

DISCUSSION PAPER / 2010.03



# Managing Multilaterals' Effectiveness: Which way forward?

Koen **De Koster** and Nathalie **Holvoet**



University  
of Antwerp



INSTITUTE OF DEVELOPMENT  
POLICY AND MANAGEMENT

Comments on this Discussion Paper are invited.  
Please contact the authors at: [koen.dekoster@ua.ac.be](mailto:koen.dekoster@ua.ac.be),  
[nathalie.holvoet@ua.ac.be](mailto:nathalie.holvoet@ua.ac.be)

Instituut voor Ontwikkelingsbeleid en -Beheer  
Institute of Development Policy and Management  
Institut de Politique et de Gestion du Développement  
Instituto de Política y Gestión del Desarrollo

Postal address:  
Prinsstraat 13  
B-2000 Antwerpen  
Belgium

Visiting address:  
Lange Sint-Annastraat 7  
B-2000 Antwerpen  
Belgium

Tel: +32 (0)3 275 57 70  
Fax: +32 (0)3 275 57 71  
e-mail: [dev@ua.ac.be](mailto:dev@ua.ac.be)  
<http://www.ua.ac.be/iob>

**DISCUSSION PAPER / 2010.03**

# **Managing Multilaterals' Effectiveness: Which way forward?**

**Koen De Koster\*** and **Nathalie Holvoet\*\***

November 2010

\*Koen De Koster is currently a staff member of the NGO Broederlijk Delen and a former student of the Institute of Development Policy and Management (IOB), University Antwerp (academic year 2008-2009). This paper was elaborated on the basis of his dissertation which he submitted for the degree of Master in Development Evaluation and Management. The author would like to thank the IOB for the research grant he obtained to write his paper. Furthermore the author wishes to express his gratitude to staff from the DGDC (the Special Evaluation Office in particular), DFID and DGIS for their time; the two referees, Prof. Dr. Robrecht Renard and Johan Debar, for their useful comments; and finally Annelies Vekeman, Robin De Boeck, Kaat Houtman and Em. Prof. Dr. Joris Duytschaever.

\*\* Nathalie Holvoet is lecturer at the IOB.



## TABLE OF CONTENTS

	<b>ABSTRACT</b>	6
<b>1.</b>	<b>INTRODUCTION</b>	9
<b>2.</b>	<b>ASSESSING MULTILATERALS' PERFORMANCE</b>	11
2.1.	The unbearable lightness of bilateral donorship	11
2.2	MOPAN - The Common Approach	13
2.3	Professional peer reviews of evaluation functions in MOs	15
2.4	CIDA's cost-effective joint evaluation of MOs	16
2.5	Towards a two-tier approach?	17
<b>3.</b>	<b>MANAGING PERFORMANCE OF MOs</b>	19
3.1	Managing performance: a framework	19
3.2	Performance administration: Belgium	24
3.2.1	Context	25
3.2.2	Measuring performance	25
3.2.3	Integrating and using performance information	26
3.3	Managements of performances: the Netherlands	27
3.3.1	Context	27
3.3.2	Measuring performance	28
3.3.3	Integrating and using performance information	29
3.4	Performance management: United Kingdom (UK)	30
3.4.1	Context	30
3.4.2	Measuring performance	31
3.4.3	Integrating and using performance information	32
<b>4.</b>	<b>CONCLUSION AND RECOMMENDATIONS</b>	34
	<b>REFERENCES</b>	37

## ABSTRACT

There has been an increase in the attention paid to the measurement of the performance of development aid and of aid agencies. While the monitoring and evaluation of aid delivered by bilateral agencies has been a well-established practice for quite some time, the measurement of the performance of Multilateral Organizations (MOs) by bilateral back-donors, and the use of this information in improved bilateral policies towards these same MOs is more recent. But of late there has been a mushrooming of bilateral initiatives that try to assess MO performance. Currently bilateral donors seem to be moving toward a more harmonized approach through the introduction of assessments that are organized by networks of donors. How the information from these assessments is being used in policy-making, however, is far from clear.

This paper analyzes the current state of affairs in the field of assessing multilaterals. It feeds into the debate on how bilateral donors should effectively and efficiently manage the performance of the MOs they fund. The analysis of three country cases (Belgium, the Netherlands and the United Kingdom) reveals significant differences in the performance management of multilaterals by bilaterals. The paper examines these differences in the light of the literature on performance measurement. It argues that while donors are increasing their efforts in measuring the performance of multilaterals, the use of this information often remains insufficient. The good example set by the UK is closely linked to administrative and political features of this country and difficult to just duplicate by other countries. The field of managing the performance of MOs remains scattered, unsystematic, and above all, challenging.

## ABSTRAIT

L'évaluation de l'efficacité de l'aide au développement et des organismes d'aide suscite un intérêt croissant. Alors que le contrôle et l'évaluation de l'aide fournie par les organismes d'aide bilatérale sont pratiqués depuis un certain temps, l'appréciation de l'efficacité des organismes d'aide multilatérale par des donateurs de soutien bilatéraux, et l'intégration des informations ainsi obtenues dans des politiques bilatérales à l'égard de ces mêmes organismes multilatéraux, s'avèrent plus récentes. Toutefois, les initiatives bilatérales visant à évaluer l'efficacité des organismes multilatéraux se multiplient ces derniers temps. Les donateurs bilatéraux semblent actuellement se diriger vers une approche plus harmonieuse, à travers l'introduction d'évaluations organisées par des réseaux de donateurs. Un manque de clarté évident apparaît toutefois quant à la façon dont les informations issues de ces évaluations sont employées dans l'élaboration des politiques.

Ce document analyse l'état actuel de la situation en matière d'évaluation des organismes d'aide multilatérale. Il alimente le débat sur les méthodes concrètes et efficaces que les donateurs bilatéraux devraient appliquer en vue de gérer l'efficacité des organismes d'aide multilatérale qu'ils financent. Les analyses de cas effectuées dans trois pays (en Belgique, aux Pays-Bas et au Royaume-Uni) révèlent des différences significatives en matière de gestion de l'efficacité des organisations multilatérales par des organismes bilatéraux. Ce document entreprend d'analyser ces différences à la lumière d'écrits sur l'évaluation de la performance. Il soutient l'idée selon laquelle les donateurs redoublent d'efforts pour mesurer l'efficacité des organisations multilatérales, tout en avançant l'emploi souvent insuffisant des informations obtenues. Le bon exemple donné par le Royaume-Uni est étroitement lié à ses spécificités administratives et politiques nationales, et peut ainsi difficilement être adopté d'emblée par d'autres pays. La gestion de l'efficacité des organisations multilatérales demeure un terrain diffus, sans systématique, qui surtout pose un véritable défi.





## 1. INTRODUCTION

In the 2005 Paris Declaration on Aid Effectiveness (OECD/DAC 2005) donors and recipients committed themselves to five distinct but interrelated principles to increase the impact of aid: ownership, alignment, harmonization, managing for results and mutual accountability. The last two principles are rather meaningless in the absence of strong performance measurement. But performance measurement is also fundamental for the enactment of alignment, whereby donors increasingly hand over responsibility for priority setting and implementation to recipient governments. Beyond the five principles of the Paris Declaration, it is also recognized that donors can improve the effectiveness of their aid by being more selective on where they spend it. Again this suggests the importance of performance as measured by results. Performance measurement is closely related to monitoring and evaluation (M&E), another key concept in the new aid architecture. Hatry (1999:3) defines performance measurement as “*measurement on a regular basis of the results (outcomes) and efficiency of services or programs*”. The United States General Accounting Office (GAO) defines it as the “*ongoing monitoring and reporting of program accomplishments, particularly progress toward pre-established goals*” (GAO, 2005:3). Performance management refers to a broader process including performance measurement and the effective utilization of performance measures (e.g. in decision making). While M&E and performance measurement have moved to the centre of attention in the case of aid to recipient countries, the focus of bilaterals on measuring the performance of the Multilateral Organizations (MOs) they generously fund is more recent. This is somewhat surprising as multilateral aid has always been and remains one of the cornerstones of the aid architecture. Bilaterals appreciate that MOs tend to be neutral, well-informed and effective development actors (OECD/DAC, 2009a) and have consistently channelled about one third of their ODA through MOs (DAC countries average; OECD/DAC, 2009b). Although the attention for MO performance is quite recent, bilateral donors have plenty of good reasons to measure the performance of MOs. Three main and interrelated arguments can be identified. A first and very important argument is accountability. Bilateral donors need to be able to justify their multilateral spending and report this to their national constituencies. In the current global economic downturn, with aid budgets coming under pressure, knowing what results are being achieved is becoming even more important. A second crucial argument why bilaterals have an interest in measuring MOs’ performances is in order to make funding decisions. As bilateral donors want ‘value for money’, i.e. development results for their funds, they need to know what results MOs are producing. On this basis bilateral donors can allocate more funds to the better performing MOs or reduce funding for bad performers. Bilaterals also need information on the effectiveness of MOs so that they can better inform their multilateral strategy. Beyond adapting the size of their funding, bilateral donors can enhance their participation in the governing bodies of MOs: hold MOs accountable, push for reforms towards greater effectiveness, making sure MOs have concurring priorities, etc. Arguably, financing MOs while influencing their management can produce greater effects than financing classic bilateral aid because of the scale, budget and geographical scope of most MOs (CIDA, 2006a; DANIDA, 2008).

But how should bilaterals measure the performance of MOs? The principles of harmonisation and alignment may well be equally relevant in streamlining the relationship between bilateral back-donors and MOs as in streamlining those between public donors and recipient governments. From this vantage point, bilaterals should rely more on the reporting function of MOs (alignment) and not try to assess MO performance using their own bilateral assessment systems. The reform agenda would then consist of MOs having to improve and expand their M&E systems

in order to bring them to international standards, while not being any longer burdened by the high transaction costs of bilateral assessments. However, MOs' M&E systems are currently not meeting international standards, while bilaterals have their own accountability requirements to deal with. As a consequence, bilaterals keep on using their own assessment systems, implying transaction costs for bilaterals who have to develop and execute assessments and for the MOs who have to comply with numerous questionnaires and other information requirements. Recently however, this seems to be changing as some influential bilateral donors are increasingly harmonizing their efforts to assess the performance of MOs.

The goal of this paper is to evaluate the current state of affairs and to feed into the debate on how bilateral donors should effectively and efficiently manage the performance of MOs. In the following chapter we will briefly review how bilateral donors are currently 'assessing' MOs' performance. To give input in the debate on how bilaterals should move forward, we subsequently analyze how donors are actually 'managing' the performance of MOs (chapter three) using three case studies. This allows us to go beyond the mere analysis of how donors are monitoring and assessing MOs' effectiveness, and include how these donors integrate and use the information they obtain (managing performance). This is an important but often neglected issue, as donors have primarily focused on developing methodologies to assess MOs' performance, often without paying attention to the usefulness of these approaches.

This paper is partly based upon on a previous paper (De Koster, 2009) which included a review of current bilateral initiatives and an in depth case-study of the Belgian context. Data sources included a literature and documentary review, and open-ended interviews with staff of the Belgian aid agency. This paper goes beyond the previous one by adding two additional case-studies of the Netherlands and the United Kingdom, drawing upon document review, open-ended interviews with staff from bilateral agencies and additional literature. Through the comparative analysis of donor practices, major differences among donors are revealed. The structure of the article is as follows: it starts with a brief overview of the most important evolutions regarding this debate and a review of the current situation. Next, it confronts the general patterns described in secondary data sources with performance management literature and evidence of the actual practices of three bilateral donors (Belgium, the Netherlands and the United Kingdom). The final section discusses some key challenges and suggests possible ways forward for bilateral agencies.

