholder perceptions survey (DFID)	Performance Management Framework (PMF) (DANIDA) Multilateral Development Effectiveness Summaries (MDES) (DFID)
International			MOPAN Survey Paris Survey SPA Budget Support Survey	[New MOPAN]
NGO			DRI	

We now briefly describe the characteristics of each of the assessment approaches identified above. Further details are provided in Annex 2.

2.3 Aspect of Effectiveness Examined

2.3.1 Evaluations

Evaluation has been described as *‘the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results’* (OECD/DAC,

2002). Evaluations examine the different levels of the logframe: inputs, outputs, outcomes and impacts, but in most cases the emphasis is on outcomes and impact. Standard criteria include relevance, effectiveness, efficiency, impact and sustainability. Evaluations cover operational projects, country programmes, specific policies and/or corporate issues. Most evaluations, however, are focused on operational work and are usually carried out after the activity or intervention has taken place. Therefore they are able to judge the likely achievement of development outcomes better than other M&E processes.

Evaluations use a range of methodologies to enable triangulation between different sources of information. These include documentary reviews, primary data collection, and secondary literature reviews. Fieldwork is carried-out involving some or all of the following: site visits; key informant interviews, focus groups, workshops and mini-surveys. Beneficiaries will also be included. Evaluations therefore offer more breadth and depth than other assessment approaches.

Most large multilaterals have evaluation offices within their institutions. The majority of these have policies guiding evaluation practice at operational and policy levels and the quality of these policies is very high (Lloyd et al, 2007). In recent years, multilaterals have begun to focus on increasing the independence of their evaluation offices by reporting to their governing bodies rather than to management and this is now the case in all MDBs.¹¹ Under the auspices of the OECD/DAC, the MDB Evaluation Group and the UN Evaluation Group (UNEG), there has been considerable effort amongst bilaterals and multilaterals to achieve consistency of standards in evaluation work. Amongst the variety of assessments of multilateral effectiveness considered here, only evaluation work has such standards.

2.3.2 Results based management assessments

RBM approaches look at the way in which the multilateral internal operational and support systems are organised towards the achievement of development results, rather than the actual results produced. For example, if good project outcomes depend on good design quality and project supervision, then systems should be in place to monitor design and supervision. The principle source of data for RBM assessments is internal M&E reporting by the multilaterals. Most agencies have incorporated an RBM focus into their project portfolio monitoring and have now expanded these systems to include country programme monitoring. Data exist on such aspects as design, supervision, disbursement, and management of project risk. The adoption of an RBM approach is less commonly applied to support systems such as human resources (HR), budget and finance, communications etc. Only a few multilaterals have a comprehensive corporate system for monitoring a results focus throughout the organisation,¹² for example through Key Performance Indicators and a Balanced Scorecard.

¹¹ For an overview of evaluation units in a range of agencies, as well as a discussion of some of the key issues, see Foresti et al, 2007.

¹² For example, the UNDP, AfDB and the European Bank for Reconstruction and Development (EBRD)

Bilateral assessments are limited in the degree to which they can independently assess the quality of multilateral RBM systems without incurring substantial transactions costs. They can ask whether the systems exist or not, but it is difficult to assess how well they are actually implemented without intensive research. They therefore have to rely on multilateral M&E reporting for this information.

2.3.3 Assessments of partnership behaviour

The assessments of the partnership behaviour of multilaterals aim to investigate the quality of aid, as manifested in their relations with partner governments and other aid agencies, particularly at country level. These assessments have focused primarily on the harmonisation and alignment agenda, particularly the criteria set out by the Paris Declaration. A number of the methodologies go beyond these criteria to look at other aspects of behaviour. The most developed is the DRI methodology, which assesses a series of criteria including donor policies as well as donor procedures (see section 3.3.6 for more detail). Other reports (e.g. DFID's Stakeholder Perceptions Survey), have confirmed the importance of these additional criteria for partner governments and wider stakeholders (see annex 3 for more detail).

The principle methodology in these assessments is surveys of perceived behaviour carried out through questionnaires or focus group discussions. Five of the six approaches identified here – the MOPAN survey, DFID's stakeholder perception survey, multilateral client surveys and the SPA budget support survey – use this methodology.¹³

The DRI methodology differs in that it uses donors' own documents to score against its key indicators. It also examines partner governments' own aid policies and works with government officials to combine the two analyses.

Partnership behaviour surveys are particularly favoured by bilateral and international groupings because of their relatively low transactions costs, and as a way of monitoring commitments to the Paris harmonisation and alignment agenda. Client surveys by multilaterals have been rare until recently.

2.3.4 Synthetic approaches

In the past, the different types of assessments described above have been conducted in a piecemeal fashion, giving a fragmented and sometimes inconsistent picture of a multilateral's effectiveness. However, bilateral policy-makers need an overall 'bottom line' conclusion on the organisation's effectiveness to facilitate their decision-making and accountability reporting. What is needed is a synthetic approach that weaves together the different types of information in order to present a more comprehensive view about effectiveness. Such an approach would combine information from evaluations, RBM assessments and partnership behaviour surveys, and therefore such separate approaches would still be necessary. Ideally, this synthesis would be provided by the multilaterals themselves, since they have the relevant information and would probably benefit from such a

¹³ A number of the other bilateral approaches not covered in this report also survey the perceptions of their own in-country staff.

synthesis as well. However, to date there is only one such example (from IFAD). In its absence, two bilaterals have produced syntheses of multilateral effectiveness, Danida (the PMF) and DFID (the MDES), but this has required an additional burden in terms of transactions costs. In order to spread this task more widely, the MOPAN group is now moving towards a common approach for combining different types of effectiveness assessment (see 3.4.4 below).

2.4 Types of Commissioning Agent

As mentioned above, the main types of commissioning agent are: multilaterals, bilaterals working on their own, international groupings, and NGOs. The main differences between them are the purpose and use of the assessments, the resource requirements and the nature of the methodology. Although in theory, the assessment of effectiveness should be driven by both accountability and learning, in practice most assessments of multilateral effectiveness are driven by accountability concerns.

2.4.1 Multilateral evaluations and RBM assessments

These have mainly focused on operational work and have been less concerned with behavioural aspects. They primarily evolved as an accountability mechanism for their Executive Boards or Governing Bodies, and increasingly are used for internal management and organisational learning. Evaluations have had more emphasis on learning, compared with portfolio reviews, which are more used by management for internal accountability and M&E reporting. Because of the importance of both types of assessments to internal stakeholders, the resources assigned are substantial and the methodologies complex. However, the multilaterals have been less inclined to conduct client surveys.

2.4.2 Bilateral assessments

These have been driven by bilateral agencies' own internal interests and public pressure to account for their multilateral spend. They have been used as an input to internal financing decisions, to inform the development of partnership strategies regarding specific multilaterals, and to cover gaps in effectiveness reporting by the multilaterals, for example on partnership (Meier, 2007a). They have used a variety of methodologies, including some innovative approaches. However, few of these have been carried out more than once because of the transactions costs involved.

2.4.3 International groupings

These reflect both bilateral domestic accountability concerns and new international pressures to monitor the Paris commitments. They rely principally on survey methodologies, focusing particularly on partnership issues. The NGO DRI has developed a different type of assessment methodology in order to help partner governments make more informed choices about their engagement with donors.

3 Mapping Different Approaches for Assessing Multilateral Effectiveness

3.1 Evaluations

3.1.1 External evaluations

There have only been a few independent external evaluations of multilaterals, and these have usually been commissioned by the executive boards of the organisation and have focused on concessional lending programmes. The scope and methodology is usually extensive and includes a wide variety of aspects and methods.¹⁴ In 2004, an independent external evaluation (IEE) was conducted on the AfDB, led by the Institute for Development Studies, Sussex UK. It covered three replenishment periods (1996-2004) and looked at the changing context being addressed by the Bank, internal institutional changes, policy responses, programming instruments and processes, and its operational effectiveness. In 2005, an IEE was conducted on the IFAD, led by the Independent Consultants ITAD Ltd., UK. It reviewed IFAD's performance between 1994-2003 and covered the relevance of IFAD's mandate, its project portfolio performance and impact, and management performance (governance, policy and strategy development, HR, partnership, knowledge management and learning, internal management). In 2007, an IEE was conducted on the Food and Agriculture Organisation (FAO), by a team independent consultants led by Leif Christoffersen. It focused on four areas: FAO's role in the multilateral system, its technical work, its management and administration, and its governance (FAO, 2007). It should be mentioned that the OECD/DAC Network on Development Evaluation, which brings together bilaterals and multilaterals, designed a peer review process of the multilateral evaluation function in order to do away with the costly institutional evaluations funded by donors. So far UNICEF, UNDP and WFP have been examined.¹⁵

3.1.2 Internal multilateral evaluations

As mentioned above, multilateral evaluations cover a variety of operational, thematic and corporate issues. It would be difficult to single out any one of these to examine here. By way of example, however, the World Bank (WB) publishes an Annual Review of Development Effectiveness (ARDE), which presents the results of its independent assessments of project performance as well as a review of some topic related to development effectiveness (which varies each year).¹⁶ Separately, it also publishes reports on its evaluations of Bank country

¹⁴ For example, documentary reviews, surveys, interviews, focus group discussions, and country visits.

¹⁵ http://www.oecd.org/LongAbstract/0,3425,en_2649_3236398_38967446_1_1_1_1,00.html

¹⁶ For the 2006 Annual Review of Development Effectiveness see <http://web.worldbank.org/WBSITE/EXTERNAL/EXTOED/EXTOEDARDE/EXT2006ANNREV/DEVEFF/0,,menuPK:3079280~pagePK:64168427~piPK:64168435~theSitePK:3079226,00.html>

programmes and other thematic evaluations.¹⁷ Finally, it publishes an Annual Report on Operations Evaluation (AROE), which examines different aspects of the Banks M&E system as a whole.¹⁸

3.1.3 UN Joint Investigation Unit

The UN's JIU has a system-wide evaluation mandate from the General Assembly to support evaluation for the UN System as a whole. The JIU is mandated to provide an independent review, through inspection and evaluation, aimed at improving management and methods, and at achieving greater coordination between organisations. The JIU produces an average of nine reports a year (ranging from 6 to 15). These are a mix of system-wide, thematic and agency specific topics. Though this is a valuable source of independent analysis and information for the General Assembly, its current programme does not constitute an adequate system-wide evaluation function because of its limited size, scope of work and ad hoc approach to topic selection (UNEG 2007). Thus, in 2006 the High-Level Panel on System Wide Coherence recommended that a UN system-wide independent evaluation mechanism should be established by 2008 to monitor how system-wide goals are being delivered.

3.2 Results Based Management Assessments

3.2.1 Common Performance Assessment System (COMPAS)

In order to promote accountability and mutual learning, the MDB Working Group on MfDR developed COMPAS in 2004 to provide the five members¹⁹ of the group with a common framework to jointly report on their own performance. COMPAS aims to provide managers and shareholders of the MDBs with information on how they are contributing to development results and improving their contributions over time (MfDR, 2006). The COMPAS focuses on RBM practices and results, but also includes aspects of partnership and evaluation. The MDBs evaluate themselves against 30 performance indicators grouped into seven categories that relate to results orientation at the partner country level, institutional level and global partnership level. These categories are: MfDR capacity building at country level; results-based country strategies; performance-based concessional lending; project management, implementation and M&E; institutional learning (including from evaluation); results-focused HR; and harmonisation among development agencies.

