

BSG Working Paper Series

*Providing access to the latest
policy-relevant research*



Assessing the effectiveness of multilateral organizations

BSG-WP-2016/013

April 2016

**Vinicius Lindoso, Alumnus, Blavatnik School of
Government, University of Oxford**

**Nina Hall, Research Associate, Global Economic
Governance Programme, Blavatnik School of
Government, University of Oxford**

Assessing the Effectiveness of Multilateral Organizations

Authors: Vinicius Lindoso (Alumnus of Blavatnik School of Government, University of Oxford)
& Nina Hall (Hertie School of Governance, Berlin)

April 2016

hall@hertie-school.org

Hertie School of Governance
Friedrichstrasse 180
10117 Berlin

Word count: 11,888 (including abstract, footnotes, appendix)

Abstract

States and multilateral organizations are increasingly assessing the effectiveness of multilateral development organizations (MOs). Assessments vary widely in their definition of effectiveness and their overall purpose. These assessments may encourage organization accountability, foster learning, and inform donor strategies and/or financing. In parallel a large body of international relations (IR) scholarship also examines multilateral effectiveness. Scholars have noted the importance of: commanding stakeholder support, effective delegation and legitimate governance, internal management of the bureaucracy and delivering results on the ground. However, this IR scholarship has not contributed to, nor been cited, in the on-going policy debates. This working paper seeks to bridge the gap between the IR and policy literature on multilateral effectiveness. It draws on a review of the primary and secondary literature, and discussions with expert practitioners and academics on the topic. The paper makes a significant contribution to the field by identifying: how policy and scholars have assessed effectiveness; the challenges of assessing effectiveness; and outlining future research avenues for scholars to pursue.

Key Words: Multilateral organizations, effectiveness, performance, assessments and evaluations.

Wider audience relevance:

Donors, recipient countries and citizens look to international institutions to provide critical public goods. We need these institutions to be effective and deliver everything from refugee protection to children's rights, from humanitarian relief to labour standards. It is thus crucial to develop methodologically robust indicators and methods of effectiveness. There is a growing field of international assessments which: inform donor strategies, share good practices between multilaterals, provide systematic information on development results, and improve performance of multilaterals. However these assessments face challenges such as: robustly and accurately measuring effectiveness, combining donors' and recipients' perspectives; and balancing universal assessments with the need to take organizational differences into account. This working paper encourages International Relations' scholars and practitioners to take up these challenges. These communities could collaborate through workshops, conferences and/or an online network of multilateral assessors, to pool resources and conduct common performance assessments of multilateral organisations. It is important to assess and enable multilaterals to operate effectively given the scale and scope of global challenges.

Target Audiences:

International organizations, donor states, aid agencies, development NGOs, performance analysts.

Biographies and Connection to BSG:

Nina Hall is a research associate with the Global Economic Governance Programme and a post-doctoral fellow at the Hertie School of Governance. She worked with Prof. Ngaire Woods and several BSG students to research and write the World Economic Report "Effective Leadership in International Organizations" (2015) and published a book on "Displacement, Development and Climate Change, International Organizations moving beyond their Mandates" (Routledge, 2016).

Vinicius Lindoso is a Communications Consultant with UNESCO. A BSG alumnus, he completed his Master of Public Policy in 2015 with a focus on multilateral effectiveness.

Acknowledgements

The authors would like to acknowledge the generous financial support of the Hertie School of Governance, which hosted a workshop on Multilateral Effectiveness in September 2015. They are also grateful to staff at the Multilateral Organizational Performance Assessment Network (MOPAN), at the Organization for Economic Cooperation and Development (OECD) for their feedback and input into this research. The authors would like to thank Ngaire Woods for her intellectual guidance and Lisa Schmid for research assistance. Thanks also to the workshop participants for their comments which greatly enriched this paper. The workshop participants included: Björn Gillsater, Rachel Arrundale, Julia Lehmann, Angel Saz-Carranza, Ngaire Woods, May Miller-Dawkins, Julian Gayfer, Julia Betts, Dorte Kabell, Johanna Mair, Per Trulsson, Katie Vanhala, Lucien Aegerter, Paul Isenman, and Miles Kellerman. The paper has not been published elsewhere and any errors are those of the authors.

Introduction

In recent years there has been a growth of comprehensive assessments, evaluations and reviews of multilateral organizations' (MOs') effectiveness in the development sector. A range of government agencies such as the former Australian Agency for International Development (AusAID), the Swedish International Development Cooperation (SIDA) and the United Kingdom's (UK's) Department for International Development (DFID) have sought to examine the performance of international organizations they fund. Governments and multilaterals are under pressure to show that development assistance produces tangible results. Donors want to see effective, accountable MOs. After all, these organizations provide critical global public goods and address urgent international crises. In fact, donor states have pooled resources to establish and conduct common assessments through the Multilateral Organisations Performance Assessment Network (MOPAN).¹ In addition, many international organizations – from the World Bank to the United Nations – are also commissioning independent assessments of their own organizational performance to illustrate how they deliver results.

As a result, we now have a multitude of assessments, reviews and evaluations carried out at the multilateral, bilateral, and organization-specific levels. These are not just technocratic exercises. They have political significance and power: they may instigate reforms, or alternatively a poor performance review can result in a decrease in the financing and legitimacy of a MO. However, these assessments do not have a common definition of effectiveness and vary in their emphasis on outputs, outcomes and alignment with donor priorities. This diversity of approaches to defining effectiveness begs the question: how should we assess the effectiveness of MOs in a more unified way? This is an important question for both academics and policy-makers working on multilateral development organizations, and is the focus of this working paper.