## 2. ASSESSING MULTILATERALS' PERFORMANCE

Since the turn of the century there has been a burgeoning of bilateral initiatives that attempt to assess the performance of MOs, despite increased efforts of MOs to improve their M&E systems (CIDA, 2006a; DANIDA, 2008; Obser, 2007). These approaches show a wide variety of methodology, scope and type of effectiveness covered, but remain rather incomplete (DANIDA, 2008). Although it would be commonsense for bilaterals to simply rely on MOs' M&E systems, it seems unlikely that bilateral assessments will disappear in the near future. This is due to the fact that MOs still report on their performance in a fragmented manner and are inconsistent in terms of regularity, quality, scope and transparency (CIDA, 2006a; DANIDA, 2008). At this point, bilaterals cannot rely on multilaterals to report on their performance yet, and thus need to find alternative ways to satisfy their information requirements (DANIDA, 2008). Notwithstanding the current need for additional M&E exercises, the final objective remains a more systematic use of MOs' own M&E (CIDA, 2006a; ODI, 2005). Examples such as the external evaluation of FAO or initiatives as the Report on IFAD's Development Effectiveness (RIDE) indicate that it is indeed possible for MOs to adequately assess their own performance and learn lessons from it (IFAD, 2007; IFAD, 2008; OECD/DAC, 2009a). The observation that effective reporting by MOs (and the use of this information) is feasible should encourage bilaterals to choose for assessments that contribute to the reinforcement of MOs' reporting systems instead of only burdening them. However, although bilateral donors generally acknowledge the importance of relying on and strengthening MOs' own M&E systems, bilaterals have so far mainly assessed MOs' effectiveness from the outside, often using non-transparent approaches (OECD/DAC, 2009a). Acknowledging the shortcomings of their individual assessment exercises, bilateral donors presently increasingly invest in more coordinated and aligned initiatives and are moving towards three harmonized approaches that just might become the only games in town: The Peer Reviews of evaluation functions of the DAC Network on Development Evaluation (EvalNet), the Common Approach of the Multilateral Organization Performance Assessment Network (MOPAN), and an approach recently proposed by the Canadian International Development Agency (CIDA). These approaches, which are the main subject of this chapter, allow donors to harmonize their measurement efforts in order to avoid needless overlap and thereby reduce overall transaction costs (for both bilaterals and multilaterals). Before reviewing these three approaches, we will briefly outline the shortfall of previous initiatives.

### 2.1. The unbearable lightness of bilateral donorship

While the rationale for assessing MOs' effectiveness is clear-cut, all actors involved acknowledge that measuring effectiveness is a challenging endeavour (see 3.1). Particularly relevant in this context is the renowned principle-agent problem, which implies less than optimal access for bilaterals to information on MOs' performance (Yang, 2008). This is particularly valid for small donors (e.g. Belgium) who lack the resources and the influence to adequately measure the performance of MOs (hence the unbearable lightness of bilateral donorship)<sup>[1]</sup>. Still, there has been a burgeoning of bilateral initiatives in recent years, of which none succeeded in assessing MOs' performance<sup>[2]</sup> in a comprehensive manner (DANIDA, 2008). Next to the evalua-

---

[1] E.g.: it will be very difficult for a small donor to influence a MO towards facilitating an assessment of its own organization.

[2] For a detailed overview of initiatives, see DANIDA (2008) and De Koster (2009).

tion of MOs, three types of assessments carried out by bilaterals have been identified (DANIDA, 2008): i) Results-Based Management (RBM) assessments<sup>[1]</sup>, ii) partnership behaviour surveys and iii) mixed initiatives.

The RBM assessments focus on the organizational effectiveness of an MO, which is considered a proxy for MOs' overall effectiveness. This type of assessment looks at the way MOs are 'managing for development results' (MfDR). Denmark for instance has used this type of assessment<sup>[2]</sup>. Information from RBM assessments tends to be very comprehensive and factual. However, the required methodology is considered as very complex and resource consuming since RBM systems themselves are very complex. Another general disadvantage of RBM assessments is their focus on the RBM at headquarters level, neglecting the MOs country level practices. Only the system as such is evaluated and not the behaviour of the MO at country level or the aid effectiveness of the MO (DANIDA, 2008; Scott, 2005).

The second type of assessments, adopted in countries such as the UK and the Netherlands, focuses on the partnership behaviour of MOs at the country level (e.g. Burall et al., 2007). This is mostly done through questionnaires sent to bilateral staff located in field offices, which questions them on several aspects of MOs' practices (e.g. does a MO align with other donors?). The advantage of this type of assessment is the relatively low transaction cost. A general disadvantage, however, is the relatively low information value as only indirect information about the relations of MOs with partner countries and other aid agencies at country level is collected (DANIDA, 2008). Thirdly, there have also been attempts by, amongst others, Denmark and Canada to approach MOs' performance in a more synthetic way. These initiatives use different measures of effectiveness (e.g. on partnership behaviour surveys and RBM), thereby triangulating data. Although the information value of these approaches is high, they are also very time - and resource - consuming (DANIDA, 2008; OECD/DAC, 2009a).

While bilateral donors have launched numerous initiatives to assess the effectiveness of MOs in recent years, none of them proved to be satisfactory. Furthermore, these assessments entailed much duplication of efforts as donors often inquire multilaterals on similar types of information, implying transaction costs for both bilaterals and multilaterals. Key decision-makers of MOs have to respond to similar questionnaires (and other enquiries) of different donors while bilaterals have to spend time and resources in developing and implementing their individual methodologies (OECD/DAC, 2009a). As already mentioned, it appears to be very difficult for individual donors to gain internal information on the functioning of an MO and overcome the principle-agent problem that exists between bilateral donors and MOs. However, by harmonizing their efforts bilateral donors could have more leverage to hold MOs accountable and can succeed better in gaining performance information.

As mentioned earlier, bilaterals are currently in the process of harmonizing their efforts. Bilaterals use two main forums to harmonize their efforts: MOPAN and DAC EvalNet. MOPAN is led by the multilateral departments within bilaterals, while the DAC EvalNet is spearheaded by the evaluation departments of bilaterals. Although these networks have been in existence for some time now, it is only recently that bilaterals effectively seem to abandon their individual assessment exercises. The three most important initiatives are discussed below.

---

[1] Results-based management is a management strategy aimed at "achieving important changes in the way that organizations operate, with improving performance (achieving better) results as the central orientation" (OECD/DAC, 2000:3).

[2] DANIDA's RBM assessment included UNDP, UNFPA, IFAD, OHCHR and UNICEF (see Danida, 2008; Obser, 2007; OECD/DAC, 2009a).

## 2.2 MOPAN - The Common Approach

The Multilateral Organization Performance Assessment Network (MOPAN) is a network of like-minded donors in which (generally) staff from bilaterals' multilateral departments are represented. It was initially founded in response to the increased focus on the performance of multilateral organizations at country level. The network was launched in 2002 by 8 donor countries<sup>[1]</sup> and has currently expanded to 16 members<sup>[2]</sup>. MOPAN used to perform in-house surveys on the perceptions of MO partnership behaviour in countries where MOPAN members had their own bilateral programs. The opinion survey assessed in particular the quality of multilateral partnership behaviour towards national and international stakeholders. In the 2002-2008 period, the MOPAN survey has covered 11 different MOs in six survey reports<sup>[3]</sup> (MOPAN 2003, 2005a, 2005b, 2006, 2007, 2008; Selbervik and Jerve, 2003).

In 2006 there was a broad consensus among MOPAN members that further evidence was required on the effectiveness of MOs in order to adequately meet the demands of accountability and organizational learning. Despite previous efforts, there still did not exist a widely accepted 'common approach' to assess the effectiveness of different MOs, as bilateral donors kept on investing in their own initiatives. The initial approach of the perceptions-based survey needed to be broadened and deepened to provide bilaterals with more robust findings. Moreover, MOPAN reports were often downplayed by MOs as being mere perception surveys. Subsequently, MOPAN members worked towards a new 'Common Approach' that was to produce more credible and robust information for MOPAN members in order to fulfil their learning and accountability requirements (RBMG, 2007). The 'Common Approach' was anticipated to provide evidence on the organizational effectiveness of MOs, which is considered an enabling condition for good performance and a proxy for development effectiveness. The Common Approach is derived from the previous MOPAN surveys, and from seven existing bilateral assessment tools<sup>[4]</sup>. The aim of the Common Approach was not only to replace the previous MOPAN approach, but to make existing bilateral assessments obsolete.

Although the 'Common Approach' also uses a perception survey, it is very different from its predecessor. An important difference is the use of a Balanced Scorecard (BSC)<sup>[5]</sup>, a widely used tool in performance management (Fryer et al., 2009; Ramalingam et al., 2009) which was introduced in the early nineties by Kaplan and Norton (1992). The Common Approach BSC uses four categories/quadrants of organizational performance: i) strategic management (corporate governance, corporate and other strategies), ii) operational management (use of performance information, financial resources management, human resources management and portfolio management), iii) relationship management (ownership, alignment, harmonisation) and iv) knowledge management (performance monitoring and evaluation, performance reporting, feedback and implementation of lessons learned). Each quadrant has a number of measures which are called Key Performance Indicators (KPI; 19 in total) and on an even lower scale every KPI has several micro-indicators (63 in total). In order to quantify the different micro-indicators, the BSC methodology uses a traffic light system. This is an objective scoring system which is cost-effective and easy to implement.

---

[1] Canada, Denmark, Germany, the Netherlands, Norway, Sweden, Switzerland and the UK.

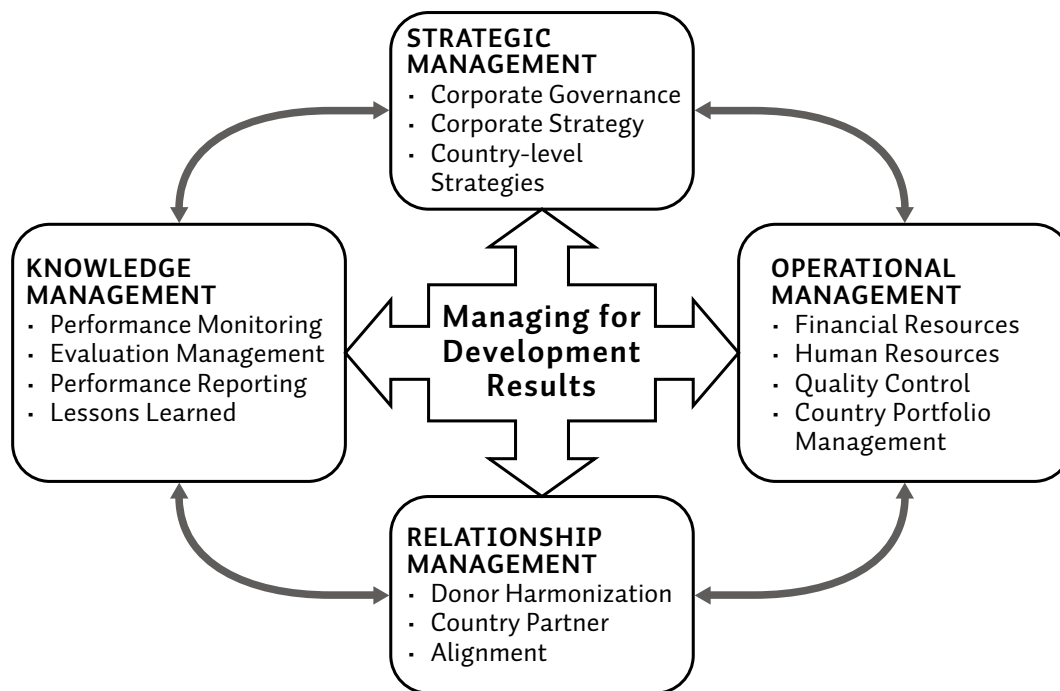
[2] New members are Australia, Austria, Belgium, Finland, France, Ireland, South-Korea and Spain.

[3] The MOs covered (some more than once): WHO, UNICEF, World Bank, UNDP, AfDB, FAO, UNFPA, UNAIDS, UNICEF, ILO, ADB, and EC.

[4] Multilateral Development Effectiveness Summaries (MDES, UK).

[5] For more info on the BSC, see [www.balancedscorecard.org](http://www.balancedscorecard.org)

**Figure 1: The four quadrants of the MOPAN Common Approach**



The 'Common Approach' collects data from three categories of respondents: MOPAN member staff at headquarters level, staff at country level and the direct (local) partners of the MO under review. These local partners can be either government, private sector organizations or civil society organizations. Respondents are also provided with 'document fact check'-documentation in order to better inform their survey responses, including the Paris Declaration survey, COMPAS<sup>[1]</sup> (only for Multilateral Development Banks) and other documents produced by MOs. The very first version of the Common Approach was published in February 2010. The report covered the African Development Bank, World Bank, UNICEF and UNDP (MOPAN, 2010a, 2010b, 2010c, 2010d). As the MOPAN network has expanded to 16 members, it has become more difficult for MOs to downplay the findings as a mere perception survey. Another advantage of the Common Approach is its methodological transparency and the fact that results are presented in both an aggregated and disaggregated way. Interestingly, the latter also elucidates diverging patterns in responses among different respondent categories: while local partners generally tend to be more positive on MOs' organisational effectiveness, MOPAN members' country staff are far more critical (MOPAN, 2010a, 2010b, 2010d). In their reactions to the 2009 MOPAN survey, MOs themselves emphasise the usefulness of a harmonised approach while they simultaneously stress a number of (methodological) concerns, including amongst others the discrepancy in responses among different samples of respondents (MOPAN, 2010a, 2010b, 2010d). In an attempt to increase the reliability of the findings, MOPAN intends to add a document review to the next Common Approach survey. While bilateral donors increasingly accept MOPAN as the way forward towards harmonisation, it remains to be seen to what extent MOPAN will indeed replace existing bilateral assessment initiatives (see also section 3).