Minor changes to the indicators were made to the second COMPAS in 2006 in order to increase specificity, reduce the room for discretion, and therefore increase the objectivity and credibility of the COMPAS. These indicators will provide the baseline for tracking future progress (MfDR, 2007). COMPAS is an important initiative in that it is the first case of a joint, publicly available self-

¹⁷ See the Independent Evaluation Group's website <http://www.worldbank.org/ieg/>

¹⁸ For the 2006 Annual Report on Operations Evaluation see <http://web.worldbank.org/WBSITE/EXTERNAL/EXTOED/EXTANNREPOPEEVA/0,,contentMDK:21115852~menuPK:3073384~pagePK:64168427~piPK:64168435~theSitePK:3073310,00.html>

¹⁹ The five members are as follows: AfDB, Asian Development Bank (AsDB), EBRD, Inter-American Development Bank (IADB) and WB.

assessment amongst multilateral peers. It generates data that could populate or complement other methodologies such as MOPAN and MDES (Meier, 2007b). In addition, the identification of common definitions and RBM practices amongst MDBs can provide the basis for a more general set of RBM best practices (DADA International, 2006).

3.2.2 Internal performance monitoring and evaluation systems

The managements of multilaterals monitor and report on the implementation of their projects and programmes (sectoral, country, and regional). These reports commonly include, for example, disbursement rates, project supervision ratings, likelihood of achieving intended development outcomes, projects at risk, etc. The coverage of these reports depends on the rate of compliance with internal monitoring procedures, but in theory they should cover all ongoing activities. Some MDBs also have arms length reviews²⁰ of project design and supervision quality. The WB's Quality Assurance Group led the way in this respect, with its annual reviews of the quality of project design, supervision and economic and sector work. In addition, some multilaterals commission special performance reviews as part of their replenishments²¹ or budget process. For example, the AfDB Review of Project Completion Reports (PCR) in 2006 focused primarily on operational effectiveness, driven by the necessity to improve the quality of PCRs and to strengthen the self-evaluation system of AfDB. This review synthesised a range of information by examining both quality of reports and the adequacy of performance ratings, AfDB's internal evaluation unit conducted a desk review of 42 PCRs submitted from 2003-2005 (these evaluate a project's results, implementation effectiveness and contribution to development and sustainability). In addition, a range of relevant reports such as the country strategy papers, appraisal reports, country portfolio reviews, supervision reports and other sector policy reports were reviewed (AfDB, 2006).

3.2.3 Multilateral Effectiveness Framework (MEFF)

During 2003-04, DFID established the MEFF for assessing the organisational effectiveness of multilaterals that it supports from headquarters. Its objective was to provide information for public accountability, and as an input into policy and financing decisions related to particular multilaterals. Using a RBM approach, the MEFF assesses the effectiveness of each multilateral from three perspectives: internal performance, focus on country-level results, and partnership. For each perspective, it evaluates eight corporate management systems by using a checklist with 72 questions, and then creates a scorecard using a traffic light system to score

²⁰ Additional reviews by internal staff who are not associated with the project. These reviews are separate from the self-assessments by project managers that form the basis of project reporting, but are not fully independent as are evaluations.

²¹ Examples would include the World Bank reviews of Country Assistance Strategies, and the budget implementation reports of some UN agencies (e.g. the International Labour Organisation (ILO) and World Health Organisation (WHO)). See for example the World Bank's website on IDA 15 - <http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/IDA/0,,contentMDK:21234677~menuPK:3492269~pagePK:51236175~piPK:437394~theSitePK:73154,00.html> or the African Development Bank's website ADF f11 http://www.afdb.org/portal/page?_pageid=473,10476268&_dad=portal&_schema=PORTAL

an agency's effectiveness in each area. DFID developed the system internally, with in-house staff conducting the assessments of twenty-three organisations.²²

Amongst bilateral donor methodologies, the MEFF comes closest to a comprehensive measurement approach and has been considered by other donors as a good starting point for building consensus and a harmonised approach towards effectiveness measurement (DADA International, 2006). However, it does not assess how well multilateral operational systems are implemented, or what results are actually achieved (Scott, 2005a). The framework was designed to be applicable to all multilateral organisations, but was slightly adapted for humanitarian organisations (Scott, 2005a). However, concerns were expressed that the framework did not allow sufficiently for the individual specificity of some organisations (DADA International, 2006).

3.2.4 RBM assessment (DANIDA)

In 2004, DANIDA developed a methodology to review the extent to which RBM principles form the basis of the work of monitoring and evaluation units within the assessed multilateral. The goals of the RBM assessment are to map, analyse and provide recommendations for the strengthening of a multilateral's RBM system. It seeks to identify what RBM tools are in place, as well as how effectively they are used in pursuit of improved organisational effectiveness, and efficiency at delivering development results (Meier, 2007a). The RBM Assessment is carried out by a peer panel and describes agencies' evaluation systems and policies using the DAC Evaluation Principles. For example, it assesses whether the evaluations are independent, credible and useable. Data are collected from multilateral documentation, staff interviews, on-site observations and discussion workshops. To date, DANIDA has completed assessments for UNDP, United Nations Population Fund (UNFPA), IFAD, United Nations Office of the High Commissioner for Human Rights (OHCHR) and United Nations Children's Fund (UNICEF).

Potentially the RBM Assessment could lessen, or even eliminate, reliance on external evaluations of multilaterals, though this will require support from a broad range of donors (DADA International, 2006). It serves the dual purpose of generating credible assessment information, while focusing management's attention on areas of critical long term importance to bilateral objectives. Thus, it has good basis for evolving into a common approach for assessing multilateral effectiveness (Meier, 2007a). However, donors still recognise the need for rigorous attention to the definition of results and an international institutional arrangement for objective verification purposes (DADA International, 2006).

²² AfDB, AsDB, EBRD, European Commission (EC), FAO, HABITAT, IADB, International Committee of the Red Cross (ICRC), IFAD, International Federation of Red Cross/Red Crescent Societies (IFRC), United Nations Office for the Coordination of Humanitarian Affairs (OCHA), OHCHR, Joint United Nations Programme on HIV/AIDS (UNAIDS), UNDP, United Nations Educational, Scientific and Cultural Organisation (UNESCO), UNFPA, United Nations High Commissioner for Refugees (UNHCR), UNICEF, United Nations Industrial Development Organisation (UNIDO), United Nations Development Fund for Women (UNIFEM), World Food Programme (WFP), WHO, World Bank

3.3 Assessments of Partnership Behaviour

3.3.1 Multilateral client surveys

To-date UNDP is the only multilateral that systematically seeks feedback on its performance from donors, partner governments, civil society organisation, through its annual stakeholder perceptions survey.²³ Other multilaterals have used client surveys on a sporadic basis (for example, as part of country programme design), and are usually country specific.²⁴ None of these are in the public domain.

3.3.2 Multilateral organisation effectiveness stakeholder perceptions survey

In 2007, DFID commissioned a pilot project to identify the views of stakeholders in recipient countries about the performance of key multilateral organisations, and their preferences for which organisations should disburse additional aid. Perceptions were sought through a locally administered questionnaire. This was completed by well-informed individuals from five stakeholder groups: business leaders, civil servants, civil society leaders, government ministers, and Members of Parliament. The survey assessed multilateral performance according to three measures from the Paris Declaration: overall development effectiveness, harmonisation and alignment. In addition, respondents were asked to rank the multilateral against fifteen performance criteria related to funding and policies and procedures (adapted from the DRI criteria), and to rank the organisation in the order of preferred disbursement channel.

Until this survey, there had been no systematic studies to seek the views of the range of stakeholders who interact with donors in-country and are responsible for implementing donor programmes. Limitations of the survey include its short data collection period and an overly long questionnaire. The survey attempted to minimise criticisms of bias and subjectivity by careful choice of respondents. Almost uniquely out of the performance assessments, data were subjected to statistical testing to ensure robust results (Burall et al, 2007).

3.3.3 MOPAN Survey

Since 2003, a network of eleven countries has conducted an annual survey of 3-4 selected multilaterals in 8-10 countries where MOPAN members have bilateral programmes. MOPAN members use the findings of the surveys for a range of reasons relating to: their own accountability for multilateral financing; as input into their policy towards multilaterals; to strengthen their relationship with multilaterals in governance and joint activities; and to contribute to the wider debate of aid effectiveness (MOPAN 2006). The MOPAN survey assesses the quality of multilateral partnership behaviour towards national stakeholders and other international development agencies at the country level. Partnership is assessed through perception surveys completed by MOPAN member country-level staff and a collective focus group discussion. As a final output, MOPAN

²³ See for example - <http://www.ke.undp.org/PartnershipsurveyKenya.pdf> - though global results do not appear to be publicly available on the internet.

²⁴ However, it is important to note that the World Bank did systematically monitor its partnership behaviour as part of the Comprehensive Development Framework.

presents and publishes an overall synthesis report that draws from the various country level surveys, giving an opportunity for the assessed multilateral to provide feedback. To date, the MOPAN Survey has released four synthesis reports assessing eleven multilaterals.²⁵

The MOPAN Survey represents a positive step towards uniting bilateral donor efforts in assessing multilaterals. It fills an important gap in understanding how multilaterals actually operate at the country level, based on bilateral experiences of working with them. To date, the MOPAN surveys provide the most important regular and publicly available source of information on multilateral partnership behaviour at country level. Because of the importance of the bilateral grouping represented by MOPAN, the surveys have been important in focusing the attention of the multilaterals on this aspect of their effectiveness. Another benefit is its design as a light and rapid exercise with minimal transaction costs, which allows for easy replication and sustainability (Meier 2007a). The methodology is currently being expanded to include other approaches (see 3.4.4)

3.3.4 Paris Declaration Survey

In the Paris Declaration, donors and partners agreed to monitor progress in improving aid effectiveness against 56 specific actions, from which 12 indicators were established and targets set for 2010. This survey was designed to generate an accurate picture of existing aid practices as a baseline for assessing progress towards the 12 indicators.²⁶ It is also intended to stimulate a broad-based dialogue on how to make aid more effective and promote agreement on specific actions that contribute to successful implementation of the Paris agenda at country level (OECD/DAC, 2007b). The Survey is managed at country level by a national co-ordinator appointed by the government and supported by donors when appropriate. The survey consists of two questionnaires sent separately to in-country donors and government which asks for both quantitative information and qualitative commentary about donors (bilateral and multilateral) relating to the relevant Paris Declaration indicators. These questionnaires are then compiled into a report at the country level.

The Baseline Survey report published in 2007 is based on evidence collected from a survey of activities in 2005 for 34 self-selected countries and 60 donor organisations. Two main constraints were identified in the survey process (OECD/DAC, 2007b). One was the separate treatment of the two main parts of the survey, which disconnected broad areas of the Paris commitments – aid delivery on the one hand and country systems on the other. The second was the broad scope of the guidance, which allowed for variations in defining the indicators in each country; this diverted discussions from factual into definitional issues. Some stakeholders also found the Baseline Survey process to be too onerous and time-consuming (OECD/DAC, 2007b). Nevertheless, more is now

²⁵ AfDB (twice), ADB, IADB, ILO, FAO, UNAIDS, UNDP (twice), UNPFA, UNICEF, World Bank, and WHO (twice)

²⁶ Indicators related to the following principles: ownership, harmonisation, alignment, results' management and mutual accountability

known about aid management systems and practices at the country level as a result of the Baseline Survey. In addition, it has contributed to generating dialogue on aid effectiveness and how to improve national monitoring efforts. Rather than discussing the multilaterals separately as a group, the report provides information on the performance of individual bilaterals and multilaterals as well as overall data relating to the Paris Declaration indicators for eight specific multilaterals.²⁷ Another survey is being conducted as this report is being written, to be presented to the High-Level Summit in Accra in September 2008. A final survey will be carried out prior to the planned Summit in 2010.