It is also imperative to connect the international relations scholarship on MOs with these policy debates. There is a large body of scholarship in international relations which explores why states use multilateral institutions, and what determines their internal effectiveness and their ability to deliver results. However, most international relations scholars have not engaged with the policy literature on multilateral effectiveness. Furthermore, the academic scholarship is strikingly absent from studies of multilateral effectiveness conducted by governments. None of the recent studies – conducted by MOPAN, Australia's Agency for International Development (AusAID) and Sweden's International Development Agency (SIDA) – cite any international relations literature.² We suggest in this paper that international relations scholarship provides insights into how we can define and understand organizational effectiveness and performance.³

International relations scholars have identified various dimensions of effectiveness. One strand of scholarship has focused on international organizations' ability to *command states' support*; others have looked at *effective delegation and legitimate governance*; another has examined their *internal bureaucratic processes* and a fourth has focused on their *results and impact*. Scholars in this fourth stream continue to debate whether we should focus on outputs or outcomes and impacts.⁴ More recently scholars have called for the integration of public administration theories, which focus on national bureaucracies, to assist our study of international bureaucracies.⁵ Why is this work not considered useful or relevant to multilateral assessment practitioners? Conversely, why aren't IR scholars taking into account the growing amount of policy, and academic literature, which evaluates multilaterals' performance and effectiveness?

¹ See MOPAN. (2012). 2012 Annual Report. Retrieved 1/23/2016 from http://www.mopanonline.org/publications/MOPAN_Annual_Report_2012.pdf.

² The one exception is a background paper to MOPAN (2013) "Common Approach Methodology" which uses a textbook written by public administration specialists, the "Handbook of Practical Program Evaluation" (Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2010). *Handbook of practical program evaluation* (Vol. 19). John Wiley & Sons.).

³ Notably there is a field of academics and practitioners who specialize in evaluation but this does not include international relations scholars. For instance see the work of: Prof. Stewart Donaldson at Claremont University, Director of the Claremont Evaluation Center and is also the President of the American Evaluation Association.

⁴ Gutner, T., & Thompson, A. (2010). The politics of IO performance: A framework. *The Review of International Organizations*, 5(3), 227-248.

⁵ Solana, J. & Saz-Carranza, A. (2015). Treating Global Governance Seriously. *Public Administration Review*, 75(5), 776-778; Woods, N. et al. (2015). *Effective Leadership in International Organizations*. Switzerland: World Economic Forum.

This paper aims to bridge the gap between international relations scholarship and policy debates on measuring, studying and assessing multilateral effectiveness.⁶ It does so by introducing international relations scholars to the challenges faced by the evaluators and assessors of multilaterals, and introducing practitioners to international relations scholarship. International relations scholars would greatly benefit from engagement with existing policy debates on the effectiveness of international organizations, in order to avoid developing research agendas that do not relate to existing policy demands and developments. Meanwhile, policy evaluators risk pursuing evaluations and assessments that are not informed by research undertaken in international relations.

The paper draws on a meta-literature review of IR scholarship and policy literature as well as extensive discussions between leading assessment practitioners and academics. A workshop between these experts was held at the Hertie School of Governance on September 17th, 2015. The working paper builds on this workshop and responds to the following questions:

1. Why do we need to measure effectiveness?
2. How are we measuring effectiveness?
3. What are current challenges of measuring effectiveness?
4. What should be the focus for future research?

Why measure effectiveness?

The way multilateral effectiveness is defined, measured and evaluated is crucially important for global governance. Donor states measure effectiveness of multilaterals to assess where to concentrate their resources, and how to strategically engage and cooperate with these organizations. Yet conceptions of effectiveness vary greatly between donors, as will be discussed in the subsequent section.

Donor states have also combined efforts and established common assessment frameworks. These efforts have led to the establishment of MOPAN in 2002, a network of 18 states (as of December 2015) with a common interest in assessing the effectiveness of MOs. MOPAN has carried out multiple assessments covering dozens of multilaterals and exists alongside assessments performed by individual donors.⁷ There are also a number of other independent assessments commissioned by the MOs themselves, many of which have set up their own independent evaluation offices such as the World Bank. Developing robust and common methods of assessing MOs is crucial for organizational learning and improvement, and of course to ensure that global institutions deliver results to recipients.

Changes in the international development landscape also make it essential for us to review definitions of performance and effectiveness.⁸ These changes include, but are not limited to:

1. The creation of the 2030 Agenda for Sustainable Development and the Sustainable Development Goals (SDGs), a set of global priorities which follow on from the Millennium Development Goals (MDGs), and guide development strategies of UN Member States with the active support of the UN System and

⁶ This working paper does not examine the public administration literature, although there is ample scope for IR scholars to learn from this. See *ibid*.

⁷ As stipulated in MOPAN's 2013 Mission Statement, the aim of these assessments goes beyond purely supporting members in assessing the effectiveness of multilaterals to include an effort to "strengthen organizations' contribution to overall development and humanitarian results" and to "contribute to organizational learning within and among MO, their direct clients/partners and beneficiaries." See MOPAN. (2014). 2014 Annual Report. Retrieved 1/23/2016 from <http://www.mopanonline.org/publications/MOPAN%20Annual%20Report%202014.pdf>.

⁸ (1)-(4): Harman, S., & Williams, D. (2014). International development in transition. *International Affairs*, 90(4), 925-941; (5): Bezanson, K. A., & Isenman, P. (2012). Governance of new global partnerships: Challenges weaknesses and lessons. *Center for Global Development*. Retrieved 1/31/2016 from <http://www.cgdev.org/publication/governance-new-global-partnerships-challenges-weaknesses-and-lessons>.

other MOs. Given their introduction in September 2015, it is imperative to ask: how can and should we assess multilateral performances towards achievement of the SDGs?