[1] The Common Performance Assessment System (COMPAS) is a system of Multilateral Development Banks that allows them to track their capacities to manage for development results.

## 2.3 Professional Peer Reviews of evaluation functions in MOs

In 2004 the DAC Network on Development Evaluation (EvalNet) and the UN Evaluation Group (UNEG) joined forces and developed an approach to assess the MOs' own evaluation capacity and performance using a 'peer review'. Pagani (2002: 4) defines peer review as "the systematic examination and assessment of the performance of an entity by counterpart entities, with the ultimate goal of helping the reviewed entity improve its policy making, adopt best practices, and comply with established standards and principles. The examination is conducted on a non-adversarial basis, and it relies heavily on mutual trust among the entities involved in the review, as well as their shared confidence in the process". The logic behind this approach (originally proposed by the DANIDA) was the observation that since MOs have their own evaluation departments, performance assessments by bilaterals are unneeded. This approach aims to assess the quality of the evaluation function of MOs, and thus determine how reliable MOs' own performance reporting is. This results in an increased understanding of MOs' evaluation systems, and a stimulus for MOs to improve their own evaluation policies and practices. Until now, five peer reviews of MOs' evaluation function were conducted: UNICEF (pilot), UNDP (pilot), WFP, OIOS and GEF (CIDA, 2006b; DANIDA Evaluation Department, 2006; SIDA, 2007; SDC, 2009; Special Evaluation Office, 2009).

The structural aspects of the evaluation function of MOs are the main focus of the examination. The peer review addresses the general evaluation policy of the MO, the various levels of evaluation in the MO (governance level, central management functions and the central evaluation unit, decentralized evaluations, etc.), the existing evaluation capacity, the use of appropriate evaluation methods and the quality of the evaluations. The 'core assessment question' of the peer reviews is: "Are the agency's evaluation function and its products: independent; credible; and useful for learning and accountability purposes, as assessed by a panel of professional evaluation peers against international standards and the evidence base" (DAC/UNEG Joint Task Force, 2007: 8). All these dimensions are assessed against three evaluation principles: independence, credibility and usefulness (OECD/DAC, 1991). In order to assess the quality of evaluation systems against these DAC criteria, the peer review teams use document review, interviews and focus group discussions. Evaluation products are assessed through a meta-evaluation<sup>[1]</sup> against mutually agreed standards (DAC/UNEG Joint Task Force, 2007). As far as 'independence' is concerned, the peer reviews focus on three dimensions: (i) structural independence in terms of reporting lines, (ii) institutional independence by means of institutional factors that increase impartiality and (iii) personal independence. The second principle, credibility, is thought to be dependent upon the expertise and independence of evaluators and the degree to which the evaluation process is transparent. The last principle, the utility of evaluations, is related to the relevance and the use of findings for the decision-making process. Of course, ensuring the usefulness of evaluations is a task that is only partly under the control of the evaluation function of MOs. It is also the responsibility of managers, and member countries through their participation in governing bodies, and in commissioning, receiving and using evaluations (DAC/UNEG Joint Task Force, 2007).

The Professional Peer Reviews have three main purposes. The primary aim of the approach is to build more knowledge (and use) of evaluation systems by management and gov-

---

[1] A meta-evaluation is "the process of delineating, obtaining, and applying descriptive information and judgemental information – about the utility, feasibility, propriety, and accuracy of an evaluation and its systematic nature, competent conduct, integrity/honesty, respectfulness, and social responsibility – to guide the evaluation and/or report its strengths and weaknesses" (Stufflebeam, 2001:185).

erning bodies, which should lead to a better understanding of current evaluation quality and give an indication of the needed improvements in evaluation. This should subsequently lead to a better integration of the evaluation function into performance management, and thus improved evaluation practice and policy. A second and equally important purpose of this approach is that through sharing good experiences, mutual learning, greater internal capacity and increased external confidence in the evaluation systems of the MOs, the demands for special external performance assessments will diminish (UNEG Joint Task Force, 2007). Third, the Professional Peer Review also wants to provide MOs with a method to 'evaluate the evaluators' (through a peer review), against recognized international standards (UNEG Joint Task Force, 2007)<sup>[1]</sup>.

This approach clearly underlines the importance of strengthening MOs pre-existing RBM and evaluation systems with the ultimate objective of enabling bilateral donors to depend on MOs' own results reporting for performance assessments and thus make bilateral assessments obsolete. This is a major difference with previous initiatives and very much in line with the Paris Declaration principles of harmonization and alignment: peer reviews are jointly undertaken by bilateral and multilateral evaluation units (harmonization) with the ultimate objective of alignment with MOs' own (improved) evaluation systems (DAC/UNEG Joint Task Force, 2007).

A peer review in itself cannot assess the effectiveness of a MO. However, by testing the capacity and quality of the organization's own evaluations of effectiveness, and thus the confidence that can be placed in them, it can be a major contribution in assessing the effectiveness of those organizations. Furthermore, although a peer review only focuses on a small fraction of the organization's performance management systems, it is likely that the peer review will also shed light on other components of the MO's effectiveness (DAC/UNEG Joint Task Force, 2007).

#### 2.4 CIDA's cost-effective joint evaluation of MOs

This last approach was presented for the first time at the 9th meeting of DAC EvalNet in June 2009 by the Canadian International Development Agency (CIDA). The approach is not yet operational but it has been used before in a different version when CIDA undertook a systematic review of the effectiveness of Canadian aid through MOs (CIDA, 2009a). The review consisted of two main components: a meta-evaluation and an assessment of each agency's organizational effectiveness. From the lessons learnt of the 2008 two-tier-review, CIDA has distilled yet another 'New Approach'. Similar to the previous review experience, the first step of the proposed approach consists of a meta-evaluation. The second step is the assessment and analysis of four key sub-systems for managing for development effectiveness<sup>[2]</sup>. The third and final step of the CIDA's approach is a field mission to validate the quality of key organizational systems and procedures (CIDA, 2009b). The proposed approach starts from the observation that most initiatives so far have focused on MOs' organisational effectiveness while there is hardly any information regarding the development results of MO interventions on the ground. In October 2009, the approach was discussed in Ottawa with stakeholders from both DAC EvalNet and MOPAN. Since the approach proposed by CIDA was quite intensive and overlapped with the Common Approach, it was decided that the assessment would be downsized to a light version. Three possible scenarios are set out.

First, if the MOPAN Common Approach and DAC EvalNet peer reviews of MOs'

[1] Depending on the type of MO under review, a different set of standards can be used: for the UN organizations for example it uses the Norms and Standards of the UN Evaluation Group (UNEG, 2005; these are regrouped according to the three criteria: independence, credibility and usefulness).

[2] Program and Project Evaluation, RBM, Reporting on Development Effectiveness and Knowledge Management.



evaluation functions<sup>[1]</sup> indicate that the MO's own reporting about development effectiveness is satisfactory, bilaterals will rely on the existing reporting system and refrain from using any additional assessments. At this stage, however, not many MOs qualify for this scenario.

Second, if an MO's own reporting function is considered reliable and when sufficient evaluation information is available but not yet geared sufficiently towards development effectiveness, bilaterals will conduct an evaluation synthesis. Evaluation syntheses combine "information from more than one evaluation in order to come to general statements about an intervention" and "are a common approach to assessing overall impact" (Cooksy and Caracelli, 2005: 32). Evaluations that could be included to provide more insights into development results achieved on the ground include, amongst others, (i) agency-wide evaluations, (ii) evaluations of agency programmes, which are global or regional in scope, (iii) evaluations of an agency's complete country programme, sector-wide or thematic, (iv) evaluations of agency programming in a given country and finally (v) evaluations of large programmes/projects at field level, which are representative of the agency's approach (CIDA, 2009b). The use of an evaluation synthesis has the advantage that it can successfully aggregate existing information (White, 2005) and provide bilaterals with a much-needed bottom-line picture of MOs' effectiveness. In order to avoid duplication of efforts, it is proposed that the evaluation synthesis would be carried out in cooperation with MOPAN, covering the same MOs in the same year (OECD/DAC, 2010a).

Third, if the evaluation function and/or its products (evaluations) are deemed inadequate, additional assessments led by bilaterals (e.g. a joint evaluation) will take place in order to strengthen the MO's own development effectiveness reporting.

While it is too early to make any assessment, the CIDA/EvalNet approach has the outlook of a sound, comprehensive approach, which is in line with the PD reform agenda: it integrates existing joint assessments of MOs' organisational effectiveness (i.e. the MOPAN Common Approach and the peer review system) and seeks joint and aligned ways to complement it with evidence regarding development effectiveness. The approach has minimum transaction costs and will only take the form of a more hands-on approach when the performance reporting of an MO is problematic.

## 2.5 Towards a two-tier approach?

In the previous sections, we have reviewed the three most innovative, contemporary initiatives to measure the performance of MOs. The Peer Reviews and the Common Approach (and now additionally the approach proposed by CIDA) are increasingly being accepted as the way forward for bilateral assessments of MOs. By harmonizing resources into joint assessments, and by pooling knowledge on MOs' performance, transaction costs will be reduced, and knowledge and information increased. Currently, bilaterals are seemingly moving towards a two-tier approach in which the DAC EvalNet takes the lead in the peer review of evaluation functions and evaluation syntheses while MOPAN (mainly the bilateral multilateral departments) spearheads the Common Approach.

The first part of the two-tier approach consists of the peer reviews and evaluation syntheses. The Peer Reviews will generally give information on the quality of the evaluation function of MOs and the evaluation synthesis will give information on the quality of the evaluation products and (when possible) provide information on the aid effectiveness of MOs.

---

[1] In the case of Development Banks, bilaterals can also use information from the Common Performance Assessment System (COMPAS).

Both approaches combined can give a strong and trustworthy indication about the MOs' effectiveness and its ability to report on it. These initiatives are exceedingly complementary and are particularly useful since the ultimate aim is to rely on MOs' own performance measurement systems. The second part of the two-tier approach is the Common Approach of MOPAN which gives donors in depth information on MOs' organizational effectiveness. As the MOPAN network has expanded to 16 members, the results from the Common Approach will be difficult to ignore by MOs. Bilaterals expect that the MOPAN approach will help the MOs in improving their organizational effectiveness, and consequently their development effectiveness.

The two-tier approach discussed here gives a diverse overview of the performances of MOs and are cost-effective (burden sharing). The greatest advantage is that they stimulate MOs to strengthen their own systems. However, so far it remains uncertain whether this two-tier approach will completely replace existing bilateral initiatives.

### 3. MANAGING PERFORMANCE OF MOs

The previous sections documented how bilateral back-donors are increasingly investing in measuring MOs' performance. So far, however, they have paid much less attention to the usefulness and effective utilisation of this information. In this section, we will briefly outline the origins of performance management in the public sector and the difficulties concerning the move from performance measurement to performance management (i.e. the effective utilization of performance measures). In order to confront theory with local realities, the next section lays out how three bilateral donors (Belgium, the Netherlands and the UK) are currently managing MOs' performance.

#### 3.1 Managing Performance: a framework

Since the 1980s and 1990s performance concerns became fundamental to public-sector practice. This focus on performance arose in response to economic and societal pressures and the New Public Management (NPM) movement which originates from the private sector. NPM basically consists of reforms to structures and processes in the public sector with the aim of improving their performance (Pollit and Bouckaert, 2004). In the international aid context, the rise of NPM has coincided with the Millennium Development Goals (MDGs) and the aid effectiveness agenda (Paris Declaration, Managing for Development Results, Accra Agenda for Action). More precisely this means a greater focus on improving performance and a shift in focus from inputs and processes to outputs and outcomes, resulting in more and better performance measurement. In the context of MOs, we can define performance measurement by bilaterals as the ongoing process of ascertaining how well MOs are providing the goods they should be providing (e.g. poverty reduction, reducing the prevalence of HIV/AIDS, efficient use of resources etc.) through a continuous collection of relevant data. The utilization of this performance information transforms performance measurement into performance management (Hatry, 1999).