3.3.5 SPA Budget Support Survey

Since 2003, the Budget Support Working Group of the SPA has engaged in an annual monitoring exercise that tracks the progress made by bilaterals and multilaterals in aligning and harmonising their budget support activities in a number of African countries. The SPA Budget Support Survey consists of questionnaires sent to development agencies and government representatives to obtain perspectives on general and sector budget support practices with regard to alignment.²⁸ In each country, a lead agency is identified with responsibility for organising the completion and return of the questionnaires. The survey evaluates the predictability of budget support, the links between budget support and Poverty Reduction Strategy (PRS) processes, support to public finance management and government views on donor behaviour in budget support relationships. This is carried out in a way that ensures that data from earlier surveys can be compared with the most recent data. In the past four years, 21 agencies²⁹ in 19 PRS countries have participated in the survey.

The process of responding to this survey has provided some important opportunities for dialogue amongst agencies and national partners. In addition, the results obtained from the SPA Budget Support Survey represent the most comprehensive data currently available on the progress of budget support alignment with the PRS and has potential for trend analysis (SPA, 2007). At the same time, it is acknowledged that there was considerable overlap between the SPA Budget Support Survey and the 2006 Paris Declaration Baseline Survey processes, and some countries experienced increased transactions costs as a result. However, it should be noted that whereas the Baseline Survey collected data across the broad range of Paris Declaration indicators, the SPA survey investigates more deeply the mechanics and relationships of budget support.

3.3.6 Debt Relief International

DRI prepared a methodology in 2004 to enable Heavily Indebted Poor Country (HIPC) governments to assess the quality of aid they receive as part of the process of designing overall national debt and new financing strategies. The goal of the methodology is to help HIPC governments identify which are the 'best' external

²⁷ AfDB, AsDB, EC, GAVI Alliance, Global Fund for Aids, TB and Malaria (GFATM), IADB, United Nations, WB

²⁸ In 2006, the survey also covered sector budget support.

²⁹ Multilateral donors include AfDB, EC, UNFP, and WB

financing sources to fund their national development programmes, and how they can improve the quality of all their external finance. Looking at both the quantity and quality of external assistance flows to African governments, the DRI framework analyses the policies and procedures of donors as well as HIPC governments. Table 2 lists the criteria for analysing donor policies and procedures.

Table 2: DRI criteria for assessing donor aid quality (source Johnson et al, 2004)

Donor policies	Donor procedures
<p><i>concessional</i>ity, as measured by the grant element;</p> <p><i>types of assistance</i>, budget/balance of payments support, project, TA or food/commodity;</p> <p><i>channel of assistance</i>, via the budget or off-budget; sectors and projects, support is for PRSP or donor-led priorities;</p> <p><i>flexibility of assistance</i>, can aid be channelled to new areas and/or used to meet unexpected financing gaps arising from economic shocks;</p> <p><i>predictability of assistance</i>, multi- or one-year commitments and whether they are fulfilled; <i>policy</i></p> <p><i>conditionality</i>, the number and degree of enforcement of policy conditions;</p> <p><i>policy dialogue</i>, the degree of donor engagement in the economic policy dialogue and independent support for government policies or alignment with Bretton Woods Institutions.</p>	<p><i>conditions precedent</i>, types and degree of enforcement and the disbursement delays arising in meeting these conditions;</p> <p><i>disbursement methods</i>, via cash direct to the budget or reimbursement claims;</p> <p><i>disbursement procedures</i>, the number and complexity and disbursement delays arising;</p> <p><i>procurement procedures</i>, the number and complexity and disbursement delays arising;</p> <p><i>co-ordination</i>, the degree of harmonisation between government and donors/creditors</p>

In the first part of the survey, Finance Ministry officials from each HIPC government assign each donor a score for 23 different evaluation criteria on the basis of objectively verifiable thresholds. The scores are aggregated to identify which aspects present the worst problems, and which donors have the best/worst quality aid from the perspective of the each government. The second part provides for the analysis of the government's own policies and procedures and the impact these have on the delivery of external assistance flows. By combining these two aspects of the survey, HIPC governments have a useful source of information to use to identify donors to provide additional aid.

Unlike most other assessments, the DRI methodology is led by HIPC governments themselves and systematically monitors whether agencies are living up to their pledges on aid quality and support for the PRSP framework consistently across a range of African countries. This assessment has only been

published once. While individual country results are not published, DRI has published the aggregate scores across all of the participating countries (after stripping out the country specific weightings). These aggregate scores cover 10 multilaterals³⁰ and 18 bilaterals (Johnson et al, 2004).

3.4 Synthetic Approaches

3.4.1 Report on IFAD's Development Effectiveness (RIDE)

IFAD has recently produced the first annual report on its development effectiveness, or RIDE (IFAD, 2007). The purpose of the RIDE is to assess the relevance of its mandate and its organisational effectiveness and efficiency. Its scope is broader than development effectiveness reports from other multilaterals in that it is not based exclusively on evaluation data. It also incorporates portfolio monitoring information, internal corporate performance monitoring, the budget, and recent initiatives relating to the broader international harmonisation and alignment agenda. The RIDE is unusual in other respects too: it assesses the relevance of its role and mandate within the changing international context. IFAD has rolled out a comprehensive corporate process to integrate an MfDR approach into all its internal operational and support systems, with quarterly monitoring of corporate key performance indicators. The RIDE reports on this process and the first results of the key performance indicators. As such, the RIDE provides a brief, synthetic report on its organisational and development effectiveness, from project to corporate level (IFAD, 2007).

3.4.2 Performance Management Framework (PMF)

In 2003, DANIDA developed the PMF as part of a larger internal reform of the management of Danish development cooperation. The objectives of PMF are to: enhance the quality of Danish development cooperation; improve management and continuous learning; and strengthen accountability. It serves internal management purposes by providing timely and reliable information on progress of activities, and it ensures accountability and reporting to external stakeholders.

The PMF uses a series of tools that combine qualitative and quantitative information gathered at three levels (DANIDA corporate, multilateral headquarter and multilateral country level) to assess multilateral development cooperation. Data sources include multilateral strategies, perception analyses (by its own embassy staff and through MOPAN), and assessments of multilateral organisations' evaluation and reporting systems. In the measurement of both organisational practices and field level results, the PMF relies on the objectives, targets and indicators established by the relevant organisations themselves as set out in the organisation's own vision and strategies (DADA International, 2006).

The benefits of the PMF are its comprehensiveness – it solicits views from multiple levels – and its dual purpose of measurement and management. The PMF systematically focuses on results with a view to providing performance

³⁰ AfDB, Arab Bank for Economic Development in Africa, EU, European Investment Bank, IDA, IFAD, IMF, Islamic Development Bank, OPEC, UN Agencies

information for management decisions that optimises value for money and ensures the prudent use of human and financial resources in support of the overall objective of poverty reduction (DANIDA, 2005). However, its complexity adds to the time and resources needed to conduct the evaluation for each multilateral (Meier, 2007a). DANIDA completes a cycle of data collection each year for the three most important multilateral in each country across 15 countries.

3.4.3 Multilateral Development Effectiveness Summaries (MDES)

In 2006, DFID devised a system to collate the variety of existing information on multilaterals and to present it within a coherent framework called MDES. The MDES are designed to inform DFID's management decision making process and to ensure senior managers have a single, consistent picture of each multilateral (Meier, 2007a). They provide a general overview about how multilaterals are performing across four key dimensions of effectiveness: managing resources; managing relationships; country/global results; and building for the future. This information is organised in the form of a "balanced scorecard", giving an integrated picture of how organisations are performing as well as their strengths and weaknesses. The MDES include basic data on the organisation, the context in which the organisation is working, as well as a summary assessment of effectiveness. The summaries are compiled by DFID desk officers and are based on published secondary data sources including agency reports, donor evaluations, perception surveys, NGO reports, and the Paris Declaration monitoring exercise. In December 2007, DFID published MDES for fourteen multilaterals.³¹

Because the MDES rely on the collation of existing data that may be collected in different ways and using different definitions, there are issues of comparability both across organisations and also across time. The MDES are not designed to produce a ranking of multilateral by their effectiveness nor intended to measure actual development impact (DFID, 2006). Also, given the variety of data sources used, the MDES are complex and more time-consuming to complete than other methodologies. Despite their level of complexity, the MDES has potential for evolving into a common approach for assessing multilateral effectiveness which focuses explicitly on accountability (Meier, 2007a). They represent the first attempt to compile data from multiple sources to assess multilateral internal performance and country level relationships and results.

3.4.4 Revised MOPAN Approach

By 2006, there was consensus that MOPAN members required further evidence on the effectiveness of individual multilateral beyond the Annual Survey. In addition, significant overlap between the various assessment approaches (both individual bilateral approaches and others) was identified and therefore that there was scope for harmonisation. To address these issues, MOPAN members are considering the introduction of the MOPAN Common Approach - Balanced Scorecard, which is based on a common frame of reference focused on strategic,

³¹ AfDB, AsDB, EBRD, FAO, GFATM, ICRC, UNAIDS, UNDP, UNESCO, UNFPA, UNHCR, UNICEF, WHO, World Bank. These are available from <http://www.dfid.gov.uk/news/files/assess-multilateral-effectiveness.asp>

operational, relationship and knowledge management categories. At this stage 35 key performance indicators have been identified, but efforts are currently underway to focus and reduce the number. Data will be collected from multilateral self-assessment checklists, MOPAN member surveys and government questionnaires at both headquarters and country levels. Secondary data collected from existing sources and multilateral reports will also play an important role in populating the MOPAN Balanced Scorecard. MOPAN makes explicit that the aim is to use existing information within the organisations RBM system. The approach is not to collect more primary data but to pull together existing data. MOPAN proposes to conduct assessments of up to six multilaterals along with one RBM organisational assessment from 2009. It is the intention that the coverage should be comprehensive within a short span of time, thereby eliminating the need for members' own assessment systems.

4 What can the Different Approaches tell us About Effectiveness?

In this section we explore some methodological issues which affect the certainty with which we can view the findings of effectiveness assessments in general, before looking at some of the strengths and limitations of particular approaches. In the final part of this section we highlight some of the major gaps that currently exist when looking at effectiveness assessments as a whole.

4.1 *Methodological Issues*

All assessments of effectiveness face difficult methodological problems, which ultimately can affect the credibility and therefore the usefulness of the exercise. Two issues should be mentioned, in particular: subjectivity and comparability. Both of these would be resolved if there were internationally agreed standards for these assessments. However, apart from the OECD/DAC Evaluation Quality Standards, such standards are lacking.

4.1.1 *Principles and standards of assessment*

All forms of assessment need to conform to certain standards and principles, particularly regarding data quality and avoidance of bias in order to ensure that they are credible. Without credibility, the assessment results will be less useful in feeding into organisational learning, or in being used as part of an accountability function.

As mentioned above, evaluations have explicit standards and principles regarding the design, the robustness of the evidence gathered, the conduct of evaluation work, and transparency. Although there may be some debate about the degree of independence of some multilateral evaluation offices from management (especially for UN agencies which don't have independent evaluation offices) there are agreed evaluation norms and standards. The closest that RBM assessments come to agreed standards or principles is the five principles of MfDR which were agreed at the Second Roundtable on MfDR in 2004. However, these are too general to provide a set of quality standards.