2. The emergence of global multi-stakeholder partnerships such as Gavi, the Vaccine Alliance; the Global Fund for the fight against AIDS, Tuberculosis and Malaria (GFATM) and the Global Water Partnership, (GWP). Are these global 'vertical' funds which focus on one sector, more effective than the larger, traditional multilateral bureaucracies which work across many sectors?
3. The combination of greater recipient country autonomy and a greater focus on state-centric development means that MOs must increase efforts to integrate recipient governments into the planning and implementation of interventions. Multilaterals should operate through national structures, and harmonize their activities with both the government and other actors operating on the field, if they wish to continue pursuing effective partnerships. How can we assess multilateral effectiveness from a recipient perspective?
4. There is also greater choice for aid recipients with the emergence of non-traditional donors, such as the BRICs (Brazil, Russia, India and China) and the proliferation of foundations and other non-state actors (the Gates Foundation) playing a more significant role in global development. This begs the question: how do existing MOs operate in a world in which there are many new, alternative sources of development finance?

In sum, these changes require multilaterals to reflect on their operating models, their contribution to global governance and the provision of global public goods. Understanding their performance and effectiveness is crucial in this context.

What are existing measures of effectiveness?

Academics and policy-makers have examined organizational effectiveness in a number of dimensions, as will be discussed here. The first section examines assessment practitioners' conceptions of multilateral effectiveness and the second outlines the academic perspectives.

Policy Assessments of Effectiveness

Multilateral effectiveness assessments are undertaken for a variety of reasons, chief among which are supporting organizational learning and enforcing political accountability.⁹ The purpose of assessments tend to correspond to which set of actors are carrying them out, whether a single donor conducting assessment at the bilateral level, or a group of donors and/or international organizations coordinating assessments following a multilateral approach.

There has been a real growth in assessments of effectiveness in the past decade: from peer reviews of the evaluation functions of international organizations to the multilateral development banks' Common Performance Assessment System (COMPAS), to broader institutional assessments such as MOPAN. They reflect a shift in the development sector towards results-based and performance management frameworks. The assessments we see today are an attempt to apply principles of traditional policy evaluation – such as theory of change – to reforming comprehensive management processes within an organization. As Obser notes, “most OECD governments see evaluation [today] as part of the overall performance management framework,” key to generating reliable performance information in an era of high accountability.¹⁰ Obser explains that evaluation at the policy level “was well established in many governments before the introduction of performance

⁹ Proudlock, K., Ramalingam, B., & Sandison, P. (2009). *Improving humanitarian impact assessment: Bridging theory and practice*. 8th Review of Humanitarian Action: Performance, Impact and Innovation: 3.

¹⁰ Obser, A. (2007). Multilateral organisations performance assessment: Opportunities and limitations for harmonisation among development agencies. *German Development Institute*: 3.

management and the new approaches did not necessarily incorporate evaluation techniques,” due to doubts over whether evaluations were too costly for their limited value added potential.¹¹

There is great variation in the measures of effectiveness. Some donors look at the role played by an institution within global governance and others focus on an organization’s internal effectiveness and relevance to national goals. Some donors compare and rank institutions and use this as a basis to allocate resources while others do not. Notably, bilateral assessments and evaluations tend to focus, at least in part, on adherence of multilaterals to donors’ priorities. Among other indicators, these assessments establish parameters that can be used to check whether multilaterals align with donor governments’ goals as well as to hold the donor government accountable to its domestic audience in regards to multilateral funding decisions. The UK’s Multilateral Aid Review (MAR), for example, has two core criteria: contribution to UK Development objectives, and organizational strength. It then assigns each of the multilaterals assessed a ‘value for money’ score that is directly related to the levels of UK financial contributions. On the other hand, multilateral reports and assessments such as COMPAS or MOPAN look at organizational and development effectiveness more broadly, not based on any single national preferences.

The development effectiveness assessment field is a mix of assessments of organizational performance, assessments of aid transparency, and multiple mutual accountability mechanisms¹². Rather than engage in imperfect cross-assessment comparisons, we list below the main bilateral and multilateral assessments that have emerged in the development effectiveness arena, with short descriptions of their purpose and indicator framework based on the assessment reports themselves, dedicated institutional websites, and on a study of multilateral development cooperation commissioned by AusAID.¹³ For a more detailed description of each assessment see the appendix.

TABLE 1: Summary of Multilateral Effectiveness Assessments

ASSESSMENT	LEVEL	ASSESSMENT CARRIED OUT BY	PURPOSE
MOPAN Organizational Effectiveness Assessments	Multilateral	19 donor countries	<ul style="list-style-type: none"> – Make available accurate, comprehensive, qualitative and quantitative data on development effectiveness of MOs – Promote dialogue between donors, multilaterals and other development partners
Quality of ODA (QuODA)	Multilateral	OECD / DAC	<ul style="list-style-type: none"> – Allow donor agencies (both bilateral and multilateral) to review where they stand against their peers on critical dimensions of development effectiveness – Serve as tools for starting dialogue and promote learning
Peer Review of Evaluation Functions	Multilateral	DAC Evaluation Network / UN Evaluation Group	<ul style="list-style-type: none"> – Share good practices, experience and mutual learning, – Build greater internal capacity and external confidence in UN evaluation systems – Reduce demands for external assessments of

¹¹ Ibid.

¹² Mutual accountability (MA) refers to mechanisms through which development partners (donor countries, international organizations and recipient countries hold each other accountable for their performance under any development project or programme.