We have already elaborated on the importance of assessing the performance of MOs. But what can donors achieve by performance measurement and management? In theory, performance measurement and management are relatively straightforward. It sets out with the selection of performance measures for a public service which should ideally align itself with government objectives and departmental strategies (Fryer et al., 2009; Thomas, 2007). Through the continuous monitoring and assessment of agencies' performance against these measures, administrators can identify good and bad performance as well as the needed improvements. The internal utilization of this information aims to improve performance, while the external usage is expected to satisfy the public services' accountability requirements (Thomas, 2007). Robert Behn (2003) differentiates between eight purposes of performance measurement. The primary and overarching objective is to improve performance (of MOs in this case) (1). All the other uses of performance measures are means for achieving this goal (Robert Behn, 2003). A second use is that donors want to know how an organization is performing (2). A third reason refers to the control of MOs: donors need to know whether MOs have the right priorities (3). A fourth purpose of performance measurement is budget allocation (4): on which MO should a donor spend taxpayers' money? Fifth, information on the performance of MOs also needs to be used to stimulate the MO to take the necessary steps to improve performance (5). In this case information on the performance of MOs is used by the representatives of donors at the board (e.g. during the replenishment process). Donors also use this information to learn which MOs' function well and which MOs do not (6). Seventh, this information may subsequently be used to

convince political leaders, parliament and taxpayers, that the funding of a certain MO is money well spent (7). For the last purpose, Behn (2003) refers to 'celebration'. When certain goals are achieved, these accomplishments can be celebrated (8). These eight purposes of performance measurement proposed by Robert Behn refer to the whole spectrum of performance management. These 'uses' can also be recapitulated by the three main purposes for performance management of MOs mentioned in the introduction: accountability (external; related to goals 2,3,7 and 7 of Behn), performance-based funding (internal; related to 1,3,4,5) and performance-based governance (internal; related to 1,2,3,5,6,8)<sup>[1]</sup>.

Although the virtues of performance management have been acknowledged, there has also been much criticism from authors who point to some dangers and limitations of performance management. This is amongst others evident from the large stream of critical literature which describes the obstacles to performance management (Bouckaert and Peters, 2002; Fryer et al., 2009; Mayne, 2007; Perrin, 1999; Taylor, 2009; Van Thiel and Leeuw, 2002; ) and/or its potential dangers and shortcomings (e.g. data misinterpretation; Schwartz and Mayne, 2005). Some authors also refer to the technical difficulties of performance measurement, which are very relevant in the context of bilateral back-donors that attempt to measure the performance of MOs. If anything, the measurement of outcomes of development efforts is inherently difficult due to numerous problems such as the lack of accurate, timely data as well as attribution and aggregation issues (White, 2002). In the case of funds channelled through MOs (and thus the 'contracting out' of public funds to an external actor) measurement becomes even more challenging because of the principle-agent problem (as mentioned earlier; Yang, 2008). Moreover, there are also consistency and comparability problems with performance measures of MOs, both over time and across different MOs as they often have very different mandates.

The greatest challenge of performance management is not the measurement of performance but the effective utilization of performance measures: the Achilles heel of public management systems (Behn, 2002; Bouckaert and Peters, 2002; De Lancer Julnes and Holzer, 2001; Feller, 2002; Kamensky, 1993; Lindgren, 2001; Mayne, 1999; Taylor, 2009; Yang, 2008). While the rhetoric of performance management is widely used at all levels and branches of government, the genuine utilization of performance measures to improve performance is often limited in reality. Often, measurement systems are justified by referring to good practices of performance measurement rather than to the 'performance of performance management' (Greiner, 1996). In line with this, some authors argue that performance measurement is primarily used to meet external accountability requirements rather than being embedded in a strategy of performance improvement (Behn, 2002; Gupta et al., 1994; Kamensky, 1993; Taylor, 2009). These systems thus focus primarily on 'collecting for collection's sake' as opposed to the appropriate use of proper performance measures (Bernstein, 1999; Feller, 2002). Additionally, some authors (De Lancer Julnes and Holzer, 2001; Feller, 2002; Lindgren, 2001; Mayne, 1999; ODI, 2005; Taylor, 2009; Yang, 2008) have documented the influence of other factors (e.g. political ones) in decisions regarding resource allocation, putting into perspective the practice of performance-based funding. Moreover, in the specific case of bilateral contributions to MOs, the scope for performance-based funding is further reduced given that a large component of these contributions is fixed.<sup>[2]</sup>

---

[1] Other authors have categorized the aims of performance management slightly different: e.g. accountability and performance improvement (Halachmi, 2005). Although the three uses of performance management suggested in this paper imply considerable overlap (performance-based budgeting is relevant to both accountability and performance-based governance), this division has proved useful in the analysis of case-studies.

[2] E.g. contributions to the European Development Fund are fixed and only revised every seven years. Contribu-

Limited utilisation of performance measures is particularly prevalent when there is no formal framework for performance management and thus no clear goal orientation (De Lancer Julnes and Holzer, 2001). In order to effectively implement performance management, clarity on the exact aims of performance measurement is an essential prerequisite (Feller, 2002; Thomas, 2007). So far however, the utility of performance measures has not received sufficient attention. This is quite surprising as performance information is not free and draws upon scarce financial and human resources. In fact, while the costs of performance management are definite and immediate, its benefits are neither direct nor certain (Bouckaert and Halligan, 2008; Feller, 2002; Gabris, 1986; Greiner, 1996; Halachmi, 2005; Wholey and Hatry, 1992; ). In this context, Greiner (1996) and Thomas (2007) argue that it should be critically examined whether performance measurement is actually feasible and worthwhile. This weighing of costs and benefits in the context of MOs and their back donors, is particularly relevant for small donors who have limited resources but the same information requirements as big donors.

When performance measures are effectively utilized, they can also cause negative effects. MOs whose funding is solely based upon their performance against targets (performance-based funding) may easily be tempted to 'perverse behaviour' (gaming). Moreover, as bilaterals and multilaterals face a (multi-actor) principle-agent relationship, MOs can easily influence the measurement of their performance. MOs are more likely do this when they can benefit from it (additional funding) and when the deception is difficult to detect due to information asymmetry (Yang, 2008). 'Gaming' can take different forms. Goal displacement ('teaching to the test') is particularly relevant, this occurs "when indicators become the objective, where the focus is on "meeting the numbers" rather than doing what the program was created to do or improving actual outcomes" (Perrin, 2002: 20). In the case of MOs, it means that MOs would focus on attaining those targets set by bilateral back-donors in order to get additional funding, thereby possibly neglecting the attainment of their actual objectives (e.g. MDGs, poverty reduction). By setting explicit targets for MOs and linking funding to the attainment of these targets, bilaterals could possibly provide incentives for this type of gamesmanship (Batterham, 1994; Bevan and Hood, 2004; Lindgren, 2001; Perrin, 1999; Perrin, 2002). However, this does not imply that target setting is unavoidably negative (Bevan and Hood, 2004). Some authors argue in that regard that involving stakeholders in developing mutually agreed-on performance indicators and targets will lead to more effective performance management and reduce the risk of cheating (Yang, 2008).

In this chapter we will make a comparative analysis of how three bilateral back-donors (Belgium, the Netherlands and the UK) are managing the performance of MOs. Applying performance management theory to analyze diverse country orientations to the management of MOs' performance has been done earlier by Andreas Obser (2007). For this purpose, Obser used a framework developed by Bouckaert and Halligan (2006; 2008) which identifies four distinct models/levels of performance management: i) Pre-Performance, ii) Performance Administration, iii) Managements of Performances and iv) Performance Management. In their work, Bouckaert and Halligan (2006) identified a number of countries and the level of performance management they use. In his paper, Obser uses the models developed by Bouckaert and Halligan to frame the practices of bilateral back donors in their wider context: the national culture and the institutions of those countries. The influence of these two factors has been put forward earlier (Perrin, 2002; Pollitt, 2006) as crucial indicators for the orientation of public agencies towards performance management. First, the administrative and cultural context of a country and more precisely the

---

tions to WB International Development Assistance (IDA) are also for the largest part fixed: in the case of Belgium 68.7%, 71.44% and 78.89% of the contributions were fixed in 2006, 2007 and 2008 respectively.

degree to which performance management is a 'way of life' in a country's public sector is considered fundamental (Bouckaert and Halligan, 2008; De Lancer Julnes and Holzer, 2001; Feller, 2002; ODI, 2005; Saltmarshe et al. 2003). Several authors highlight that the development of such a performance culture is the greatest challenge for performance management to be met (e.g. Perrin, 2002; Ramalingam et al., 2009). Second, the degree of performance management implementation is also often associated with institutional factors. Drawing upon comparative research and Historical Institutional Theory (Lijphart, 1999), Pollitt (2006) for instance points out that it is easier for majoritarian systems of government to implement public management reforms (e.g. performance management) than for consensualist forms of government. He argues that a consensualist form of government is expected to be less suitable to effective performance management because of its lower capacity to implement coherent, result-oriented policymaking as well as to make straightforward use of performance measures and performance management reforms in general. This stems from the fact that they have to operate in a highly fragmented political arena characterised by many veto points which makes strong evidence-based policymaking more difficult (Pollitt, 2006). In contrast, majoritarian forms of government are able to push performance management (reforms) further and faster because they operate in a one-party system of government with fewer veto points which facilitates forceful policymaking (Lijphart, 1999; Pollitt, 2006; Pollitt et al. 2010; Varone et al., 2005). Therefore, these forms of government find it easier to set clear, useful and measurable government targets. Another important factor is the extent to which a national parliament is active and has the power to hold the government accountable. When a parliament is weak, the executive government (and subsequently government agencies) will have fewer incentives to measure performance as a strong call for accountability is lacking (Hofstede, 2001; Lijphart, 1999; Varone et al., 2005; Weaver and Rockman, 1993).

In their framework Bouckaert and Halligan (2006, 2008) set out 4 levels of performance management. The first level of performance management has been dubbed 'Traditional/Pre-Performance' and would in this context refer to development agencies in which multilateral performance goals can be found but which cannot be considered as fully developed. In these cases, multilaterals' performance will be generalized and diffuse with goals not defined in exact performance indicators, but in a more general and vague manner. Data on performance is rather scarce and has little information value, will not be integrated into documents and procedures, and will consequently have a low utility. In this type of performance management, ideas about performance are rather intuitive as there is no actual strategy available. Genuine awareness about the importance of managing performance is lacking at this level (Bouckaert and Halligan, 2008; Obser, 2007).

**Table 1: Four models of performance**

	<b>Level 1: Traditional/Pre- Performance</b>	<b>Level 2: Performance Administration</b>	<b>Level 3: Managements of Performances</b>	<b>Level 4: Performance Management</b>
<b>Measuring</b>	Intuitive	Administrative data registration	Specialized performance measurement systems	Hierarchical performance measurement system
<b>Integrating</b>	None	Some	Within different systems for specific management functions	Systematically Integrated
<b>Using</b>	None	Limited	Disconnected	Coherent, integrated, comprehensive, consistent
<b>Limitations</b>	Functional Awareness	Ad hoc selective	Incoherence	Complex

Source: Bouckaert and Halligan (2008)

At the second level of performance management, 'Performance Administration', a commitment to performance management is present, but the relationship between performance measurement and the utilization of these measures may not be explicit or well developed and its application is ad hoc and rather incoherent. The measuring of MOs' performance is part of another administrative procedure that may be part of an administrative setting, but not of a managerial or policy context. Put differently, information on the effectiveness of MOs generated from these administrative procedures is disconnected from performance improvement strategies. Sophisticated rules developed for collecting information on MOs' performance (e.g. institutional strategies) do not generate information to be used in performance-based governance or performance-based funding. It is much more about the registration of resources used and the way in which procedures are being implemented. The information is mostly used for internal reporting purposes and is rather limited (Bouckaert and Halligan, 2008 ; Obser, 2007).