There are no agreed standards and principles for the type of perception based survey that is most commonly used to assess the partnership behaviour of multilateral organisations beyond that of good social survey practice. It is not surprising therefore that these types of approaches are sometimes criticised for being biased, subjective, and superficial.³²

³² See for example the organisational responses to the MOPAN surveys - <http://www.acdicida.gc.ca/CIDAWEB/acdicida.nsf/En/JUD-5292536-HRK>

4.1.2 *Subjectivity*

As already mentioned, the primary method for assessing multilateral partnership behaviour, except in certain aspects of the Paris Survey, is to investigate the perceptions of the participants in the relationship or event through the use of a client or partner survey. By necessity therefore, assessments of partnership, ownership, and participative orientations are always going to be subjective.

However, good survey design can ensure good quality attitudinal data. Key issues relating to the quality of surveys includes: sample size, sample quality (i.e. how knowledgeable the respondents are), question design, response rates, adequate piloting, and careful presentation of results. It is also important that the questions are valid, i.e. if the indicators selected are good measures of performance and the question wording is unambiguous.

The use of staff surveys to elicit views about RBM systems is much more questionable, because people outside of the multilaterals themselves are unlikely to have detailed knowledge of such systems, especially at country level. Few country level respondents are likely to be familiar with the HQ RBM systems, and may lack detailed knowledge of multilateral projects and programmes at country level.³³ Even the HQ bilateral staff responsible for multilateral relations may have patchy knowledge of some aspects of a multilateral RBM system.³⁴ Moreover, there is always an element of judgement involved in the document reviews typically used in RBM assessments, especially if this involves a rating scale. For this reason, it is important to have adequate guidelines and quality control in the use of such scales.

It is not appropriate to contrast quantitative and qualitative information in terms of the degree of subjectivity. Statistics can convey an impression of objectivity, but if the data are based on individual perceptions, they are still subjective. Quantitative data can be more reliable if there are large numbers of informants, but qualitative data can yield more detailed information. They are complementary methods.

It is easier to reduce subjectivity in RBM assessments based on documentary analysis. For example, by focusing on whether a certain RBM practice was in place or not or was under development, the MEFF was able to insist on an objective, factual approach. However, an assessment of the *quality* of an RBM practice would require more subjective judgement. There is of course an element of subjectivity in almost all assessment exercises. Even independent evaluations (which are often perceived as ‘objective’ because they are independent) usually contain survey data, as well as interviews and site visits, which will involve subjective interpretation. The issue is not to avoid subjectivity; but rather to use adequate quality controls.

³³ Early rounds of the MOPAN survey found low response rates from Embassy/Delegation staff on RBM questions, and these questions were deleted in later rounds.

³⁴ This was revealed during the completion of the MEFF checklists. Desk officers were most knowledgeable about aspects that related to their institutional strategies, but less so in other areas where they had less direct contact.

4.1.3 Comparability

By their very nature, bilateral assessments imply comparisons between multilaterals. However, only two of the approaches reviewed by Meier, CIDA's MERA³⁵ and DFID's MEFF, made explicit inter-agency comparisons (Meier, 2007a). Even where these are not explicit, the use of effectiveness ratings makes comparisons implicit, and if this information is made public, other actors will make the comparisons even if the commissioning agent does not.

There are two main issues relating to multilateral comparisons. The first issue concerns whether multilaterals are themselves comparable, given different structures and functions. The second issue relates to whether the results from different studies are comparable, given differing methodologies.

There are important differences between agencies in terms of mandate, governance structure, role, method of funding etc. These differences can limit comparability. The systematic approach provided by the MEFF revealed unexpected details as to these differences. For the analysis, DFID tried to deal with these by grouping the organisations according to their different functions (Scott, 2005b).³⁶ However, even within these groupings there were still important differences.

The DFID-commissioned survey of stakeholder perceptions took a much cruder approach to comparison, splitting the organisations into two groups: those that have broad portfolios and are able to channel large sums of money, and the more specialised agencies (Burall et al, 2007); and this too had its limitations. The MOPAN questionnaire has been revised several times over successive years in order to adapt to differences between different agencies, and by 2005 a separate questionnaire tailored to each agency had been produced.³⁷

Methodological differences are a major impediment to comparisons between multilaterals. Even amongst similar multilaterals, there are different definitions of key performance terms, as the COMPAS exercise and the Paris Baseline survey found (see sections 3.2.1 and 3.3.4 above). Different definitions make it difficult to compare the results of different approaches in order to triangulate the results and check for bias. For example, both the MOPAN survey and the stakeholder perceptions survey assess perceptions of performance of different agencies against a number of the Paris Declaration principles. Yet while the questions ask about the same general principle, such as ownership, the specific wording in both questionnaires makes valid comparison very difficult.

Reports carried out in different time periods may also limit comparability. Variations in the timeframes of multilateral assessments are very likely because of

³⁵ Not covered here because neither its methodology nor results are made public.

³⁶ For example, MDBs, UN Specialised Agencies, UN Funds and Programmes, Humanitarian Agencies, Coordination Agencies, and the EC.

³⁷ See the annual synthesis reports for copies of the questionnaires - <http://www.acdi-cida.gc.ca/CIDAWEB/acdicida.nsf/En/JUD-5292536-HRK>

the ‘once-off’ nature of many of the assessments, and are particularly problematic in terms of RBM practices which can change rapidly from year to year.

4.2 Assessing the Strengths and Limitations of the Different Approaches

In this section we now highlight some of the general strengths and limitations of the different approaches. Details for particular studies within each approach are provided in Annex 2.

4.2.1 Evaluations

Evaluations are critical techniques for generating substantive information about outputs, outcomes and impacts. Carried out well, they can be useful for the purposes of management, governing boards and key stakeholders. Evaluations can ask questions about efficiency and effectiveness, and help to cast light on the critical issue of the counter-factual; what would have happened without this intervention?

Although the quality of evaluations is generally high, they are usually time-consuming, resource-intensive and limited in their coverage of the multilaterals’ total operations. Since most of them are ex-post studies, there is also a lag between the project life and the evaluation results. While regular project monitoring is an ongoing function that generates a continual stream of data on operations, the coverage of evaluations is discontinuous and much more limited in scale. In small organisations, this may affect the representativity of the sample evaluated. Independent external evaluations, of the kind described above, are very rare because of the resources required.

4.2.2 Results based management assessments

The major attraction of RBM assessments is their comprehensive and factual nature. However, there are some disadvantages. First, the methodology is necessarily complex and resource intensive, because RBM systems are themselves complex. Second, they tend focus on the headquarters and give less account of what happens at country office level.³⁸ Third, they do not cover behaviour and practices, which may diverge from what systems predict. Finally, since RBM systems are always changing, the approaches rapidly become out of date.

Assessing the *quality* of multilateral MfDR systems is difficult. It requires clear criteria against which judgements can be made. Good practice benchmarks have been provided by the UN Joint Inspection Unit (Joint Inspection Unit, 2004) and specialist RBM consultancy firms have their own frameworks for assessing the RBM systems of a variety of private and public sector firms.³⁹ However, there is no established consensus amongst development agencies themselves regarding

³⁸ However, a few UN organisations (WFP, UNDP) have performance indicators at country level, which are reported on annually.

³⁹ Some multilaterals (ILO, UNDP) have commissioned the JIU and external consultants to evaluate their RBM systems.

best RBM practice.⁴⁰ It is unlikely therefore, that the subjective opinions of bilateral staff would be a good basis for evaluating RBM quality.

4.2.3 *Partnership behaviour*

Perceptions surveys are virtually the only way of obtaining information about multilateral *behaviour*, which, as already mentioned, is an important aspect of aid quality. However, they do not measure outputs and outcomes and thus give limited information about development effectiveness. Given the nature of these methodologies, there is a risk of bias and concern about subjectivity. The guidance given to those completing the questionnaires for partnership behaviour surveys is critical in reducing subjectivity, and was a major weakness in the Paris Baseline Survey.

Despite a number of challenges, these methodologies are probably the easiest and quickest to administer. However, they imply transactions costs for those completing the surveys, especially for government officials or other key stakeholders whose time is scarce. This concern is magnified because a number of the surveys cover some of the same ground, with the same government stakeholders (particularly the Paris Baseline and SPA surveys).

4.2.4 *Synthetic approaches*

Synthetic approaches would not replace the separate approaches outlined above; rather they would represent an additional effort to pull together the existing information from these different sources. However, they are not without problems. By their very nature, bilateral synthetic approaches require significant staff time and resources to digest and produce clear summaries of the large amounts of source material. Some subjectivity will be involved in summarising these data. They are also very dependent on the quality of the original source material. Synthetic approaches also suffer from problems of non-comparability of different assessments due to differing methodologies and the time lags between them.

DFID's MDES offers perhaps the best attempt to integrate all publicly available knowledge in one place within a schema, the balanced scorecard, which synthesises the information in an easily digestible way. Yet it suffers from some of the difficulties mentioned above, especially combining data from different sources, with different definitions and time periods. It is also very time consuming, and would be better to be carried out collectively by a group of donors. It is encouraging therefore that the MOPAN group are currently considering adopting a similar approach (Meier, 2007b). By settling on one assessment methodology which combines a significant proportion of the available information would reduce the transaction costs which occur at the moment as a result of multiple, overlapping assessment methodologies (Scott, 2005b). Ideally though each multilateral would provide a synthesis containing this information itself.

⁴⁰ The DAC sponsored MfDR principles are too general to form a basis for RBM evaluations.

4.3 Gaps

It is clear from the descriptions above that the different assessment approaches are extremely variable in the information they generate, the purposes to which they are put and their potential to add to knowledge about multilateral effectiveness. In the final part of this section we highlight a number of gaps before moving on in section 5 to make suggestions for ways forward for the assessment of multilateral effectiveness.

4.3.1 *Where is the information about relevance?*

As defined above (section 2.1.9), there are three aspects, relevance to: bilateral donors own priorities; the MDGs; and the multilateral's own mandate. The issue of relevance to bilaterals' own priorities is a policy matter for each bilateral, and should not drive any assessment other than their own. Multilaterals should respond to priorities that are collectively agreed amongst the shareholders through their governance mechanisms and not to the priorities of individual donors. However, the other two aspects of relevance (to MDGs and to mandate) are critical for their partner countries, the financiers of multilaterals as well as for the efficiency of the global aid architecture. The issue of relevance to the MDGs is primarily addressed by multilateral evaluations, but not by other assessment approaches. The question of relevance to mandate is crucial to the efficient division of labour of the international system, yet only IFAD's RIDE report explicitly addresses it. It should be noted however, that it would be difficult to assess relevance to MDGs and to mandate through the current methods as it would require detailed knowledge of the content of a multilateral's programmes.

4.3.2 *Where is the information about comparative advantage?*

All donors should be able to identify the sectors, countries or thematic areas where they will have the greatest impact, if acting alone. This information about their *absolute advantage*, will tell them nothing about how to operate in a country or sector where there are multiple actors. For this they will need information about their *comparative advantage*.⁴¹ None of the approaches makes any attempt to assess multilateral performance in relation to a division of labour based on comparative advantage. Yet again, this information is critical for both bilateral donors and partner governments attempting to make decisions about which multilateral agencies to engage with, and it is also important for global aid efficiency. The multilaterals would need to provide a lead in defining and measuring their comparative advantage, but they have not been inclined to do this so far.