¹³ Dinham, M. (2011). Study of AusAID’s Approach to Assessing Multilateral Effectiveness: A study commissioned by the Independent Review of Aid Effectiveness to assist in their overall analysis of the effectiveness and efficiency of the Australian aid program. *ALNAP*. Retrieved 1/31/2016 from <http://www.alnap.org/resource/9851>.

organizations' performance and effectiveness

- Facilitate cross-organization dialogue and stimulate organizations to change

Development Effectiveness Review	Multilateral	OECD / DAC Network on Development Evaluation	<ul style="list-style-type: none"> – Provide systematic information on the development results of multilaterals – Complement DAC/UNEG Peer Reviews and MOPAN periodic assessments – Reduce bilateral demands for ad hoc information
Common Performance Assessment System (COMPAS)	Multilateral	Group on Managing for Development Results (WG-MfDR)	<ul style="list-style-type: none"> – Improve public information on performance of Multilateral Development Banks (MDBs) – Bolster constructive dialogue within and among MDBs and partners to sharpen their focus on MfDR
Publish what you Fund	Multilateral	NGO	<ul style="list-style-type: none"> – Stimulate aid transparency among reporting agencies – Improve public accountability and awareness around development aid
Monitoring reports for the Global Partnership	Multilateral	Global Partnership for Effective Development Cooperation	<ul style="list-style-type: none"> – Track behavior change in development cooperation – Support accountability based on Busan standards – Highlight opportunities and challenges for further progress – Stimulate inclusive dialogue on improving the effectiveness of development cooperation – Promote agreement on actions to use the Busan principles successfully
UNDCF's mutual accountability survey	Multilateral	UN Development Cooperation Forum	<ul style="list-style-type: none"> – Promote candid and inclusive exchange among aid donors and recipients – Review progress made in implementing mutual accountability
Australian Multilateral Assessment (AMA)	Bilateral	Australia	<ul style="list-style-type: none"> – Provide a firm base of information about the effectiveness and relevance of multilateral organizations, from the perspective of the Australian aid program – Inform Australian funding decisions – Design rating system for annual reviews
Multilateral Aid Review (MAR)	Bilateral	United Kingdom	<ul style="list-style-type: none"> – Assess the value for money offered by all the multilateral organizations that receive significant core aid funding – Support domestic accountability – Promote dialogue with multilaterals and develop a focused reform agenda – Monitor progress of reforms
Multilateral Review	Bilateral	Denmark	<ul style="list-style-type: none"> – Support multilaterals aligned with Danish development cooperation priorities – Support future multilateral engagement

			policy, including funding
Multilateral Review	Bilateral	Sweden	<ul style="list-style-type: none"> – Guide decisions on multilateral funding – Provide basis for organizational strategy and informed policy dialogue with MOs
Multilateral Review	Bilateral	Norway	<ul style="list-style-type: none"> – Inform funding decisions and accountability – Promote dialogue with multilaterals
Multilateral Review	Bilateral	Germany	<ul style="list-style-type: none"> – Inform strategic engagement and cooperation with multilaterals – Assess organizations direction of travel: history of and potential for implementing reforms to achieve more effective development cooperation

Assessing Effectiveness

Different donor rationales for assessing multilaterals and the specificities of multilaterals' mandates and operating contexts lead to different definitions and interpretations of effectiveness. For example, Sweden's Multilateral Review defines effectiveness as both alignment with Swedish goals and effective internal organizational processes. On the other hand, the Danish and OECD Development assessments relate effectiveness to a multilaterals' ability to produce needed public goods such as: public health, knowledge production, financial security, and environmental protection. We can see four main purposes of effectiveness assessments: transparency, development results, alignment with donor goals and priorities, and finally internal organizational effectiveness.

Effectiveness vs. Performance

To further complicate the definitional debate around multilateral effectiveness, the term is often used interchangeably with *performance*. The Multilateral Organisations Performance Assessment Network (MOPAN), for example, focuses most of its assessments on "five dimensions of organizational effectiveness."¹⁴ MOPAN assesses organizational effectiveness by looking at four dimensions (strategic management, relationship management, operational management, performance management and results) and in addition assesses development effectiveness. Each of these dimensions of effectiveness is measured via specific Key Performance Indicators (KPIs). In 2012, a specific assessment of multilateral performance in select countries (measured using perception surveys of multilateral country staff and stakeholders) was added to MOPAN assessments.

These indicators formed MOPAN's "Common Approach" method, established in 2009 and discontinued recently in favour of a new MOPAN 3.0 methodology. The Common Approach indicators were originally heavily slanted toward measuring performance in the sense of whether multilaterals have the necessary elements (behaviours, systems and processes) to achieve results, without assessing whether the results were actually being achieved. This changed in 2012, with the inclusion of an expanded methodological framework that included "development results" indicators to assess multilaterals' progress in achieving their stated goals. The focus on development results – strengthened further in the MOPAN 3.0 methodology – shows a trend within MOPAN of increasing focus on development results (*effectiveness*) as well as *performance mechanisms*.

¹⁴ MOPAN. (2015). Synthesis Report: United Nations High Commissioner for Refugees (UNHCR). Retrieved 1/23/2016 from <http://www.mopanonline.org/publications/MOPAN2014UNHCR-Executivesummary.htm>.