Model 3, entitled 'Managements of Performances', is in between Performance Administration and Performance Management. This model performs better than the previous one but also has its limitations. In countries which fit the 'Managements of Performances' model, the link between performance measurement and the utilization of these measures is present but the connection is underdeveloped. At this level of performance management, there are several types of performance and measurement systems, going beyond performance administration. A diverging set of performance measurement systems is thus feeding information into a disconnected set of management functions (e.g. different departments), resulting into different performance measures for different purposes, which are not necessarily linked in a hierarchical or logical way, leading to a multiplicity of managements of performances (e.g. financial management, management on country level, etc.). Put differently, performance is measured extensively, but the utilization of performance measures remains limited and inconsistent. These function-based measurement systems are rather inconsistent, incoherent and disintegrated (Bouckaert and Halligan, 2008; Obser, 2007).

The fourth level, 'Performance Management', is characterized by a solid, systematic measurement system that goes beyond administration. There is a hierarchical, coherent system

that measures MOs' performance. It includes a systematic integration of performance information into documents and procedures, with the purpose of using it in a coherent strategy to improve performance and in order to better inform funding decisions (i.e. performance-based funding and performance-based governance). It considers different assessment systems for different purposes, which are all connected. At this level of performance management, an explicit policy on performance management is present (Bouckaert and Halligan, 2008; Obser, 2007).

The following section explores to what extent different bilateral donors (UK, Netherlands, Belgium) are measuring, integrating and using information on MOs' performance. We review how different bilaterals are dealing with MOs' performance and which factors contribute to their relative success or failure. The section documents experiences of performance management in a very particular area of government and qualifies the sometimes vague rhetoric of performance management. In doing this, it complements other studies which have examined performance management experiences of a government as a whole and/or particular branches of a government, or which have invested in international comparisons (Bevan and Hood, 2006; Bouckaert and Halligan, 2008; Gao, 2009; Obser, 2007; Pollitt et al., 2010; Taylor, 2007). The use of performance management literature and the framework produced by Bouckaert and Halligan, allows us to analyse the practices of bilateral donors and detect the causes of their success or failure in different component of performance management (measurement, integration, use). Our comparative analysis draws upon insights from theoretical literature, document review (including annual reports, legislative texts, performance reporting, parliamentary reports, strategy papers) as well as primary data collection. Primary data was gathered through open-ended interviews with staff of the multilateral departments and evaluation offices of the three bilateral donors as well as through participant observation<sup>[1]</sup>. Based on insights from theoretical literature, we anticipate performance management to be most advanced in the UK: it is characterised by a majoritarian system and literature has documented the strong performance culture in other branches of UK public services (Pollitt, 2006). We expect that performance management will be least developed in Belgium: it uses a consensualist system and financial resources are scarcer in comparison with the Netherlands and the UK.

### 3.2 Performance Administration: Belgium

Belgium, in contrast with the Netherlands and the United Kingdom, was not included in the analyses of Bouckaert and Halligan (2008) or Obser (2007). However, from the analysis below of the context of Belgian aid to MOs, it would seem that Belgium is to be located in the second model: Performance Administration<sup>[2]</sup>. Following Bouckaert and Halligan (2008), the performance administration model is characterized by a low focus on results, ad hoc measuring, a focus on inputs and a limited use of performance information, which fits the Belgium case (in the context of MOs) quite well. This conclusion can be drawn from multiple observations. Some stem from cultural and institutional causes, others from financial and methodological factors. Before analysing the way in which Belgium is measuring, incorporating and using performance information, we will briefly outline the Belgian context.

---

[1] One of the authors did a two-months internship within the Special Evaluation Office of the Belgian Directorate for Development Cooperation (DGCD).

[2] Although literature (e.g. Varone et al., 2005) hints at this level of performance management, locating the whole of the Belgian public sector at the performance administration level on the basis of the analysis below, would be far-fetched. In the case study, the focus is therefore exclusively on the actors involved with MO' performance, although some of the causes of Belgium's position are valid for the whole federal Belgian level.



**Table 2: Performance Administration - Belgium**

	Measuring	Incorporating	Using	Limitations
<b>Model 2: Performance Administration</b>	Administrative data registration: limited and selective	Some	Limited: reporting, internal	Ad hoc selective

### 3.2.1 Context

The Directorate-Generale of Development Cooperation (DGDC) is the federal body that controls the majority of Belgian ODA. Belgium has always been a strong contributor to the multilateral system, channelling an average of 37% of its ODA through MOs in the 2006-2008 period (OECD/DAC, 2009b). In 1999, as a part of a new law on development cooperation, the policy towards MOs was renewed (Belgisch Staatsblad, 1999). The law announced the reduction of the amount of MOs Belgium funds to around 20, in order to avoid the fragmentation of the multilateral budget. The Royal Decree of 7 April 2000 noted that MOs' performance would be crucial in their selection (Belgisch Staatsblad, 2000). In 2008 Belgium announced to move towards a full core policy, cutting back its earmarked contributions and giving as much 'core' contributions as possible<sup>[1]</sup>. The Belgian state expects that this will increase Belgium's influence in the governing bodies of MOs, and enable Belgium to push for a more result-oriented policy of MOs. Furthermore, Belgium reduced the number of MOs receiving voluntary contributions from 23 to 21, which should allow a better follow-up (DGDC, 2008, 2009).

We can define two broad sets of actors within the DGDC that are directly involved in the (performance) management of MOs and will be included in this analysis: the 'Multilateral Department (D4)' and the 'Special Evaluation Office'. The D4 department (heads of unit, attachés<sup>[2]</sup>, file administrators) is responsible for the follow-up and the overall management of MOs. The Special Evaluation Office is responsible for the evaluation of Belgian ODA.

### 3.2.2 Measuring Performance

There are several elements which elucidate the fact that the DGDC is mainly 'administering' performance and as a result only measures MOs' performance in a partial manner. A first and more general reason stems from the institutional context of Belgium, which leads to a deficient performance or 'evaluation' culture in Belgium<sup>[3]</sup> (Varone et al., 2005). In the context of measuring the performance of MOs, this is illustrated by the absence of any formal strategy on the matter. If we compare Belgium to other countries (Varone et al., 2005), Belgium is clearly lagging behind in the institutionalization of monitoring and evaluation, which also points to a deficient performance orientation. Although improvements have been made in the DGDC, a genuine performance culture remains largely absent as three elements that caused this initial laggardness remain.

[1] Core contributions are not earmarked by programme, sector, theme, country or region, and are more difficult to evaluate than earmarked contributions (also called multi-bi aid).

[2] Attachés represent Belgium in countries receiving Belgian ODA and at the HQ of key Multilateral Organizations (e.g. at FAO in Rome).

[3] Although literature (e.g. Varone et al., 2005) hints at this level of performance management, locating the whole of the Belgian public sector at the 'Performance Administration' level on the basis of the analysis below, would be far-fetched. In the case study, the focus is therefore exclusively on the actors involved in MOs' performance management, although some of the causes of Belgium's position are valid for the whole federal Belgian level.

A first cause stems from the fact that the political spectrum is exceedingly fragmented in Belgium which leads to fragile coalitions which are consequentially not eager to commission critical policy evaluations (Greiner, 1996; Varone et. al., 2005). The MPs of coalition parties have little to gain with objective evaluations or strict monitoring (political stability is more important) and have lost their watchdog role. A related root cause stems from the limited parliamentary control over the government: a majority of the bills passed in parliament are introduced by the government itself, meaning that MPs have long lost their agenda-setting role. A last and crucial factor leading to a weak results orientation is Belgium's Federalism. From the 1980s onwards, competencies are being devolved to lower echelons of power, leading to the de facto fragmentation of power. This creates many overlapping competencies among different levels, which in turn makes evaluation and coherent policymaking in general more difficult (Varone et al., 2005). The amalgamation of the DGDC and the Federal Public Service of Foreign Affairs, and the resulting overlap and fragmentation of competences related to the performance of MOs (both often have relations with the same MOs), is illustrative here. This leads to a disconnected set of functions which results into a situation in which nobody knows who is doing what.

As a consequence of this absent 'performance' culture, the Multilateral Department has not set up any measurement system for MOs, nor has the National Audit Office evaluated MOs or the Belgian policy towards MOs (OECD/DAC, 2009a). The D4 department only collects performance information from second degree performance documents (e.g. reports to the board) and day to day interaction with MOs. Genuine in-depth performance information is only collected in an ad hoc manner, e.g. in the event of a parliamentary question, or when a specific evaluation leads to the reform of a MO. The extent to which the performance of a MO is monitored is also largely dependent on the particular staff involved. Some staff members simply have more time (e.g. because they have fewer or smaller MOs in their portfolio) for performance measuring or attribute greater importance to the issue of MOs' effectiveness. This is a result of the understaffing of certain positions in D4 and more importantly because of the absence of a formal strategy on the management of MOs' performance. The Multilateral Department does have Institutional Strategy papers for the MOs it supports, but these do not include explicit performance measures. Recently however, (February 2010) a shift towards more systematic performance measurement occurred, as Belgium became the most recent MOPAN member. Belgium will thus participate in the MOPAN 2010 Common Approach, whose report will be published by January 2011.

The Special Evaluation Office too, has no particular strategy for the evaluation of MOs. Although there is a clear awareness of the importance of performance measurement, the only activity related to the assessment of MOs (and thus measuring) has been the lead function by the Belgian Evaluation Department in a Professional Peer Review of the evaluation function of the Global Environment Facility (GEF) (Special Evaluation Office, 2009). Though, the participation in DAC EvalNet is a positive element since it is a forum in which ideas about assessing MOs can be exchanged (e.g. on the new CIDA initiative).

### **3.2.3 Integrating and using performance information**

The lack of any formal assessment system logically also curtails the integration and utility of performance data in the DGDC. The usage of available data (e.g. reports made by file administrators or attachés) is perceived as rather limited since it is mostly used internally (D4). When information about the effectiveness of MOs is being channelled (i.e. integration) to higher echelons (e.g. the cabinet), top-down feedback on the final use of this information is lacking,

resulting in diminished incentives to closely monitor the performance of multilaterals. Belgium has made funding decisions on the basis of performance (performance-based funding) in the past, but evidence on the link between allocation of funds and performance measures is absent. Performance information is rarely used externally by attachés (performance-based governance) and is obviously limited since information on the performance of MOs is rare. Even when information is available, it remains grossly underused due to a relatively low results focus.

### 3.3 Managements of Performances: the Netherlands

In the Bouckaert and Halligan (2008) framework, the Netherlands is classified under the third model: Managements of Performances. This is characterized (see 3.1 for further clarification) by a high commitment to performance, with specialized performance measurement systems being present. However, the link between performance measurement and the utilization of performance measures is underdeveloped and parallel systems of performance management exist.

The Netherlands has a long tradition of performance measurement initiatives and policy evaluation. In 1999 a new outcome-oriented budget structure was launched named ‘from policy budget to policy accountability’ (VBTB). This initiative aimed at making the budget and annual ministerial reports more transparent and more performance-based in order to increase accountability. The new budget is divided into policy lines and ministers have to explain the allocation of resources by means of strategic and operational objectives and efficiency and effectiveness measures (Bouckaert and Halligan, 2008; Pollit and Bouckaert, 2004). However, since 2006 the Comply-or-Explain principle was put into force. This means that a minister can choose not to comply with the requirement to report on performance as long he or she explains why it is not useful or relevant to include performance information. Since the cost of collecting valid, accurate and useful performance information is very high, the coverage of performance measures is very variable between ministries. Prior to the analysis of how the Netherlands is measuring, incorporating and using performance information, we will briefly outline the Dutch context.

**Table 3: Managements of Performances – the Netherlands**

	Measuring	Incorporating	Using	Limitations
<b>Model3: Managements of Performances</b>	Specialized performance measurement systems	Within separate systems for different management functions	Disconnected	Incoherence

#### 3.3.1 Context

The Netherlands is considered a frontrunner in the field of development cooperation, and has a strong commitment to the international aid effectiveness agenda. This is illustrated by the Policy Memorandum ‘Mutual Interests, mutual responsibilities: Dutch Development Cooperation en route to 2015’ which was published in 2003. In this memorandum, the boosting of the quality and effectiveness of aid is identified as an essential element in achieving Dutch priorities (OECD/DAC, 2006a). The Dutch ODA/GNI coefficient amounts to an impressive 0.8% of which most is controlled by the Ministry of Foreign Affairs (80%; MFA). The DG of International Cooperation is part of the MFA, which however does not lead to a duplication of functions with

Foreign Affairs (as is the case in Belgium) since both are integrated in mixed departments<sup>[1]</sup>.