4.3.3 *Where is the information on governance and management?*

Theoretically, governance mechanisms should play a key role in steering the performance of the multilaterals. However, it is well known that these mechanisms are weak and indecisive. Executive boards and governing bodies are prone to micromanaging the institution and are often inconsistent in the advice

⁴¹ For a more detailed discussion of the issues of comparative vs absolute advantage see the companion paper to this one (Graves et al, forthcoming)

they give.⁴² This makes it difficult to manage an institution coherently. There have been few studies of the governance and management of the multilaterals and none of the assessment approaches apart from the MEFf have covered it.⁴³ Again, it would be difficult to include governance and management with conventional assessment methodologies, but it would be worth commissioning special reviews as an input to an analysis of multilateral effectiveness.

4.3.4 What is the role of partner governments in the assessment of multilateral effectiveness?

Of the assessment approaches considered here, the DFID stakeholder survey, the Paris Declaration Survey, and the SPA and DRI surveys have involved partner governments. While it is important to ensure that they have a voice in such assessment exercises,⁴⁴ we also highlight that partner countries have information needs which are different to those of donors and these are currently not be met in a systematic way.

The above analysis shows that none of the current approaches are able to give a full picture of multilateral effectiveness on its own and all of them have strengths and weaknesses. The development of quality standards would help to address some of these problems, but even then a more accurate picture of multilateral effectiveness would require some gaps to be addressed and a wider range of approaches to be combined. However, the transactions costs of a more comprehensive approach, especially for external agencies such as bilaterals, would be considerable.

⁴² See for example, the comments of IFAD's IEE on the ineffective decision-making of its Executive Board (Report on the Independent External Evaluation of IFAD, August 2005. EB 2005/84/R.2/Rev.1 pp80-81

⁴³ But DFID considered that the governance questions in the MEFf checklist were weak (Scott 2005b)

⁴⁴ The MOPAN survey has been criticised for not involving partner governments for example.

5 Ways Forward for Multilateral Effectiveness Assessments

In this section we go beyond the description and review of the different assessment approaches to explore the implications for the ways forward for the assessment of multilateral effectiveness. We consider:

- Possible reliance on multilateral assessments and eventual abandonment of separate bilateral assessments;
- Whether there is a continued need for bilateral assessments in the short term;
- Ways to meet the information needs of partner governments; and
- Ways to use effectiveness assessments to improve the overall aid architecture.

5.1 Relying on the Multilaterals to Assess Effectiveness

Even as multilaterals have been increasing their effort to monitor their own effectiveness, there has been an increasing proliferation of bilateral assessments of multilateral effectiveness. Ideally, it should not be necessary for bilaterals to conduct these assessments because this is not the most efficient way of obtaining the information. Whether carried out by individual agencies or as part of an international grouping, bilateral assessments are problematic because:

- They are carried out by a number of agencies in an unharmonised way, risking duplication and increasing transaction costs, for both bilaterals and multilaterals;
- It is difficult for an external actor to gain information about internal processes such as RBM systems. Efforts to do so are very time and resource intensive; and
- It is more appropriate for the multilaterals themselves to define, monitor and improve their own effectiveness, thereby improving the overall effectiveness of the international system.

Rather than seeking to assess multilateral effectiveness themselves, bilateral donors should obtain the necessary information on effectiveness from reports produced by multilaterals. IFAD's RIDE is provides a good demonstration of a possible way forwards. Yet in the short run, bilateral assessments will continue to be necessary because of the additional information requirements of bilaterals (see section 5.1).

Multilateral agencies are responsible for reporting on their own performance for accountability and management purposes, and a large number of reporting instruments exist. Yet this reporting is fragmented and its quality is variable, with large differences in the coverage, regularity, quality and transparency of reports.

Despite some improvements, particularly by the MDBs, many multilateral reports are still primarily activity and input oriented. Most only provide qualitative descriptions of their activities rather than a systematic reporting of results. Many multilaterals do not systematically track the outputs and likely outcomes of their projects or have corporate RBM targets to be monitored and reported against. Only a few systematically track their commitments to the harmonisation agenda. As members of the governing boards of the multilaterals, bilaterals should seek improvements in reporting in all these areas to assist with the challenge of establishing comparable information among agencies.

However, there is some evidence that bilaterals do not make good use of effectiveness information even when it is available, because of a lack of clarity over their own information needs (Kabell and Balogun, 2004). The root of this problem is the current lack of consensus over what exactly is required in order to assess effectiveness and the absence of common standards for assessing multilateral effectiveness. Agreement is needed on the basic definitions, minimum criteria for the assessment of RBM systems and good practice guidance for partnership survey design. This could proceed by distilling the experience of existing assessment approaches, rather than creating new ones. Such standards would help guide demands made of the multilaterals for their internal performance reporting.

Bilaterals are now considering more collective and comprehensive approaches towards assessing multilateral effectiveness, for example through the new MOPAN Common Balanced Scorecard approach. Such a development is to be welcomed as it would reduce duplication and could form the basis for the development of consensus on information needs, agreed quality standards, and common advocacy and influencing objectives as regards effectiveness assessment.

However, it is conceivable that bilaterals and partner countries would still want to conduct separate assessments, at least in the short run until multilateral reporting had improved. Given the transactions costs and other inefficiencies implied, it is worth examining further the rationale for these separate assessments.

5.2 *Bilateral Assessments of Multilateral Effectiveness*

5.2.1 Purposes of bilateral assessments

Bilaterals carry-out assessments for a variety of reasons, including:⁴⁵

- To inform multilateral policy, institutional strategy and financial allocation decisions;
- To exercise accountability and reporting to national constituencies, and to justify multilateral spending;
- To strengthen participation in the governance of multilateral organisations;
- To strengthen relationships with multilateral organisations at country level;
- As input to joint advocacy by like-minded bilateral groups such as Utstein;

⁴⁵ This list is adapted from Obser, 2007

- As input into wider debates about aid effectiveness;
- To assess the relevance of the multilateral to their bilateral priorities;
- To cover gaps in reporting, particularly on multilateral behaviour; and
- To provide a synthesis of performance drawing on a variety of sources.

Most of these reasons relate to internal policy and decision-making, strongly linked to domestic accountability pressures and decisions on financial allocations to multilaterals. This is no doubt related to the high proportion of bilateral ODA that is allocated to multilaterals.⁴⁶ The internal orientation of these motivations is demonstrated by the low transparency of the assessments, only a few of which are publicly available. However, there is some doubt about whether the bilateral assessments carried out do in fact fulfil the original purposes that were intended.

5.2.2 Bilateral financing decisions

One of the key reasons for carrying-out multilateral assessments is to inform bilateral allocation decisions. This would require the following conditions:

- The ability to quantify the relationship between spending and ‘results’;
- The room for manoeuvre dictated by financing modalities for different multilaterals and extent of flexibility to either expand or contract bilateral programmes;
- The degree of ‘fit’ between the mandates of different multilaterals and thematic or regional objectives of bilateral donors and, related to this, effectiveness of multilaterals relative to bilateral programmes; and
- The extent to which bilateral influencing objectives are linked to the level and terms of financing provided to multilaterals.

However, the ability to use the information on effectiveness to address these issues is limited for the following reasons:

Quantifying the relationship between spending and ‘results’

The ability to quantify the relationship between spending and development results implies the possibility of attributing results to particular agencies and providing some comparability among them. As discussed above, both are problematic first, because of differences in the mandates, functions and deliverables of the multilaterals, and second because it is very difficult in practical terms to attribute results directly to individual agencies.

Room for manoeuvre

The room for manoeuvre available to bilaterals in deciding where to channel resources varies greatly. One constraint common for all bilateral donors is that replenishment rounds for multilateral organisations are a series of disconnected events. For example, the replenishment of the MDB funds is tied to three year cycles which loosely coincide with each other, while non-thematic funds, as well as

⁴⁶ While about 70% of ODA is provided through bilateral agencies and 30% through multilateral agencies as a whole, there is considerable variability among bilateral agencies in relation to their bilateral : multilateral contributions, which ranged from 9 to 64 % for the 2000-2005 period (IDA, 2007)

global funds such as GFATM, are replenished relatively frequently. The Global Environment Facility is on a four year cycle, and the European Development Fund budget is set to 2013 (though there are also opportunities to provide supplementary funding). These disconnected events, together with their frequency, risk sidetracking donors from taking a step back in order to consider the overall balance between their multilateral and bilateral spending (Burall and Maxwell, 2006).

Bilateral donors also respond to a series of domestic pressures (Martens, 2005) which affect their political priorities, and further limit their room for manoeuvre in relation whether or not they change their financial allocations to individual multilateral agencies.

In short, bilaterals take funding decisions on a piecemeal agency-by-agency basis at different moments in time. This contributes to the complexity of the decision-making process which bilaterals face. Multilateral assessments could have an important role in informing these sequential decisions about allocation, but may have limited traction if the overall envelope being made available for multilateral financing by bilaterals is set on a less systematic or even ad hoc basis.

Mandate and effectiveness

The different mandates of multilaterals are critical to financing decisions. For example, the outreach of MDBs provided by their global or regional mandates provides bilaterals with a useful means of contributing resources without the need to establish a country presence. The legitimacy of UN, European Union (EU) and Bretton Woods institutions can provide particular advantages, for example in terms of policy dialogue at the country level, compared with bilateral programmes where ultimate accountability is to domestic political structures. Global funds provide close association with particular MDGs, bringing political appeal and visibility, and can introduce greater voice by incorporating civil society in their governance structures.

Although multilateral assessments throw some light on the issue, they rarely address the way in which different mandates contribute to achievement of the MDGs, leaving considerable room for judgement. Nor do multilateral assessments provide a sufficient basis for judging the comparative advantage of multilaterals with overlapping mandates.

In summary, although information on effectiveness is useful for discussion of financial allocations, it is not likely to be determinant in any systematic way. Thus it is highly unlikely that bilaterals could develop a meaningful performance-based allocation model for funding the multilaterals. Bilaterals should acknowledge that effectiveness assessments only play a minor part in informing decisions about particular levels of funding and pay more attention to using them for other objectives.

5.2.3 Bilateral advocacy and influencing objectives

It can be argued that influencing the multilaterals towards bilateral objectives can produce far greater effects than by financing through bilateral channels alone,

because of the larger size, budget and geographical scope of most multilateral agencies.⁴⁷ Decisions to route finance through multilaterals can be, and is in practice, closely allied to influence. Bilaterals may seek to wield influence by rewarding reforms (ex post) or leveraging reforms (ex ante) through more generous commitments, allied with the possibility of threats of withdrawal if promised actions fail to materialise. Political pressure on European governments to lever changes in WB lending policy during the IDA 15 negotiations, and withdrawal of DFID funding to the WB linked to progress against conditionality policy⁴⁸ provide recent examples.

Bilaterals have sought to influence the multilaterals through partnership strategies which include joint operations, policy work, and extra budgetary financing. The content of influencing strategies remains, in practice, an internal matter for bilateral agencies to pursue in the light of their aid policies and domestic constituency needs. However, information on effectiveness can point to particular areas where efforts should be concentrated. The question is whether separate assessments are still needed as an input to influencing strategies, in addition to any information the multilaterals produce. In the short run this is still likely to be the case given the issues around multilateral reporting mentioned above. However, if the uses of this information was more clearly specified, the assessments could be better focused and less onerous, reducing transactions costs and duplication. Separate bilateral assessments are valuable in providing a common basis for joint bilateral influencing strategies, and broadening the scope for consensus. The joint discussions of the MOPAN and Utstein groups with the multilaterals provide examples of more joined up advocacy and influencing around a focus on effectiveness. Bilateral donors wishing to influence multilateral donors more effectively may want to consider strengthening the consensus within these types of networks.