Within MOPAN as well as other assessments, focus on development results follows a general trend in general public policy approaches that see multilateral assessments as *public accountability*, and seek to give the general public a picture of multilaterals that clearly link financial allocations to results achieved.¹⁵

International Relations' and Multilateral Effectiveness

Meanwhile, international relations scholars have identified at least five different conceptions of multilateral effectiveness. Firstly, scholars have looked at whether MOs *command states' support*. Much of traditional international relations scholarship has examined to what extent and why states act through MOs, and support the organization's mandate and budget. Secondly, principal agent scholars have looked at *effective delegation* while others have focused on the *legitimacy* of multilaterals. Sociological institutionalists and constructivists have looked at the *internal bureaucratic* dimension. They have examined whether or not the staffing, organization, and operations of an agency lead to 'pathologies' or results that contradict with an organization's mandate. Finally, some international relations scholars have examined the *results* of international organizations, and if their outputs actually have an impact. Here we will briefly examine these five dimensions and then focus most substantively on how to measure results.

Commanding Stakeholder Support

A powerful body of international relations scholarship identifies international organizations as arenas of inter-state competition, control and strategy and probes the conditions under which a MO is likely to pursue the goals either of a powerful state, or of its membership as a whole. Scholars have probed and debated to what degree the existence and design of an institution can alter the willingness of governments to cooperate and to alter their behaviour as a result of cooperation. Ultimately, however, effectiveness is conceived as the ability of the organization to deliver political outcomes which command the support of member states. The literature usefully asks: how extensively are states choosing to use the institution?¹⁶ How do states ensure that an organization delivers on its delegated tasks?

By way of example, Jupille, Mattli and Snidal propose that states either *Use, Select, Change or Create* international organizations, depending on the nature of the cooperation problem and the institutional status-quo.¹⁷ They argue that states, are boundedly rational, 'satisficers,' rather than 'optimisers,' when selecting which institutional strategy and prefer to 'stick with the devil they know.'¹⁸ A state's choice to use a particular international organization will depend on organizational fit (does the organization have the 'right' properties) and role (is the organization 'entrepreneurial in raising or seeking to host the issue'). For instance: Henry Dunant, the founder of the International Committee of the Red Cross (ICRC) raised the idea of treating battlefield medical personnel and the wounded as non-combatants on the international agenda. Dunant also lobbied successfully for the ICRC to be the institution to address this issue, thus determining what institution states 'used' to solve a new policy problem. This work highlights the importance of commanding stakeholder support for multilaterals.

This literature could also examine if, and how, the perceived effectiveness of the institution translates into the selection, use or change of an organization. The DFID Multilateral Aid Review for instance suggests that states are less likely to use institutions that do not perform well in multilateral assessments. International relations scholars could also consider how to measure variations in commanding stakeholder support and how this translates into multilateral effectiveness. One tentative hypothesis is that institutions which command greater support from their states will perform better on their behalf.

¹⁵ Obser, A. (2007). Multilateral organisations performance assessment: Opportunities and limitations for harmonisation among development agencies. *German Development Institute*: 6.

¹⁶ Stone, R. W. (2011). *Controlling institutions: International organizations and the global economy*. Cambridge University Press; Jupille, J., Mattli, W., & Snidal, D. (2013). *Institutional choice and global commerce*. Cambridge University Press.

¹⁷ Jupille, J., Mattli, W., & Snidal, D. (2013). *Institutional choice and global commerce*. Cambridge University Press: 50.

¹⁸ *Ibid*: 7.

Effective Delegation

Other strains of IR scholarship have examined how effectively and legitimately multilaterals are governed by their member states. Principal-agent scholars suggest that states (principals) delegate autonomy to international organizations (agents) to perform specialised tasks on their behalf.¹⁹ Scholars have explored the conditions under which multilaterals fulfil these tasks and identified why they may not. They propose that multilaterals may deviate from their delegated tasks (and engage in so-called ‘agency-slack’) as agents try to maximise their own autonomy and preferences, which are assumed to be distinct from those of the principal. A multilateral’s ability to deviate from delegated tasks depends on factors such as: heterogeneity between states’ preferences, voting structures, asymmetry of information, proximity of preferences (or how strongly states feel about a particular issue) and/or whether states delegate a high degree of discretion.²⁰ Principals will also monitor multilaterals to ensure they stick to their tasks – by controlling the budget, senior staff appointments and/or vetoing proposed plans or programmes. Problems can arise if principals delegate tasks that conflict with existing mandates – so-called “antinomic delegation.”²¹

The principal-agent literature is primarily concerned with the relationship between states and multilaterals. It assesses their performance based on whether they follow states’ instructions and deliver on mandated tasks. These insights may be useful for the policy evaluation literature which rarely explores the conditions under which tasks are delegated to agents. Yet the degree of consensus between board members (heterogeneity) and/or how strongly states monitor a multilateral may influence an organization’s actions. The evaluation literature could also examine if and how multilateral Executive Boards (comprised of member states) function. In particular: are the Boards delegating mandates which there is strong support for and are they giving states sufficient financing to perform these tasks?²²

Legitimate Governance

In parallel to the principal-agent literature, IR scholars have also explored the governance of multilaterals by states. Scholars have identified core principles of good governance such as: *participation* (do all affected stakeholders have access to decision-making); *accountability* (are organizations run by the most powerful for the most powerful? How is the institution run to reflect all member states’ interests?); and *fairness* (both procedural fairness in process of decision-making, and substantive fairness in outcomes).²³ This scholarship points to broader normative questions of legitimacy alongside effectiveness.²⁴ For instance, who, in addition to member states, should have a voice in decision-making in multilaterals (NGOs, civil society and/or affected populations)? How should boards distribute voting power to ensure newly emerging powers have a proportionate influence alongside traditional powers? These are critical questions for the on-going relevance, effectiveness and legitimacy of multilateral development and humanitarian organizations.