In the 2006-2008 period, on average 24% of ODA was channelled through MOs as core funding, making the Netherlands the 8th largest contributor to MOs (OECD/DAC, 2009b, 2010b). Although it was proposed to increase the proportion of multilateral aid to 50%, the funding of MOs remained relatively stable as political support for this proposal was limited and because of parliamentary concern over the effectiveness of MOs (OECD/DAC, 2006a, 2009a). In a strategy paper on multilateral aid published in 2009, however, the Netherlands announced to increase its support to MOs. In this paper, obtaining information on and improving of MOs' performance is labelled a key priority (MFA, 2009a).

In the Netherlands we can define different actors who are involved in the performance management of MOs. Similar to Belgium, there is an external evaluation office, i.e. the Policy and Operations Evaluation Department (IOB), which is responsible for the evaluation of the MFA, and thus for contributions to MOs. Additionally there is one department responsible for the overall management of UN agencies and IFIs, the UN and International Institutions Department (DVF), and a distinct department responsible for collaboration with the EU (European Integration Department (DIE))<sup>[2]</sup>. Furthermore there is also an effectiveness and quality department (DEK) which reports on the quality of Dutch development cooperation (DEK). Since it has not reported on the effectiveness of multilateral aid, it will not be included in this analysis (MFA, 2009b). During an interview it was noted that this department will include multilateral aid and its effectiveness in future reports.

### 3.3.2 Measuring performance

Although there is strong public support for development cooperation in the Netherlands, there has been a growing demand for results by parliament and the public in general in recent years (OECD/DAC, 2006a; Tweede Kamer der Staten-Generaal, 2009). The Netherlands has an institutionalized 'performance' culture (Varone et al., 2005) and it is, ipso facto, also committed to measuring and managing the performance of their multilateral contributions. As early as 1999, the National Audit Office made an evaluation of Dutch multilateral policy in which the lack of evidence on the effectiveness of MOs was denounced (Tweede Kamer der Staten-Generaal, 1999). In 2001, the MFA launched a monitoring system for MOs focusing on their partnership behaviour at country level: the Multilateral Monitoring System (MMS). This MMS consists of a questionnaire sent to MFA staff at country level (e.g. embassy staff) and gives the Netherlands information on partnership behaviour of MOs and relevance to the Dutch priorities as outlined in the 'Mutual Interests, mutual responsibilities' memorandum (Obser, 2007; OECD/DAC, 2009a). The MMS is executed every two years in 36 partner countries. In the future, the MMS will cease to exist as the Netherlands are expecting the Common Approach (MOPAN) to render the MMS obsolete (indicators of the MMS were integrated in the Common Approach). The Netherlands has been a member of MOPAN since its establishment in 2003 and is strongly committed to the Common Approach.

As mentioned, since 2009 the Netherlands has a single strategy document which outlines its overall policy and interaction with MOs (MFA, 2009b). Guidelines on the type of information to be collected are, although rather vaguely, articulated in this paper. Next to the MMS, the DVF retrieves information from the reports of the board, audit reports, strategic plan-

[1] I.e. one department working on the same issue with representatives from both Foreign Affairs and International Cooperation.

[2] This department has not been included in this analysis due to a lack of information.

ning, evaluations, results frameworks, thematic policy areas (e.g. HR), day to day interaction with MOs (e.g. by file administrators) and MOPAN (MFA, 2009b). The main focus is on organizational effectiveness while evidence of aid effectiveness remains scarce. On the basis of these different information sources, the DVF produces a scorecard for every MO. These scorecards provide general information on the mandate of the MO, its organization and on the relevance of its work for Dutch aid priorities. These scorecards are continuously being updated, but are not publicly accessible since the method needs further refinement and because they are primarily meant for internal use.

The second key actor in the Netherlands, the external evaluation office (IOB), is often asked to assist the DVF in judging the quality of evaluations produced by MOs. IOB is a strong supporter of the peer reviews of evaluation functions organized by DAC EvalNet and has taken the lead in such peer reviews in the past. According to IOB staff, evaluations produced by MOs themselves have so far remained underused due to a lack of time, trust and interest of bilaterals. They strongly advise the multilateral department (and bilateral donors in general) to make more use of the evaluations produced by MOs in cases where peer reviews have judged the evaluation functions of MOs to be satisfactory. While it is too early to make any definite conclusions, the IOB also considers the CIDA-initiated approach (with its focus on development effectiveness) valuable.

### 3.3.3 Integrating and using performance information

Clearly, the Netherlands invests considerably in the measurement of MOs' performance. Despite these investments, systematic integration of this performance information seems to be lacking since a formal framework does not exist. For the evaluation department, integration (and use) of performance information is limited to transferring the peer reviews of the evaluation function of MOs to the involved departments. The DVF department integrates most of its performance information in the Scorecards of every MO. Furthermore the DVF department reports its findings to other stakeholders in and outside the Ministry, to their permanent representations and to the Minister. Information on the performance is not (yet) used in the annual reports of the DEK department (see supra), and thus the use of performance information for accountability purposes is not explicitly present.

Performance information feeds into Dutch policy and practice towards MOs in two major ways. Firstly, performance information is used to make performance-based allocation decisions. MOs that prove to be effective can be rewarded by additional funding while weak performers will receive reduced funding, or no funding at all. The direct link between performance information and funding decisions is however not clear as formal mechanisms do not exist. Allocation decisions are additionally not only made on the basis of performance: the politics of development, the perception and priorities of the minister and the relevance to Dutch priorities also play an important role. The significance of MOs priorities being concurrent with Dutch development priorities is also outlined in the 2009 strategy note (MFA, 2009b). If the work of an MO on country level is not relevant to these priorities, funding can be suspended as has happened in the past.

Secondly, the DVF furthermore uses performance information for their governance towards MOs (e.g. in the MOs' boards). The Netherlands strives to influence MOs towards greater effectiveness and better reporting, and tries to ensure that Dutch objectives are served as much as possible. However, the lack of a formal framework makes the link between performance measurement and its use in performance based governance ad hoc and dependent upon

the person representing the Netherlands in the board of the MO (or the concerned file administrator) and the importance that this person attributes to the performance of MOs. Information on the effectiveness of MOs produced by MOs themselves also remains largely unused or under-used, as DVF staff question the reliability and credibility of this information.

The Dutch case clearly goes well beyond that of Belgium, both in terms of assessment and monitoring and in terms of effective utilisation. While the linkage between measurement and utilisation is more developed than in the case of Belgium, a formal framework to move beyond performance measurement is currently lacking, which leads to an ad hoc use of existing information on MOs' performance.

### 3.4 Performance Management: United Kingdom (UK)

The Bouckaert and Halligan framework links the UK to the last model: performance management. The UK model of public service reform is based on top-down performance management, competition and contestability in service provision, citizen choice and voice, and strengthening capability and capacity of officials; all of which have performance elements (Bouckaert and Halligan, 2008). This has added up to a comprehensive, hierarchical model based on Cabinet office and Treasury agendas, which has at its centrepiece the Treasury's regular spending reviews and the Public Service Agreement (PSA) framework. The framework has allowed extensive steering and coordination of public activity and the integration of central government under a system-wide performance regime that supports Treasury's role in priority setting.

**Table 4: Performance Management – United Kingdom**

	Measuring	Incorporating	Using	Limitations
<b>Model 4: Performance Management</b>	Hierarchical performance measurement systems	Systematic internal integration	Coherent, comprehensive and consistent	Complex

#### 3.4.1 Context

In 1997 the Department For International Development (DFID) was created as a freestanding Ministry, controlling most of the UK's ODA. Similar to the Netherlands, DFID is internationally considered as a frontrunner in the development effectiveness agenda. In the 2006-2008 period, an average of 36% of UK's ODA was registered as multilateral ODA making the UK the largest contributor of multilateral aid among DAC members (OECD/DAC, 2009b, 2010b). In 2002 the International Development Act was launched which provided a clear legislative mandate on poverty reduction. Crucial for the direction of UK development policy are the White Papers (DFID, 2006, 2009a) in which the overall UK development objectives and strategies are made public (see e.g. DFID 2006, 2009a). DFID's objectives are furthermore articulated in the Public Service Agreements (PSAs). These two-yearly agreements between the Treasury and UK government departments (DFID in this case) set out policy objectives and targets for the next three years. This has pushed UK government departments to develop performance measurement systems according to the Treasury's instructions and guidelines. DFID is the lead department for PSA 29: Reduce Poverty in poorer countries through quicker progress towards the Millennium Development Goals (MDGs). In order to ensure delivery of its PSA, DFID has also set

out a number of Departmental Strategic Objectives (DSOs; e.g. “make multilateral donors more effective”), which are subsequently linked to indicators (e.g. “improved effectiveness of the UN system” or “improved effectiveness of the International Financial Institutions”) and sub-indicators (e.g. Paris Declaration targets; DFID, 2009b). The relationship between DFID and specific MOs is framed in Institutional Strategies which set out how DFID aims to contribute to achieving its White Paper and PSA objectives in partnership with MOs. These Institutional Strategies are updated every three years through a consultative process involving DFID, the MO and other stakeholders (OECD/DAC, 2006b, 2009a).

We can identify multiple actors in DFID that have a clear interest in managing the performance of MOs. Similar to the Netherlands and Belgium there is the evaluation department and an International Department. The International Department consists of three departments responsible for Europe, the International Financial Institutions (IFIs) and United Nations Organizations respectively. The International Department also contains subdivisions working specifically on the effectiveness of multilaterals (on the allocation of funds to MOs, on MOPAN).

### 3.4.2 Measuring performance

The institutionalization of performance management is very strong in the UK in general. In DFID there is also a very strong focus on measuring the performance of MOs with the aim of increasing the value for money that DFID can get from them. DFID has developed an extensive and complex measurement system for this purpose. Where DFID cannot rely on MOs’ own performance reporting (which is often the case), DFID focuses on the enabling conditions, i.e. the organizational effectiveness of MOs. For this reason DFID’s International Division initially created the Multilateral Effectiveness Framework (MEFF). This initiative used a Balanced Scorecard approach and made a strong contribution to the design of the Common Approach (see Scott (2004) for a description of MEFF and its methodology). The MEFF is not being used anymore since it has been formally integrated into the Common Approach. The MEFF used a RBM approach, assessing the effectiveness of each MO from three perspectives: internal performance, country-level results and partnership. Subsequent to the MEFF initiative, in 2007 DFID produced Multilateral Development Effectiveness Summaries (MDES). These reports collect evidence on the effectiveness of MOs drawing upon a wide array of sources. Next to (joint/donor) evaluations MDES also use MOs’ own reporting systems by utilizing the managements’ reports to the board which draw upon a vast amount of information. This information is accompanied by an estimate of DFID staff about the extent to which the organizations’ structure is appropriate to be used for accountability purposes. Information on the multilaterals’ own results framework, typically on their reporting function, is also included in the MDES. All this information is furthermore cross-referenced with information from the Paris Survey. Collectively these information sources feed into MDES and give extensive information on the organizational effectiveness of an MO.

In 2007, DFID furthermore commissioned a partnership behaviour survey in which five stakeholder groups (business leaders, civil servants, civil society leaders, government ministers, and MPs) were consulted about seven organisations<sup>[1]</sup>. This client survey examined stakeholders’ perceptions of MOs’ performance against three measures: overall development effec-

[1] African Development Bank, Asian Development Banks, European Commission, the Global Fund for AIDS, TB and Malaria, United Nations Children’s Fund, the United Nations Development Fund and the World Bank.

tiveness, harmonisation with other donors, and alignment with government priorities (Burall et al., 2007). Additionally, DFID has played a leading role in the development of the MOPAN Common Approach. DFID's evaluation division also participates in the DAC peer reviews of MOs' evaluation functions. Similar to its Belgian and Dutch counterparts, DFID's Evaluation Department also shows enthusiasm for the evaluation syntheses proposed by CIDA. The evaluation department considers the MOPAN initiative and the initiative led by the DAC as a promising two-tier approach.

The importance attached to performance measurement in DFID is also apparent in the Institutional Strategies (IS) which set out the framework for future engagement between DFID and MOs. Recently, the main element of the IS became the performance framework, which includes performance targets and indicators in areas where DFID wants to see improvement. Progress against these indicators is assessed annually. While some indicators are related to final development outcomes, many of them are measures of organisational effectiveness<sup>[1]</sup>. Monitoring progress against these indicators is achieved by relying on MOs' own reporting (where possible), MDES, feedback from DFID country offices, MOPAN reports, joint donor discussions and meetings with MOs.