5.3 Improving the Bilateral Governance Role

One criticism that has been made of individual bilateral influencing strategies is that they undermine the collective governance process. In acknowledging this criticism, and taking into account the problems arising from the multiplicity of approaches to assessing multilateral effectiveness, a strong case can be made that bilaterals should fundamentally change the way they advocate for multilateral reform. Instead of using non-transparent approaches for assessing multilateral effectiveness from the outside, they could more usefully use their positions within the governance mechanisms of the multilaterals to advocate for change from within the multilaterals.

The ultimate objective should be to move beyond reliance on individual bilateral assessments by giving more attention to multilaterals developing their own approaches in ways that meet the needs of shareholders/stakeholders. Once again,

⁴⁷ For a more detailed discussion of the potential synergies between the activities of bilateral and multilateral agencies see the companion paper to this one (Graves et al, forthcoming)

⁴⁸ See <http://www.dfid.gov.uk/News/files/world-bank50.asp>

however, we come back to the need for bilaterals need to be clearer and more consistent about the standards they expect. Collective groupings such as MOPAN could play an important role in getting consistency and clarity amongst members, *inter alia* by leading a debate on principles and standards for assessing and reporting on effectiveness. Such a debate would be best founded on collective assessment experiences, such as the MOPAN Survey and the proposed new approach.

In summary, bilateral needs for separate information on multilateral effectiveness are likely to continue in the short run, but more to support their influencing and governance roles than as an input to financial decision-making. However, in order to reduce the risk of duplication and escalation of transactions costs, it is desirable that:

- Bilateral assessments are conducted on a collective basis rather than separately; and
- A consensus is forged on quality standards for the assessment of effectiveness, to guide advocacy for improvements in multilateral reporting on effectiveness.

5.4 Meeting the Needs of Partner Countries

It is clear from the analysis above the majority of the approaches to assessing multilateral effectiveness are carried out for and by donors, primarily the bilaterals. Even the Paris and the SPA Surveys, which include a role for partner governments, are driven by the donors. The information that is generated therefore primarily meets their needs, rather than the very different information needs of partner countries. For example, partner countries are interested in bilateral as well as multilateral assessments; the comparative efficiency of different agencies in terms of transactions costs; the local relevance of their lending or grant policies, different levels and types of conditionality etc. As part of the Paris mutual accountability agenda, any joint assessments conducted through the OECD/DAC, should meet the needs of *all* stakeholders in the development relationship, rather than those of development agencies alone.

Only one of the methodologies described above, that of DRI, is designed primarily to meet the needs of partner governments. Its aim is to provide them with the information they need about donor policies and procedures in order to develop clearer policies to help to guide their engagement with donor agencies. The methodology developed by DRI, and some of the Overseas Development Institute's (ODI) own work, indicates that partner governments and wider recipient stakeholder groups (such as Members of Parliament, Business Leaders and Civil Society Leaders) are interested in a broader range of criteria as they make decisions (Johnson et al, 2004; Burall and Maxwell, 2006; Burall et al, 2007) (see Annex 4).

In the context of Paris and the moves towards promoting greater country ownership, we suggest that those developing and revising multilateral assessments should firstly consider ways to include more input from partner governments, and

secondly how to meet their specific information needs. Particular attention should be paid to synthesising information in a form that is useful for partner governments. The DRI methodology offers one possible template that could be used and adapted. However, it has only been applied to HIPC countries and it could be rolled out more widely. Any assessment developed should include not just assessments of multilateral agencies, but also of bilaterals, as they can provide a significant proportion of overall ODA levels in some countries.

5.5 Implication of Effectiveness Assessments for the Global Aid Architecture

Important commitments have been made, for example in the G8 and the EU, both to increase the quantity of global aid and, through the Paris Declaration and the EC code on division of labour, to improve its quality. The effect of these commitments remains an open question though there are clear processes in place for taking stock of progress and holding donors to account, for example the Third High-Level Forum on the Paris Declaration in Accra in September 2008.

By contrast, the international context is changing with the growth of China and India, trade and security challenges, and concerns over global/regional public goods which could knock progress against MDGs off-track and raise huge new demands for additional financing. In addition, questions about where, and for what purposes, global aid needs to be allocated remain unanswered. This is preventing effective co-ordination at the international level. Critical weaknesses in the aid delivery system, such as proliferation and verticalisation, have been identified as reducing the effectiveness of the system taken together and there is little or no consensus on how the system should evolve to address these. With their overwhelming focus on *multilateral* effectiveness, none of the approaches considered in this paper contribute to a deeper understanding about the effectiveness of the aid system as a whole. Bilateral agencies should therefore consider focusing more energy looking at effectiveness across a broader range of actors (multilaterals, old and new bilaterals, private foundations, etc.) Again, this would imply an existing consensus around common standards. Groups like MOPAN may offer one venue for these discussions to be held and consensus to be developed.

6 Conclusion

This review of different approaches to assessing multilateral effectiveness reveals a picture of complexity, fragmentation and inefficiency. The approaches differ widely in the aspect of effectiveness studied, the methodology employed and the scope of the findings. All approaches have strengths and limitations. No single approach is able to capture all the relevant aspects of effectiveness and all approaches cast useful light on one aspect of effectiveness or another, although there are also gaps that need to be addressed. This is a reflection of the complexity of effectiveness itself.

In the face of this complexity, there is a need for synthesis. Bilateral decision-makers find it difficult to reconcile the fragmented and often conflicting information generated by bilateral and multilateral approaches for assessing multilateral effectiveness. Such a synthesis would draw on, rather than replace, existing assessment approaches as the transactions costs of a single comprehensive approach would be too great for any single organisation to bear. There is a trade off between comprehensiveness and sustainability in this respect. The question is, who should conduct such a synthesis and how would the criteria be agreed? The current MOPAN initiative to provide a synthesis through a Balanced Scorecard approach is to be welcomed, although ideally each multilateral should produce this itself.

Amongst the variety of assessment approaches described here there is considerable duplication of effort. Separate bilateral assessments often require similar information from the multilaterals, and partnership behaviour surveys often ask similar questions of development stakeholders at country level. Such a duplication of effort is inefficient and time-wasting for key decision-makers; time that could be better spent getting on with the business of improving effectiveness rather than answering questions about it.

There are many reasons for the proliferation of assessments of effectiveness by both bilaterals and multilaterals. For the most part this is based on accountability pressures, either by members of governing bodies in the case of the multilaterals or domestic constituencies in the case of bilaterals. Yet there is much uncertainty about what is expected in terms of effectiveness reporting to meet these accountability requirements. This is because of a lack of international consensus about minimum criteria for the assessment of effectiveness and good practice standards for assessment methodologies. As a result, bilaterals continue to carry-out their own assessments of multilateral effectiveness, in part because they themselves are unsure about what they require in this respect. They are therefore ineffective in pressing for improvements through their governance role and

instead conduct separate bilateral assessments, thereby incurring substantial transactions costs.

Bilaterals justify their separate assessments in terms of their use for internal decision-making. However, they need to be more self-critical of these internal purposes and specify the role of effectiveness information more precisely. Our analysis shows that although bilateral accountability is primarily rooted in justifying the decisions on financing the multilaterals, in practice the scope for using information on relative effectiveness for these decisions is limited. On the other hand, there is a case for using such information for influencing and governance objectives, although this use is weakly specified at present.

Ultimately, the ideal scenario would be that multilaterals' reporting on their own effectiveness were sufficiently comprehensive and yet synthetic to satisfy bilateral information requirements, making separate bilateral assessments unnecessary. However, in the short run or until such a scenario were reached, these separate assessments will continue to be needed. Nevertheless, there is a strong case for;

- Conducting them collectively in order to reduce duplication and transactions costs: and
- Developing a consensus on minimum requirements and standards to guide the design of such assessments and common advocacy for improving multilateral reporting itself.

Such fora as the MOPAN, the DAC Joint Venture on MfDR and the SPA provide important mechanisms for these kinds of joint ventures. In particular, the current initiative within MOPAN for a Common Balanced Scorecard Approach offers potential to meet both these requirements.

This report is one of a number of recent reports evaluating the different approaches to assessing effectiveness. In all of these reports, the bilateral approaches that are covered are those from 'like-minded' donors. If standards are to be developed for assessing donor effectiveness, there is a critical need to understand the purposes for which, and how, a broader range of actors currently make their assessments. This understanding will be needed if a broad consensus is to be built around new standards for assessment.

To date, the debate about multilateral effectiveness has largely been conducted amongst development assistance agencies, whether bilateral or multilateral. And the concerns with accountability are largely those of the Western governments which provide the finance for both types of agencies. It is surprising however, that in the context of international exhortations about the importance of mutual accountability between development agencies and partner governments, that there is little voice for the latter in the debate about effectiveness. It is possible and desirable that future assessments will be led by partner governments in order to address their rather different information needs and restore a more balanced approach to mutual accountability.

Finally, important as the debate about multilateral effectiveness is, it is only a small part of the international aid system, which is seeing growing importance of Global Funds, non-Western bilaterals, and private Foundations in the overall aid architecture. Bilateral and multilateral agencies need to consider how to extend the debate about effectiveness in one small part of the system to a wider range of actors and broader systemic issues. A shift in discussion is desirable to explore whether, for example, proliferation is best addressed through planned rationalisation of aid delivery channels, disciplining mechanisms at country level; or more radical ‘marketisation’ of the aid system.

This report concludes that bilaterals should pursue the following strategy for future assessments of multilateral effectiveness:

- Use collective fora such as MOPAN and the DAC Joint Venture on MfDR to develop a consensus on criteria for the assessment of multilateral effectiveness and common good practice standards for assessment methodologies
- Use these standards as a basis for clarifying their information requirements from the multilaterals and advocating improvements in multilateral reporting on effectiveness, such that separate bilateral assessments eventually become unnecessary. In particular, they should advocate more synthetic reporting from the multilaterals.
- Clarify the rationale for conducting separate bilateral assessments in order to inform financing decisions, influencing strategies and an enhanced role on the governing boards of the multilaterals in such a way that the use of this information is specified more precisely
- Conduct separate bilateral assessments only on a collaborative basis with other bilateral donors through mechanisms such as MOPAN
- Promote mutual accountability with partner governments by ensuring that their particular information needs are adequately catered for and their participation in survey instruments is ensured.
- Jointly with other bilaterals and multilaterals, seek to extend the debate on multilateral effectiveness to wider systemic aid architecture issues.