Effectiveness of internal management and bureaucratic processes

A third body of IR scholarship focuses on the bureaucratic processes within international organizations. Sociological institutionalists have explored the internal, bureaucratic reasons for organizational ineffectiveness. Scholars in this vein have argued that MOs can be dysfunctional due to their internal structure, staffing and organizational culture, not because they are quietly perverted by their most powerful member state.²⁵ Scholars have identified organizational pathologies which lead them to “violate the self-understood core goals of the

¹⁹ Hawkins, D. G., Lake, D. A., Nielson, D. L., & Tierney, M. J. (2006). Delegation under anarchy: states, international organizations and principal-agent theory. *Delegation and Agency in International Organizations*. Cambridge, UK.

²⁰ Nielson, D. L., & Tierney, M. J. (2003). Delegation to international organizations: Agency theory and World Bank environmental reform. *International Organization*, 57(02), 241-276.

²¹ Gutner, T. (2005). Explaining the gaps between mandate and performance: agency theory and World Bank environmental reform. *Global Environmental Politics*, 5(2), 10-37.

²² Note there is a broader literature that looks at governance.

²³ Woods, N. (1999). Good governance in international organizations. *Global Governance*, 39-61.

²⁴ Hurd, I. (2008). Myths of Membership: The Politics of Legitimation in UN Security Council Reform, *Global Governance*, 199 – 217.

²⁵ Barnett, M., & Finnemore, M. (2004). Rules for the world: International organizations in global politics. Cornell University Press.

organization.”²⁶ These include: exceptions to the rule becoming the norm; universalism (whereby multilaterals work in numerous local contexts at once and assume similarities between); or organizational insulation (there is a lack of feedback so organizations do not learn lessons from the past). In addition, there may be cultural contestation between different organizational units with different priorities or bureaucratic rationalization (rules become so entrenched and powerful that they determine the ends and the way the organization defines its goals). So “rather than designing the most appropriate and efficient rules and procedures to accomplish their missions, bureaucracies often tailor their missions to fit the existing, well-known and comfortable rulebook.”²⁷ These are important bureaucratic tendencies that policy evaluators should look out for.

Scholars have also argued that contradictory results (or ‘hypocrisy’) are a predictable feature of organizational behaviour.²⁸ This occurs because international organizations are trying to satisfy the interests of many different states and stakeholders. The next step for IR scholars is to consider how to measure variations in internal bureaucratic effectiveness and the managerial performance of international organization. To do so they could look to the public administration literature as well as the policy and evaluation literature.

Organizational Results

Scholarship also focuses on results as an indication of the effectiveness of international organizations. Here scholars have examined whether specific international organizations have delivered concrete changes on the ground. For instance: Do IMF programmes reduce government budget deficits and if so how?²⁹ Do IMF and World Bank programs affect the likelihood of major government crises, if so under what circumstances?³⁰ What is the impact of UN Security Council decisions?³¹ These studies are extremely useful for examining the impact of specific institutions (and could be informative for policy evaluations). However, IR scholars have tended to focus on just a select few institutions – the World Bank, the IMF and the UNSC – and not examined the full range of multilateral development and humanitarian organizations. Also their metrics of effectiveness are not intended to be generalizable for a broader population of organizations and thus are of limited use to those wanting to evaluate a larger population of multilaterals.

In fact IR scholars still debate how to measure and rigorously capture organizational performance and results.³² Two useful typologies from the IR literature for measuring effectiveness of MOs are Biermann and Bauer (2004) and Gutner and Thompson (2010). They both focus on a MO’s ability to deliver on its mandate, or stated goals, and propose a similar three-dimensional measurement of results.³³ Biermann and Bauer distinguish measurements of multilateral effectiveness along the *output* (activities), *outcome* (behavioural changes pursued by the organization), and *impact* (accomplishments vis-à-vis policy targets). According to this typology, in an ideal world multilateral effectiveness should be measured at the *impact* level, but acknowledging that

²⁶ Ibid.: 38.

²⁷ Barnett, M. N., & Finnemore, M. (1999). The politics, power, and pathologies of international organizations. *International Organization*, 53(04), 699-732.

²⁸ Weaver, C. (2008). *Hypocrisy trap: The World Bank and the poverty of reform*. Princeton University Press.

²⁹ Nooruddin, I., & Simmons, J. W. (2006). The politics of hard choices: IMF programs and government spending. *International Organization*, 60(04), 1001-1033.

³⁰ Dreher, A., & Gassebner, M. (2012). Do IMF and World Bank programs induce government crises? An empirical analysis. *International Organization*, 66(02), 329-358.

³¹ Voeten, E. (2005). The political origins of the UN Security Council's ability to legitimize the use of force. *International Organization*, 59(03), 527-557.

³² Note that even in the literature on corporate performance, the issue of which numerical value best captures corporate performance is hotly debated. Nohria, N., & Khurana, R. (Eds.). (2013). *Handbook of leadership theory and practice: An HBS centennial colloquium on advancing leadership*. Harvard Business Press. More recently, measures such as earnings per share have been criticized by those pressing companies more comprehensively to assess their performance by measuring the value it has created and estimating its ability to create more: http://www.mckinsey.com/insights/corporate_finance/measuring_long-term_performance

³³ Notably Biermann and Bauer (2004) prefer to talk about effectiveness while Gutner and Thompson (2010) focus on performance. See Gutner, T., & Thompson, A. (2010). The politics of IO performance: A framework. *The review of international organizations*, 5(3), 227-248.; Biermann, F., & Bauer, S. (2004). Assessing the effectiveness of intergovernmental organisations in international environmental politics. *Global Environmental Change*, 14(2), 189-193.