Improving the effectiveness and efficiency of MOs is clearly a key priority for DFID, since it is continuously looking for ways to improve the value for money it can get from MOs. Accordingly DFID is constantly looking to improve the quality of the available data. Although DFID is primarily interested in the development outcomes of their funding, it is considered more pragmatic and realistic to focus on the organizational effectiveness of MOs for now. As one interviewee put it: "DFID wants to know which MO is most likely to use funds optimally and contribute to the outcomes we want to see".

### 3.4.3 Integrating and using performance information

Similar to the Dutch case, DFID extensively measures the performance of MOs. Unlike the Netherlands, the integration and use of information on MOs' performance is formally structured within DFID. The performance frameworks embedded within the Institutional Strategies do not only function as tools to measure the performance of MOs, but they are also the instruments par excellence to successfully utilise the information collected. DFID uses performance information to make funding decisions, to inform their multilateral policy (and thereby increase the effectiveness of MOs) and for accountability purposes.

Performance-based funding is very much present in DFID's use of performance information and is mainly based on the progress that MOs make against the indicators of the IS. DFID mainly utilizes it to reward the good performers (additional funding) and not so much to punish weak performers (withdrawing as a funder). Weak performers, or MOs with different priorities as DFID, will for example only receive fixed contributions and no voluntary contributions. DFID follows this strategy as they believe that an organization can only become more effective if they receive enough assistance. In the context of continuing proliferation of MOs, DFID retains the MOs it funds. For example: if health care is a priority of DFID and it supports the WHO for this cause, they do not shift their funding to another (health) organization when WHO is not performing as expected as this leads to further proliferation of MOs. Likewise, when a MO is very self-critical when it reports on its own performance, DFID would not cut funding since this

[1] Examples of such indicators include e.g. "WHO produces clear reporting on results achieved at the level of i) Strategic Objectives, ii) Organization Wide Results and iii) Office Specific Results" (DFID, 2009c) and "Percentage of evaluation recommendations translated into UNDP decision making and development of Country Plans" (DFID, 2009d).



would undermine the self-reporting function of MOs. Instead, good performers are rewarded by DFID through additional funding (positive incentive). In their financing scheme of UN agencies for example, DFID uses a bonus system in which 10% of the core funding for UN agencies is dependent on their performance as measured against the indicators set out in the IS.

Furthermore, DFID actively uses the information it collects in its multilateral strategy (performance-based governance). Through representation in the board and especially during replenishment negotiations, DFID uses its weight as a large donor to push for reform in a constructive manner (using information from primarily IS and MDES, MOPAN, day to day interaction, and so on). When a good performance framework is absent within a MO, DFID mainly focuses on improving the performance framework through dialogue and makes sure the MO is moving towards a system that can report on its own outcomes. Where these frameworks are present and MOs report well on their effectiveness, DFID will focus on improving the MOs' aid effectiveness.

The IS are furthermore integrated into the department's annual report which feeds into DFID's annual report. This is basically a statement to the parliament and the public and serves accountability goals. On the basis of available information and personal appreciation (in the case of lacking performance information) MOs receive a score on the progress made against indicators of the Departmental Strategic Objectives. This score indicates the likelihood that a project will deliver its results at any stage. When good performance information is lacking this assessment is done on the basis of personal appreciation. The performance of MOs is integrated in DFID's annual report about the progress made against its Departmental Strategic Objectives and which serves its own accountability goals towards the national constituencies and Treasury.

## 4. CONCLUSION AND RECOMMENDATIONS

This paper has provided an overview of the current state of affairs. We have described how donors are currently harmonizing their efforts to measure the performance of multilaterals and start to abandon their individual assessment systems. In doing this, bilateral donors are moving towards a two-tier approach. Firstly, the MOPAN Common Approach, led by bilaterals' multilateral departments, was designed to replace the numerous uncoordinated bilateral initiatives and to provide in depth information on the organizational effectiveness of MOs through a perception survey. However, establishing causal linkages between the Common Approach and its anticipated findings, and the development results MOs produce, remains particularly challenging (White, 2005). However, measuring and subsequently improving the enabling conditions will likely improve MOs' development effectiveness. Secondly, the initiatives that are being facilitated by the DAC EvalNet (evaluation synthesis and peer reviews) focus on the quality of MOs' M&E systems and products. Although the findings of these assessments and the subsequent use of this information will not directly improve the aid effectiveness of MOs, it will help improve the M&E function of the MOs under review (and the use of the results it produces) and is very much in line with the PD principles. Furthermore, the peer reviews and evaluation syntheses can fulfil accountability and learning needs of both MOs themselves and their (bilateral) back-donors. The approach proposed by CIDA in the DAC EvalNet is particularly interesting given the fact that it focuses on measuring development effectiveness, contrary to the peer reviews and the MOPAN Common Approach which focus on organizational effectiveness. It is from this vantage point that one can understand the enthusiasm among bilateral donors for this approach.

Although the harmonisation by bilaterals is advantageous and will produce more valid information in the future, donors still only have a vague understanding of why particular results occur and what to do to improve them. For bilaterals it remains very challenging to link inputs, processes, outputs and outcomes and elucidate why certain results occurred or which aspects of MOs' organizational effectiveness need to be enhanced in order to improve performance in terms of development effectiveness. Still, in the current context of the aid effectiveness agenda which imposes an ambitious reform agenda of harmonisation and alignment upon donors, the MOPAN and DAC EvalNet initiatives are definitively laudable ways forward.

Findings from the case studies have shown that, in spite of their commitments towards more harmonised and aligned approaches, the current level of harmonisation and alignment on the ground is still relatively low. In order to fill the information gap about actual bilateral practices in the area of measuring and utilizing information regarding MOs' performance, we have documented and analysed the cases of Belgium, the Netherlands and the UK. Our case studies highlight that the three donors reviewed measure and manage the performance of MOs to varying degrees. On one end of the spectrum, there is the case of Belgium, a small bilateral donor which does not strongly invest in the measurement of MOs' performance. Despite some significant ongoing changes on the ground, such as its recent MOPAN membership, Belgium still has no clear strategy for the measurement of MOs' performance or for the effective utilization of this information. The UK aid agency has, on the other end of the spectrum, a strong performance culture. Since the turn of the century DFID has invested in different ways to adequately measure the performance of MOs. DFID's approach has been strongly influenced by the broader UK context which is characterised by an institutionalised performance culture which pushes all public agencies to set targets and measure performance. Similar conclusions can be drawn regarding

the actual utilization of performance measures. While Belgium has no strategy for the utilization of performance information, in DFID measurement and utilization in the context of performance-based funding and governance are formally linked through procedures and documents. The main tools are the performance assessment frameworks embedded within the Institutional Strategies which formally structure the relationship between DFID and the specific MOs. The Netherlands are somewhat in between. Similar to the UK, they heavily invest in the measurement of MOs' performance, yet the linkage to an effective utilisation of effectiveness measures remains underdeveloped which results in an ad hoc and non-transparent use of performance measures. It has become apparent that our findings, on the measurement and use of performance measures are largely in line with the Historical Institutional theory. Belgium, a highly federalized and consensualist country, has the least developed performance management system: there is no strong call for accountability or for an increase of Belgian ODA's efficiency. While the Netherlands too has a consensualist government, discussions in Dutch Parliament illustrate a much stronger call for accountability (Tweede Kamer der Staten-Generaal, 2009). The UK, on the other hand, is able to push performance management reform further and faster because it operates within a majoritarian form of government which facilitates forceful policy-making. The importance of context and institutional setting obviously puts into perspective the degree to which DFID's approach towards MOs' performance management may be easily extrapolated to other bilateral donors. For a small donor with limited capacity such as Belgium, a cost-benefit analysis about the set-up of a similar performance management system could turn out negatively. Within the current setting however, there still remains room for improvement. Participation in the initiatives of MOPAN and the EvalNet is a laudable first step as it provides bilaterals with essential information, while each additional member boosts the credibility of these initiatives. Utilising the obtained performance measures in a strategy aimed at improving MOs' performance is a difficult and challenging next step. An interesting option might be to harmonise with other donors by organising in coalitions in the governing bodies of MOs, an option already explored by some donors (e.g. the Netherlands). More formal linkages between assessment and utilisation could also be established through the joint adoption of performance frameworks and institutional strategies. The joint creation of an Institutional Strategy for the UNDP by Denmark and the UK is an interesting precedent for other donors. The Netherlands has already announced that it will work towards joint performance frameworks with the UK and Denmark in the future (Tweede Kamer der Staten-Generaal, 2009). While it is too early to draw final conclusions in this regard, the joint, or at least coordinated, usage of performance information might not only lower costs but also increase benefits. Organising in coalitions and setting up joint performance frameworks might increase bilaterals' weight to push for reforms in the governing bodies of MOs. On the other hand, using joint performance frameworks in a joint system of performance-based funding might increase gamesmanship among MOs (e.g. goal displacement). This may partly be circumvented through setting mutually agreed targets, or by selecting different measures for different donors. The alignment to MOs' own performance measurement systems could provide a different, possibly more optimal, way to circumvent possible gaming.



## REFERENCES

- Batterham, R. (1994) **“How Can We Argue with Performance Indicators? An Analysis of the Sources of the Power and Dysfunction in the Use of Performance Indicators”**, Centre for Program Evaluation, The University of Melbourne, Melbourne.
- Behn, R. (2002) “The psychological barriers to performance management: Or why isn’t everyone jumping on the performance management bandwagon?”, **Public Performance and Management Review** 26 (1): 5-25.
- Behn, R.D. (2003) “Why Measure Performance? Different Purposes Require Different Measures”, **Public Administration Review** 63 (5): 586-606.
- Belgisch Staatsblad (1999) **“25 mei 1999 - Wet betreffende de Belgische internationale samenwerking”**, FOD Justitie, Brussels.
- Belgisch Staatsblad (2000) **“27 april 2000 – Koninklijk besluit tot vastlegging van de internationale partnerorganisaties van de multilaterale samenwerking”**, FOD Justitie, Brussels.
- Bernstein, D.J. (1999) “Comments on Perrin’s “Effective Use and Misuse of Performance Measurement””, **American Journal of Evaluation** 20: 85-93.
- Bevan, G. and Hood, C. (2004) “Targets, Inspection and Transparency”, **British Medical Journal** 328 (7440): 598.
- Bevan G. and Hood, C. (2006) “Have targets improved performance in the English NHS?”, **British Medical Journal** 332 (7538): 419-22.
- Bouckaert G. and Halligan J. (2006) “Performance and Performance Management” in: Peters B. G., Pierre, J. (Eds) **Handbook of Public Policy**, London, Sage.
- Bouckaert, G. and Halligan, J. (2008) **Managing Performance: International Comparisons**, London, Routledge.
- Bouckaert, G. and Peters, G. (2002) “Performance Measurement and Management: The Achilles’ Heel in Administrative Modernization”, **Public Performance & Management Review** 25 (4): 359-362.
- Burall, S., Mease, K., Mall, P. and Datta, A. (2007) **“Assessing Key Stakeholder Perceptions of the Effectiveness of Multilateral Organizations”**, Overseas Development Institute: London.
- CIDA (2006a) **“Bilateral Methodologies for Assessing Multilateral Performance: Survey Findings”**, CIDA, Vancouver.
- CIDA (2006b) **“Peer Review of Evaluation Function at United Nations Children’s Fund (UNICEF)”**, CIDA, Québec.
- CIDA (2009a) **“Review of the effectiveness of CIDA’s Multilateral Delivery Channel”**, CIDA, Gatineau.
- CIDA (2009b) **“More timely, credible and cost-effective performance information on Multilateral Partners: a proposed approach”**, paper presented at the 9th meeting of the DAC network on Development Evaluation 15-16 June 2009, OECD, Paris.
- Cooksy, L. and Caracelli, V. (2005) “Quality, Context and use: issues in achieving the goals of metaevaluation”, **American Journal of evaluation** 26: 31-42.
- DAC/UNEG Joint Task force (2007) **Framework for Professional Peer Reviews**, unpublished paper, [www.uneval.org/documentdownload?doc\\_id=103&file](http://www.uneval.org/documentdownload?doc_id=103&file) (Last consulted on July 21, 2009)