Annex 1 - References

- AfDB (2006) *Synthesis Report on the Review of 2003-2005 Project Completion Reports*, Tunisia, Operations Evaluation Department, AfDB.
- Burall, S. and Maxwell, S. (2006) *Reforming the international aid architecture: Options and ways forwards*, London, Overseas Development Institute.
- Burall, S., Mease, K., Mall, P., and Datta, A. (2007) *Assessing Key Stakeholder Perceptions of the Effectiveness of Multilateral Organisations*, London, Overseas Development Institute.
- Collier, P. and Dollar, D. (2001) 'Can the World Cut Poverty in Half? How Policy Reform and Effective Aid Can Meet International Development Goals,' *World Development*, 29(11), pp 1787-1802.
- DANIDA (2005) *Performance Management Framework for Danish Development Cooperation 2006-2007*. Quality Assurance Department, KVA, DANIDA.
- DFID (2007) *Multilateral Development Effectiveness Summaries – Frequently Asked Questions*, [online] [Accessed on 9 January 2008] Available from World Wide Web: <<http://www.dfid.gov.uk/aboutdfid/dfidwork/multilateral-eff-faq.asp>>
- DFID (2006) 'Multilateral Development Effectiveness Summaries,' *DFID Methodology Note*, October 2006.
- Duska-Anema Development Associates (DADA) International (2006) *Bilateral Methodologies for Assessing Multilateral Performance: Survey Findings*, Vancouver, CIDA.
- FAO (2007) *FAO: The Challenge of Renewal, Report of the Independent External Evaluation of the Food and Agriculture Organisation of the United Nations*, Rome, Council Committee for the Independent External Evaluation, of FAO.
- Foresti, M., Archer, C., O'Neil, T. and Longhurst, R. (2007) *A Comparative Study of Evaluation Policies and Practices in Development Agencies*, Paris, Agence Française de Développement.
- Graves, S., Wheeler, V., Foresti, M., Burall, S., Highton, N. (forthcoming) *Synergies between bilateral and multilateral activities*, London, Overseas Development Institute.
- IDA (2007) *Aid Architecture: an overview of the main trends in official development assistance flows*, Washington, DC, International Development Association.

- IFAD (2007) *Report on IFAD's Development Effectiveness*, Rome, Strategic Planning and Budget Division, IFAD.
- Johnson, A. and Martin, M. (2005) *Empowering Developing Countries to Lead the Aid Partnership: A Background Paper for the UNDP Human Development Report*, London, Development Finance International.
- Johnson, A., Martin, M. and Bargawi, H. (2004) *The Effectiveness of Aid to Africa Since the HIPC Initiative: Issues, Evidence and Possible Areas for Action*, London, Development Finance International.
- Joint Inspection Unit (2004) *Overview of the Series of Reports on Managing for Results in the United Nations System*, Geneva, JIU.
- Kabell, D. and Balogun, P. (2004) *Selected Bilateral Donors' Current and Future Performance Information Needs from Multilateral Organisations: An Assessment*, Denmark, DANIDA
- Lloyd, R., Oatham, J. and Hammer, M. (2007) *Global Accountability Report*, London, One World Trust
- Martens, B. (2005) 'Why Do Aid Agencies Exist?', *Development Policy Review*, 23 (06), London, Overseas Development Institute
- Meier, W. (2007a) *Assessing Multilateral Organisation Effectiveness: A Comparative Analysis of Assessment Tools and Development of a Common Approach*, Stockholm, RBMG.
- Meier, W. (2007b) *Assessing Multilateral Organisation Effectiveness: Implementing the MOPAN Common Approach*, London, RBMG.
- MfDR. (2006) *Multilateral Development Banks Common Performance Assessment System 2005 Report*, Thailand, Asian Development Bank.
- MfDR. (2007) *The 2006 Multilateral Development Banks Common Performance Assessment System*, Washington DC, Inter-American Development Bank.
- Multilateral Organisations Performance Assessment Network (2006) *The MOPAN Survey 2006: Donor Perceptions of Multilateral Partnerships at Country Level – Synthesis Report*, Zurich, MOPAN.
- Obser, A. (2007) *Multilateral Organisations Performance Assessment: Opportunities and Limitations for Harmonisation among Development Agencies*, Bonn, German Development Institute.
- OECD/DAC (2007a) *Managing for Development Results Principles in Action: Sourcebook on Emerging Good Practices (Second Edition)*, Paris, OECD/DAC.
- OECD/DAC (2007b) 'Survey on Monitoring the Paris Declaration: Overview of the Results', *OECD Journal on Development*, 8(2), pp i-133.

- OECD/DAC (2005) *Paris declaration on aid effectiveness, Ownership, harmonisation, alignment, results and mutual accountability*, Paris, OECD/DAC.
- OECD/DAC (2002) *Glossary of Key Terms in Evaluation and Results Based Management Terms*, Paris, OECD/DAC Working Party on Aid Evaluation.
- OECD/DAC (2000) *Results Based Management in the Development Co-Operation Agencies: A Review of Experience*, Paris, OECD/DAC Working Party on Aid Evaluation.
- Scott, A. (2005a) *DFID's Assessment of Multilateral Organisational Effectiveness: An Overview of Results*, London, IDAD, DFID.
- Scott, A. (2005b) *The 'MEFF' Methodology: A Review of DFID's Multilateral Effectiveness Framework*, London, IDAD, DFID.
- SPA (2007) *Survey of Budget Support, 2006, Volumes 1 & 2*, Washington DC, SPA.
- UNEG (2007) *Oversight and Evaluation in the UN*, New York, United Nations.

Annex 2 – Terms of Reference

Danida file: 104.A.1.e.81.

17-10-07

Terms of Reference

Approaches to assessing multilateral performance

Background

The performance of multilateral organisations has for many years been debated and analysed, but in particular the last five years have seen the development of several different approaches to performance assessment of multilateral organisations. These approaches usually focus on both relevance and effectiveness of the activities of the individual multilateral organisations, but no generally accepted approach has been established so far.

The appearance of these different approaches is a consequence of donors' stronger focus on management for results as well as on division of labour and comparative advantages.

In addition to the existence of these different approaches to performance assessment several evaluations of multilateral organisations have been carried out. Most of these evaluations have been done by the evaluation units of the multilateral organisations. The organisations within the UN-system have recently developed a standard approach to be used in evaluations – an approach which is based on DAC evaluations principles - but at present the bilateral donors do not have a common approach to evaluations of multilateral organisations.

When bilateral donors assess multilateral organisations they usually base their assessments on both one or more of the above mentioned performance assessment methods as well as on available evaluations.

The number of existing approaches calls for an analysis of how the various approaches are related and of their relative strengths and weaknesses.

Objective

The objective is

- a)
 - to identify and compare available key approaches to performance assessment of multilateral organisations
- b)

- based on (a) above to identify gaps for future analyses

Output

The output will be a report which will not exceed 30 pages, annexes excluded. The report will be published by the Evaluation Department as a Danida Evaluation Study. The Evaluation Study will also be made available on the Internet (www.evaluation.dk).

Scope of work

It is envisaged that the analyses will have to be based on a distinction between different kinds of multilateral organisations. The scope of work will at least cover the following issues:

- Conceptual and methodological issues. This part of scope of work will identify the main conceptual and methodological challenges related to multilateral performance assessment. Most performance assessments seem to be based on criteria like relevance and effectiveness. To which extent are these concepts applicable to all multilateral organisations? The relevance of quantitative and qualitative methods? Perceptions studies are popular, but how valid are they? Can comparative assessments of multilateral organisations be carried out?
- A mapping of existing key approaches to assessment of multilateral performance. The mapping, which will include a short description, will cover a selected number of approaches. The sample will be agreed between the Evaluation Department of Danida and ODI.
- Overall assessment of the selected approaches. Based on the conceptual and methodological analysis and the mapping above an overall assessment of the approaches will be made.
- Ways forward for approaches to assessing multilateral performance. Gaps, which can be covered by future analyses, will be identified. This part will also include a discussion of the relationship between multilateral performance management and evaluations of multilateral organisations. Finally, it will consider the implications of assessments of individual multilateral organisations for the overall aid architecture.

[... practical details follow ...]

Ole Winckler Andersen/EVAL

Annex 3 – Comparing the Different Approaches to Assessing Multilateral Effectiveness

Approach	Description	Purpose	Criteria Measured	Data Sources	Strengths	Limitations	New information generated	multilateral assessed
<i>RBM Assessments</i>								
COMPAS	Common framework for MDBs to jointly report on their own performance	To provide managers and shareholders of the MDBs with baseline information on how the MDBs are contributing to development results and how they may improve their contributions over time	30 performance indicators grouped into 7 categories that relate to results orientation at the partner country level, institutional level and global partnership level	Self-assessment	Generates data that could serve to populate or complement other methodologies	Only 5 MDBs involved	Common definitions and results-based methodologies between MDBs	5: AfDB, AsDB, EBRD, IABD and WB.
DFID MEFF	Based on RBM approach, a framework that distinguishes between multilateral categories to assess organisational effectiveness	To provide information for public accountability and input for policy and financing decisions related to multilateral	Internal operational effectiveness from 3 perspectives: internal performance, focus on country-level results, and partnership that evaluates 8 corporate management systems	Checklist with 72 questions; a scorecard that uses a simple traffic lights system to score an agency's effectiveness in each area; and a summary report carried out by DFID staff	Comes close to a comprehensive measurement approach Considered by other donors as a good starting point for progress towards harmonisation	Does not assess how well multilateral operational systems are implemented, or what results are actually achieved		23: AfDB, AsDB, EBRD, EC, FAO, HABITAT, IADB, ICRC, IFAD, IFRC, OCHA, OHCHR, UNAIDS, UNDP, UNESCO, UNFPA, UNHCR, UNICEF, UNIDO, UNIFEM, WFP, WHO, and WB

Approach	Description	Purpose	Criteria Measured	Data Sources	Strengths	Limitations	New information generated	multilateral assessed
DANIDA RBM Assessment	Methodology to review the extent to which RBM principles form the basis of the work of monitoring and evaluation units within the multilateral.	To identify both what RBM tools are in place as well as how effectively they are used in pursuit of improved organisational effectiveness and efficiency at delivering development results	Independence, credibility, and usability of evaluations through assessment of the internal systems and processes of a multilateral's own central evaluation office	multilateral documentation, staff interviews, on-site observations and discussion workshops carried out by Peer Panel	Good potential for evolving into a common approach for assessing multilateral effectiveness credible Focuses attention of multilateral management on areas of importance as highlighted by bilateral donors	Need for more rigorous attention to the definition of results and an international institutional arrangement for objective verification purposes	Understanding of how RBM principles are incorporated into multilateral evaluation units	5: UNDP, UNFPA, IFAD, OHCHR and UNICEF

Approach	Description	Purpose	Criteria Measured	Data Sources	Strengths	Limitations	New information generated	multilateral assessed
<i>Partnership Behaviour Assessments</i>								
Partner Stakeholder Perception Survey	Pilot project to identify the views of stakeholders in recipient countries about the performance of key multilateral organisations, and their preferences for which organisations should disburse additional aid	To seek the perceptions of key stakeholders in aid recipient countries about the effectiveness of multilateral.	multilateral performance according to 3 measures from the Paris Declaration: overall development effectiveness, harmonisation and alignment; ranking against 15 performance criteria related to funding and policies and procedures	Questionnaire seeking views from five stakeholder groups that includes ranking organisation in the order of preferred disbursement channel.	Considers views of stakeholders in aid partnership outside of government and donors	High transaction costs for stakeholders Challenging to identify well informed respondents to reduce subjectivity	Views of a range of stakeholders who interact with donors in-country and are responsible for implementing donor programmes.	7: AfDB, AsDB, EC, GFATM, UNICEF, UNDP and WB
MOPAN Survey	Annual Survey on 3-4 selected multilateral in 8-10 countries where MOPAN members have bilateral programmes	For accountability on multilateral financing, for input into their policy towards multilateral, to strengthen their relationship with multilateral in governance and joint activities and to contribute to the wider debate of aid effectiveness	Quality of multilateral partnership behaviour towards national stakeholders and other international development agencies at the country level	Perception surveys of MOPAN member embassies and country offices regarding the performance of different multilateral	Positive step towards uniting bilateral donor efforts in assessing multilateral Light and rapid exercise with minimal transaction costs	Highly subjective Does not measure development impact or multilateral effectiveness achieved on the ground	Fills an important gap in understanding how multilateral operate at the country level and sharing experiences in working with multilateral	11: AfDB, ADB, IDB, ILO, FAO, UNAIDS, UNDP, UNPF, UNICEF, WB, and WHO