“linking observable environmental improvements to the specific influence [of an international organization] is virtually impossible,” the authors focus on organizational *outcomes*.³⁴

Gutner and Thompson’s typology is somewhat similar, but drawn along a continuum from “process-based” to “outcome-based” approaches, where they distinguish between measurements of *processes* (routine activities) at the ‘micro’ level, *outputs* at the intermediate level, and *outcomes* at the macro level. The authors do not place much confidence in macro level measurements, under which they conflate both outcome and impact-type assessments. They argue that “performance measured in terms of outcomes may not be appropriate when [multilaterals] are constrained by various political and other factors outside of their [the organization’s] control.”³⁵ MOs cannot control the environments in which they work. For instance, the spike in global food prices undermined the effectiveness of the World Food Programme at delivering food aid, but this was outside of their control. Gutner and Thompson are supportive of micro measurements of process performance as a means to “appreciate the context ... and observe contingencies that constrain [a multilateral’s] conduct,” but they warn that process performance is an insufficient condition to reach desired outcomes.³⁶ They argue that ‘midterm outputs’ might be the best solution: “while [they] do not always lead to problem solving, they are likely to be associated with organizations that perform well.”³⁷

TABLE 3: Competing Typologies of Multilateral Effectiveness Measurement

Biermann & Bauer (2004)		Output	Outcome	Impact
Gutner & Thompson (2010)	Processes	Output	Outcome	

The Process-Outcome measurement continuum arising from both typologies casts a net wide enough to fit all of the multilateral effectiveness and performance definitions identified on the various existing assessments. The following *Figure 1* adapts Gutner and Thompson’s graphic representation of their measurement typology to integrate the definitional approaches listed on *Table 3*. These IR typologies draw on a much broader literature which also explores the relationships between inputs, outputs, outcomes and impact.

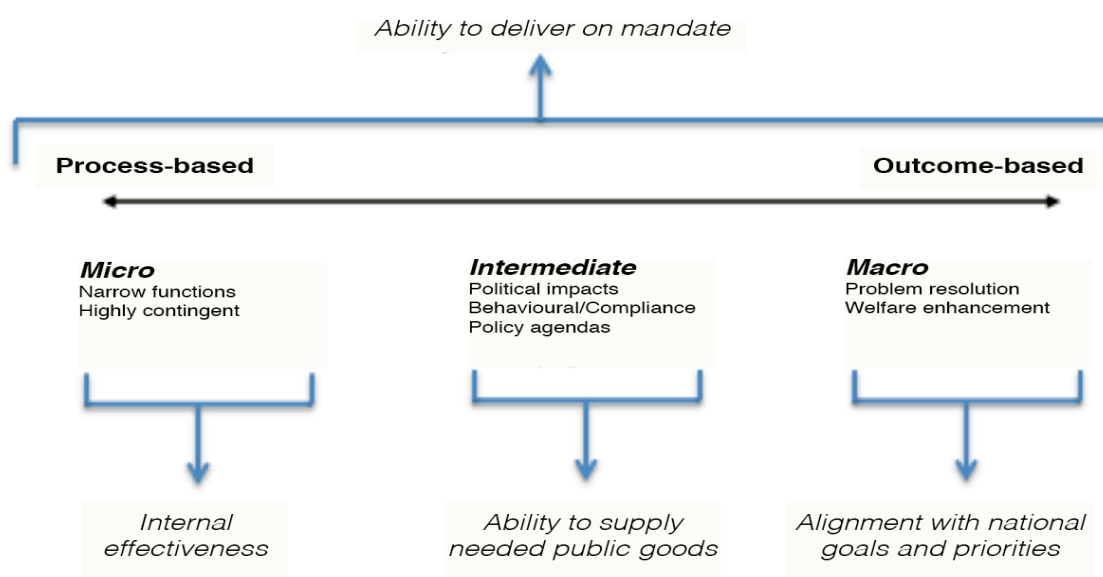
³⁴ Biermann, F., & Bauer, S. (2004). Assessing the effectiveness of intergovernmental organisations in international environmental politics. *Global Environmental Change*, 14(2), 189-193: 191.

³⁵ Gutner, T., & Thompson, A. (2010). The politics of IO performance: A framework. *The review of international organizations*, 5(3), 227-248: 235.

³⁶ *Ibid.*: 236.

³⁷ *Ibid.*: 235.

FIGURE 1: Typology of Performance Measurement³⁸



The IR debate on approaches to measuring international organization performance and effectiveness is ongoing. Some argue we should focus on the process level and confront the fact that “studying broader organizational impacts presents difficult methodological problems.”³⁹ Meanwhile, proponents of outcome-focused assessments fear that focusing too much on outputs and process might lead to what Robert Picciotto has called the ‘development outcome-agency performance disconnect’ – a situation in which multilaterals excel at completing tasks and delivering outputs, but have little to no impact on the wide issues they are trying to solve.⁴⁰ Picciotto also suggests that results-based assessments cannot as yet fully respond to the “host of measurement problems, contractual dilemmas and principal-agent constraints” that shape multilateral performance. Furthermore, others have argued that it is problematic to assume effective internal governance, or that delivery of outputs has a clear impact in terms of change on the ground.⁴¹ This will be discussed further in the subsequent section.

In summary, IR scholars have looked at various dimensions and explanations of effectiveness. Scholarship has paid particular attention to the governance of multilaterals – do multilateral command stakeholder support? How do states monitor and ensure multilaterals deliver on their mandated tasks? How effective and legitimate are MOs? They have examined the internal bureaucratic functioning of international organizations and their dysfunctions, pathologies and contradictions. They have examined the impact and results of many multilaterals (particularly the World Bank, the IMF and the UN Security Council). IR scholars are divided over what dimension of effectiveness to study (internal; output or outcome) and what the relationship is between them.

³⁸ Adapted from Gutner, T., & Thompson, A. (2010). The politics of IO performance: A framework. *The Review of International Organizations*, 5(3), 227-248: 234.