- DANIDA (2008) **“Assessing Multilateral Organization Effectiveness”**, Overseas Development Institute, London.
- DANIDA Evaluation Department (2006) **“Peer assessment of evaluation in multilateral organizations: United Nations Development Programme”**, Ministry of Foreign Affairs Denmark, Copenhagen.
- De Koster, K. (2009) **“Assessing Multilaterals’ effectiveness: Which way forward for Belgium?”** Master dissertation, IOB, Antwerp.
- De Lancer Julnes, P. and Holzer, M. (2001) “Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study of Factors Affecting Adoption and Implementation”, *Public Administration Review* 61 (6): 693-708.
- DFID (2006) **“White Paper on International Development: Making Governance work for the poor”**, DFID, London.
- DFID (2009a) **“White Paper on International Development: Building our Common Future”**, DFID, London.
- DFID (2009b) **“Annual Report and Resource Accounts 2008-2009: volume II”**, DFID, London.
- DFID (2009c) **“World Health Organization UK Institutional Strategy (2008-2013)”**, DFID, London.
- DFID (2009d) **“UNDP/Denmark/UK Institutional Strategy 2008-11”**, DFID, London.
- DGDC (2008) **“Algemene beleidsnota van de minister van ontwikkelingssamenwerking”**, Belgische Kamer van Volksvertegenwoordigers, Brussel.
- DGDC (2009) **“Strategienota multilaterale samenwerking”**, Draft version, DGOS Brussels.
- Feller, I. (2002) “Performance Measurement Redux”, *American Journal of Evaluation* 23: 435-452.
- Fryer K., Antony J. and Ogden S. (2009) “Performance management in the public sector”, *International Journal of Public Sector Management* 22 (6): 478-498.
- Gabris, G. (1986) “Recognizing Management Techniques Dysfunctions: How Management Tools Often Create More Problems than They Solve”, *Public Productivity Review* 10 (2): 3-19.
- GAO (2005) **“Performance Measurement and Evaluation: Definitions and Relationships”**, GAO-05-739SP, GAO Report, Washington.
- Gao, J. (2009) “Governing by Goals and Numbers: A Case Study in the Use of Performance Measurement to Build State Capacity in China”, *Public Administration and Development* 29 (1): 21-31.
- Greiner, J. (1996), “Positioning performance measurement for the twenty-first century”, in: Halachmi, A. and Bouckaert, G. (eds), **Organizational Performance and Measurement in the Public Sector: Toward Service, Effort and Accomplishment Reporting**, Westport, Quorum Books.
- Gupta, P., Dirsmith, T. and Fogarty, T. (1994) “Coordination and Control in a Government Agency: Institutional Theory Perspectives on GAO audits”, *Administrative Science Quarterly* 39: 264-284.
- Halachmi, A. (2005) “Performance Measurement: test the water before you dive in”, *International Review of Administrative Sciences* 71 (2): 255-266.

- Hatry, H. (1999) **Performance measurement: getting results**, Washington, Urban Institute Press.
- Hofstede, G. (2001) **Culture's consequences: Comparing values, behaviours, institutions and organizations across nations**, 2nd ed., Thousand Oaks, CA, Sage.
- IFAD (2007) **"Report on IFAD's Development Effectiveness"**, IFAD, Rome.
- IFAD (2008) **"Report on IFAD's Development Effectiveness"**, IFAD, Rome.
- Kamensky, J. (1993) "Program Performance Measures: Designing a system to Manage for Results", **Public Productivity & Management Review** 16 (4):395-402.
- Kaplan, R. and Norton, D. (1992) "The balanced scorecard - Measures that drive performance", **Harvard Business Review** 70 (1): 71-79.
- Lijphart, A. (1999) **Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries**, New Haven, Yale University Press.
- Lindgren, L. (2001) "The Non-profit Sector Meets the Performance-management Movement: A Programme-theory Approach", **Evaluation** 7 (3): 285-303.
- Mayne, J. (1999) **"Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly"**, Discussion Paper, Office of the Auditor General of Canada.
- Mayne, J. (2007) **"Challenges and Lessons in Implementing Results-Based Management"**, **Evaluation** 13 (1): 87-109.
- MFA (2009a) **"Samen Werken aan mondiale uitdagingen: Nederland en multilaterale ontwikkelingsamenwerking"**, Ministry of Foreign Affairs, Den Hague.
- MFA (2009b) **"Resultaten in ontwikkeling 2007-2008"**, Ministry of Foreign Affairs, Den Hague.
- MOPAN (2005a) **"The MOPAN Survey 2004: synthesis report"**, MOPAN, Zürich.
- MOPAN (2005b) **"The MOPAN Survey 2005: Perceptions of multilateral partnerships at country level"**, MOPAN, Zürich.
- MOPAN (2006) **"The annual MOPAN Survey 2006: Donor perceptions of multilateral partnerships at country level"**, MOPAN, Zürich.
- MOPAN (2007) **"The annual MOPAN Survey 2007: Donor perceptions of multilateral partnerships behaviour at country level"**, MOPAN, Zürich.
- MOPAN (2008) **"The annual MOPAN Survey 2008: Donor perceptions of multilateral partnerships behaviour at country level: synthesis report"**, MOPAN, Zürich.
- MOPAN (2010a) **"Common Approach: United Nations Development Program (UNDP) 2009"**, MOPAN, n.p.
- MOPAN (2010b) **"Common Approach: UNICEF 2009"**, MOPAN, n.p.
- MOPAN (2010c) **"Common Approach: African Development Bank (AfDB) 2009"**, MOPAN, n.p.
- MOPAN (2010d) **"Common Approach: World Bank 2009"**, MOPAN, n.p.

- Obser, A. (2007) **“Multilateral Organisations Performance Assessments: Opportunities and Limitations for Harmonisation among Development Agencies”**, Discussion Paper, German Development Institute, Bonn.
- ODI (2005) **“Relationships of other donor organizations with multilaterals”**, ODI, London.
- OECD/DAC (1991) **“Principles for Evaluation of Development Assistance”**, OECD, Paris.
- OECD/DAC (2000) **“Results Based Management in the development co-operation agencies: a review of experience”**, OECD, Paris.
- OECD/DAC (2005) **“Paris Declaration on Aid Effectiveness. Paris: OECD/DAC (2006a) DAC Peer Review: Netherlands”**, OECD/DAC, Paris.
- OECD/DAC (2006a) **“DAC Peer Review: Netherlands”**, OECD, Paris.
- OECD/DAC (2006b) **“DAC Peer Review: United Kingdom”**, OECD, Paris.
- OECD/DAC (2009a) **“2008 DAC report on multilateral aid”**, OECD, Paris.
- OECD/DAC (2009b) **“OECD Stat Extracts: ODA by Donor”**, OECD, Paris.  
[http://stats.oecd.org/Index.aspx?DatasetCode=ODA\\_DONOR](http://stats.oecd.org/Index.aspx?DatasetCode=ODA_DONOR)  
 (Last consulted on 30 July 2009)
- OECD/DAC (2010a) **“Proposed Approach to Strengthening Information on the Development Effectiveness of Multilateral Organizations”**. 10th meeting of the DAC Network on Development Evaluation, 10-11 January, OECD/DAC, Paris.
- OECD/DAC (2010b) **“DAC Work on Multilateral Aid. Presentation given at the 10th meeting of the DAC Network on Development Evaluation”**, 10-11 January 2010, OECD/DAC, Paris.
- Pagani, F. (2002) **“Peer Review: a tool for co-operation and change: An analysis of an OECD Working Method”**, OECD, Paris.
- Perrin, B. (1999) “Performance Measurement: Does the reality match the rhetoric? A rejoinder to Bernstein and Winston”, **American Journal of Evaluation** 20:101-111.
- Perrin, B. (2002) **Implementing the vision: addressing challenges to results-focused management and budgeting**, Paris, OECD.
- Pollitt C. and Bouckaert G. (2004) **Public Management Reform: A Comparative Analysis**, Oxford, Oxford University Press.
- Pollitt, C. (2006) “Performance Management in Practice: A Comparative Study of Executive Agencies”, **Journal of Public Administration Research and Theory** 16 (1): 25-44.
- Pollitt, C., Harrison, S., Dowswell, G., Jerak-Zuiderent, S. and Bal R. (2010) “Performance Regimes in Health Care: Institutions, Critical Junctures and the Logic of Escalation in England and the Netherlands”, **Evaluation** 16(1):13-29.
- Ramalingam, B.; Mitchell, J.; Borton, J. and Smart, K. (2009) “Counting what counts: performance and effectiveness in the humanitarian sector”, **ALNAP Review of Humanitarian Action**, ALNAP, London.
- RBMG (2007) **“Assessing Multilateral Organization Effectiveness: implementing the new MOPAN common approach: option paper”**, paper presented to the MOPAN Working Group on December 6 2007, RBMG, Ottawa.
- Saltmarshe, D., Ireland, M. and McGregor, J. (2003) “The performance framework: a systems approach to understanding performance management”, **Public administration and development** 23 (5): 445-456.



- Schwartz, R. and Mayne, J. (2005) "Assuring the quality of evaluative information: theory and practice", *Evaluation and Program planning* 28: 1-14.
- Scott, A. (2004) "Assessing multilateral effectiveness", working paper, DFID, London.
- Scott, A. (2005) "DFID's Assessment of multilateral organisational effectiveness: an overview of the results", DFID, London.
- Selbervik, H. and Jerve, A.M. (2003) "MOPAN: Report from the 2003 pilot exercise: final report", Chr. Michelsen Institute, Bergen.
- SIDA (2007) "Peer Review of the Evaluation Function at the World Food Programme (WFP)", Edita, Stockholm.
- SDC (2009) "Peer Review of the Evaluation Function of the Office of Internal Oversight Services of the United Nations (OIOS)", Nicolussi, Zürich.
- Special Evaluation Unit (2009) "Peer review: The evaluation function of the global environment Facility (GEF)", Foreign affairs, foreign trade and development cooperation, Brussels.
- Stufflebeam, D. (2001) "The metaevaluation imperative", *American Journal of Evaluation* 22: 183-209.
- Taylor, J. (2009) "Strengthening the link between performance measurement and decision making", *Public Administration* 87 (4): 853-871.
- Thomas, P. (2007) "Why is Performance-Based Accountability So Popular in Theory and Difficult in Practice?", Paper presented to the World Summit on Public Governance: Improving the Performance of the Public Sector, May 1-3, Taipei City.
- Tweede Kamer der Staten-Generaal (1999) "Rapport: Bijdragen aan Internationale Organisaties", Sdu Uitgevers, 's-Gravenhage.
- Tweede Kamer der Staten-Generaal (2009) "Beleidsbrief Ontwikkelingssamenwerking", Sdu Uitgevers, 's-Gravenhage.
- UNEG (2005) "Standards for Evaluation in the UN System", [http://www.uneval.org/papersandpubs/documentdetail.jsp?doc\\_id=22](http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=22) (Last consulted July 19, 2009)
- Van Thiel, S. and Leeuw, F.F. (2002) "The performance paradox in the public sector", *Public Performance & Management Review* 25 (3): 267-281.
- Varone, F.; Jacob, S. and De Winter, L. (2005) "Polity, Politics and Policy Evaluation in Belgium", *Evaluation* 11:253-273.
- Weaver, R.K. and Rockman, B.A. (eds). (1993) *Do institutions matter? Government capabilities in the United States and abroad*, Washington DC, Brookings Institution.
- White, H. (2002) "The road to nowhere? Results based management in international cooperation: current issues, food for thought, background information", Euforic ([http://www.euforic.org/detail\\_page.phtml?lang=en&page=resource\\_briefing\\_evaluat\\_edit3](http://www.euforic.org/detail_page.phtml?lang=en&page=resource_briefing_evaluat_edit3), accessed November 2009).
- White, H. (2005) "Challenges in Evaluating Development Effectiveness", in: Pitman, G.K., Feinstein, O.N. and Ingram, G.K. (eds), *Evaluating Development Effectiveness, World Bank Series on Evaluation and Development* Vol. 7. New Brunswick, New Jersey: Transaction Publishers.

Wholey, J. and Hatry, H. (1992) "The case for performance monitoring", *Public Administration Review* 50 (6): 604-640.

Yang, K. (2008) "Examining Perceived Honest Performance Reporting by Public Organizations: Bureaucratic Politics and Organizational Practice", *Journal of Public Administration Research and Theory* 19 (1): 81-105.





University  
of Antwerp



INSTITUTE OF DEVELOPMENT  
POLICY AND MANAGEMENT