Approach	Description	Purpose	Criteria Measured	Data Sources	Strengths	Limitations	New information generated	multilateral assessed
Paris Declaration Baseline Survey	Survey conducted by OECD to monitor progress towards goals of Paris Declaration	To generate an accurate picture of existing aid practices as a baseline for assessing progress towards the 12 indicators as well as stimulate a broad-based dialogue on how to make aid more effective and promote agreement on specific actions that contribute to successful implementation of the Paris agenda at country level	Indicators related to the following principles: ownership, harmonisation, alignment, results' management and mutual accountability	2 questionnaires sent separately to in-country donors and government which asks for both quantitative information and qualitative commentary about donors relating to the relevant Paris Declaration indicators.	Contributed to generating dialogue on aid effectiveness and progressing national monitoring efforts	Disconnected analysis of aid delivery and country systems Definitions for indicators too flexible Onerous and time-consuming	More is now known about aid management systems and practices at the country level	8: AfDB, AsDB, EC, GAVI Alliance, GFATM, IADB, UN and WB
SPA Budget Support Survey	Annual monitoring exercise conducted by African governments to assess aid relationship with donors	To evaluate the progress made by bilateral and multilateral donors in aligning and harmonising their budget support activities in several PRS countries in Africa	Predictability of budget support, the links between budget support and PRS processes, and government views on donor behaviour in budget support relationships	Questionnaires sent to donors and government representatives to obtain perspectives on general budget support practices with regard to alignment	Provides some important opportunities for dialogue amongst donors and development partners Considerable potential for trend analysis	Considerable overlap with the Paris Declaration Baseline Survey which increased transaction costs for participating countries Limited to aspects of budget support, and does not provide HIPC with the tools to assess the performance of donors and their own priorities for mobilising funding	Most comprehensive data currently available on the progress of budget support alignment with the PRS	4: AfDB, EC, UNFP, WB

Approach	Description	Purpose	Criteria Measured	Data Sources	Strengths	Limitations	New information generated	multilateral assessed
DRI	Methodology developed for HIPC governments to assess the quality of aid they receive as part of the process of designing overall national debt and new financing strategies	To help HIPC governments identify which are the 'best' external financing sources to fund their national development programmes, and how they can improve the quality of all their external finance.	Quantity and quality of external assistance flows to HIPC governments, through analysis of the policies and procedures of donors as well as HIPC governments.	Scores given by HIPC governments on 23 evaluation criteria for donors; self-assessment by HIPC governments	Led by HIPC governments themselves Systematically monitors whether donors are living up to their pledges on aid quality and support for the PRSP framework or behaving consistently across a range of African countries	Time and capacity intensive	Identifies which qualitative aspects are the worst problems, and which donors have the best/worst quality aid.	10: AfDB, Arab Bank for Economic Development in Africa, EU, European Investment Bank, IDA, IFAD, IMF, Islamic Development Bank, OPEC, UN Agencies
<i>Synthetic Approaches</i>								
DANIDA PMF	Management tool that provides timely and reliable information on progress of activities, and it ensures accountability and reporting to external stakeholders.	To enhance the quality of Danish development cooperation; to improve management and continuous learning; and to strengthen accountability	Organisational practices and field level results based on objectives, targets and indicators established by the relevant multilateral themselves	multilateral strategies, perception analyses, and assessments of multilateral' evaluation & reporting systems gathered at three levels (DANIDA corporate, multilateral headquarter and multilateral country level)	Comprehensive Dual purpose of measurement and management	Complexity adds to the time and resources needed to conduct the evaluation for each multilateral		3 most important multilateral per country each year

Approach	Description	Purpose	Criteria Measured	Data Sources	Strengths	Limitations	New information generated	multilateral assessed
DFID MDES	Balanced scorecard that collates the variety of existing information on multilateral to provide overview of multilateral performance	To inform DFID's management decision making process and to ensure senior managers have a consistent picture of each multilateral	Four key dimensions of effectiveness: managing resources, managing relationships, country/global results and building for the future	Published secondary data sources including agency reports, donor evaluations, perception surveys, NGO reports, and the Paris Declaration monitoring exercise Compiled by DFID staff	Issues of comparability both across organisations and also across time Does not produce a ranking of multilateral by their effectiveness nor intended to measure actual development impact Complex and more time-consuming to complete	Strong potential for evolving into a common approach for assessing multilateral effectiveness which focuses explicitly on accountability	First attempt to compile data from multiple sources to assess multilateral internal performance and country level relationships and results	14: AfDB, AsDB, EBRD, FAO, GFATM, ICRC, UNAIDS, UNDP, UNESCO, UNFPA, UNHCR, UNICEF, WHO, WB

Table 3: Comparison of Multilateral Assessment Methodologies

Annex 4 – The Importance of Additional Partnership Behaviour Criteria for Partners

The DFID commissioned Stakeholder Perceptions Survey (see section 3.3.2) included a question asking respondents to rank the importance of a range of fifteen effectiveness criteria drawn from research carried out by both ODI and by DFI. The questionnaire also offered respondents an opportunity to add additional criteria. The fifteen effectiveness criteria in the questionnaire were divided into two categories: *Funding*, and *Policies and Procedures*. These two categories have seven and eight criteria respectively.

Respondents were asked to rank the importance of the criteria using a scale of high, medium and low. The total score was calculated for each criterion. The scores were then used to rank the criteria in order, from the highest to the lowest. The two figures below show the rankings for the Funding and for the Policies and Procedures criteria (figures and discussion drawn from Burall et al, 2007).

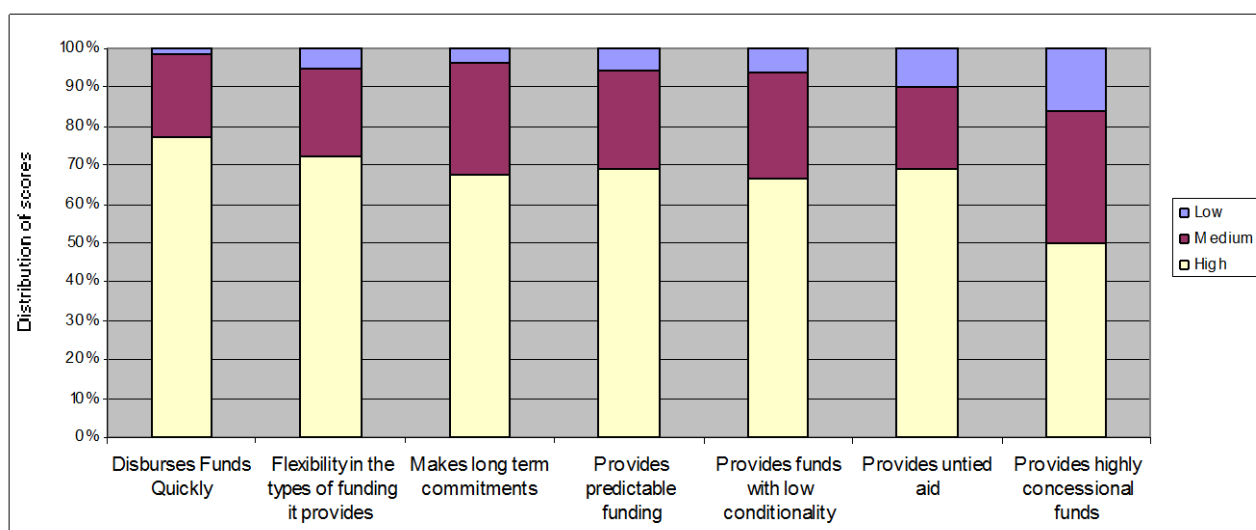


Figure 2: Ranking the importance of the seven funding criteria

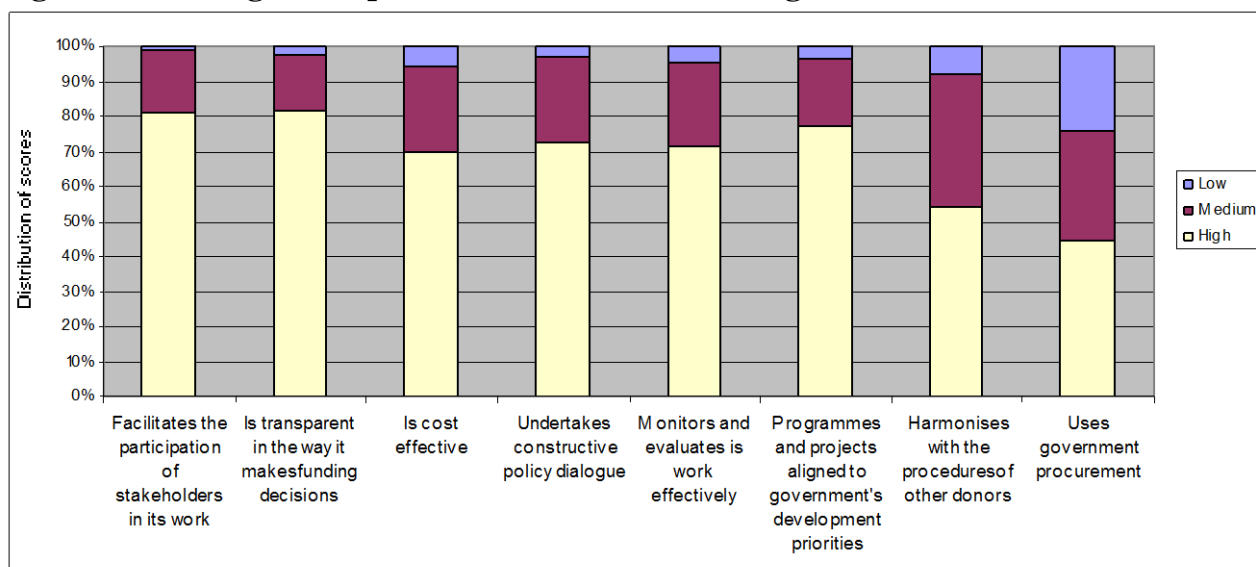


Figure 3: Ranking the importance of the eight policy and procedure criteria

The figures show that most of the criteria are considered to be either high or medium importance, with only one, uses government procurement procedures, receiving less than 50% of the respondents' high-importance rankings. In addition, very few respondents added new criteria to the list and no consensus about new criteria emerged.

Published by:
Ministry of Foreign Affairs of Denmark
Evaluation Department
Asiatisk plads 2
1448 Copenhagen K
Denmark
E-mail: eval@um.dk

The publication can be downloaded from:
www.evaluation.dk

ISBN: 978-87-7667-896-8
ISBN: 978-87-7667-897-5 (internet version)

Other Evaluation Studies:
2007/1: Assistance to Development Research
2007/2: Synthesis of Evaluations on Technical Assistance
2007/3: Evaluating the Impact of Rural Roads in Nicaragua
2008/1: Synthesis of Evaluations of HIV/AIDS Assistance
2008/2: Synergies between Bilateral and Multilateral Activities

Responsibility for the content and presentations of findings and recommendations rests with the author.
The views and opinions expressed in the report do not necessarily correspond to the views of the Ministry of Foreign Affairs of Denmark.