³⁹ Gutner, T., & Thompson, A. (2010). The politics of IO performance: A framework. *The Review of International Organizations*, 5(3), 227-248: 244.

⁴⁰ Picciotto, R. (2006). Development effectiveness at the country level. *German Development Institute*.

⁴¹ Ramalingam, B. (2013). *Aid on the edge of chaos: rethinking international cooperation in a complex world*. OUP Oxford.

What are the challenges of measuring effectiveness?

Here we highlight and discuss the key challenges of measuring effectiveness, many of which arose from the workshop held at the Hertie School of Governance on September 17th, 2015. This is not necessarily an exhaustive list but provides a useful starting point for practitioners and IR scholars to reflect on where and how they can contribute to furthering the debate.

Do we need a common definition of effectiveness?

There are many definitions of effectiveness used across the academic and policy sphere. Effectiveness will differ in different national contexts, different organizations and from different perspectives (donor versus recipient). Most practitioners agree that it is unlikely we would ever reach a common definition of effectiveness. States will want to assess if multilaterals align with their national priorities. A common framework of effectiveness, or a shared set of indicators on organizational performance, may be more useful. Such a framework of common indicators could then serve as the basis for national assessments. It is important to note the difference between a “common definition” and a “baseline of common indicators” for assessing effectiveness: the latter takes into consideration that different actors might always have different definitions of effectiveness, but seeks to give them a shared baseline of indicators and data from which to build their individual definitions. That is what MOPAN seeks to provide to all of its Member States: quality baseline assessments from which Member States can build their bilateral assessments according to their conceptual definitions and priorities. Despite differing notions of what effectiveness means, MOPAN Member States’ views on effectiveness have nevertheless tended to converge when working from MOPAN’s shared baseline of effectiveness indicators.

How can we measure organizational effectiveness accurately?

Identifying measurable and robust indicators of organizational effectiveness is challenging as the previous section illustrated. It is difficult to translate multilateral mandates and mission statements into clear measurable targets. Furthermore, academics and practitioners have highlighted the difficulties of evidencing a causal link between a multilateral’s processes, outputs and outcomes. There are a wide-range of intervening and contextual variables – from levels of corruption in a country, to the global financial crash – which may impact on a multilateral’s operations.

In addition, some scholars and practitioners are sceptical that any causal link can be made between output and impact or outcomes.⁴² They argue it is difficult to identify if and how a particular output of a MO (such as a capacity building programme) leads to significant social, political or economic change in a country. Furthermore, practitioners have questioned how important organizational effectiveness is for organizational results. A ‘poorly run’ organization may still achieve results. There may be other ways to achieve development and humanitarian results which do not coincide with views of what a good, ‘effective’ organization is. Although results-based management and many assessments assume there is a causal link between process, outputs and outcomes, the causal link between these is often ambiguous.⁴³

Many practitioners and assessors highlighted the challenge of demonstrating a link between organizational effectiveness and results on the ground. For instance, multilateral effectiveness at the country level may differ greatly depending on the political context of a country and whether an organization is operating in a democratic regime or not. Humanitarian and development organizations may have greater access and lobbying power in democratic regimes than the authoritarian, or vice-versa. Practitioners have emphasised the need for attention to what is realistic in a given context.

⁴² See Ramalingam, B. (2013). *Aid on the edge of chaos: rethinking international cooperation in a complex world*. OUP Oxford.2

⁴³ Lipson, M. (2010). Performance under ambiguity: international organization performance in UN peacekeeping. *The Review of International Organizations*, 5(3), 249-284.

Yet, donors' political need for accountability and the policy community's interest in results-based management mean that institutional evaluators will continue to look for outcome- and impact-based measurements. They can use theories of change to draw causal links between outcomes and impacts. Theories of change should be "clear and explicit, not over-ambitious and based on a solid understanding" of organizational mandates.⁴⁴ Assessments should be based on a theory of change which connects the multilateral's actions to broader social, political and economic change. It remains an open question how institutional evaluators can balance the need for methodological rigor while assessing organizational outcomes and impact.

How can we balance donors', recipients' and other stakeholders' assessments of effectiveness?

Multilateral performance assessments should all capture recipients' and other stakeholders' perceptions of effectiveness. Assessments which already undertake wide stakeholder participation, such as MOPAN, could increase the representation of country-level respondents, and in particular facilitate the inclusion of civil society and beneficiary groups (i.e. refugee groups, when assessing the performance of the UNHCR). For bilateral assessments, donors should bear in mind that greater involvement from beneficiary populations adds credibility to assessments. It is critical that assessments include the voices of beneficiaries and those affected as these may challenge the conceptions of what counts as successful development or intervention. The World Bank's Report, *Voices of the Poor*, for instance documented the views of more than 60,000 women and men from 60 countries to understand poverty from the perspective of the poor themselves. The report challenged World Bank preconceptions about what constituted poverty, and how poor people conceive their chances.⁴⁵ Research by the Overseas Development Institute on recipient stakeholder perceptions of MOs identified "predictability, flexibility, and alignment with recipient-country priorities" as main characteristics, for recipient stakeholders, of well-performing organizations.⁴⁶ One recommendation would be to include more individuals from recipient countries who had experience with multilateral institutions in assessments.

Should we aim to harmonize bilateral assessments?

Donors and multilateral officials recognise that the multiplicity of multilateral assessments imposes severe transaction costs and takes up substantial institutional resources that could be more effectively deployed.⁴⁷ Multilaterals are expected to contribute significant time and staff resources to joint performance assessments and bilateral assessments. Former OECD/DAC Chair Sir Richard Manning explains donors' preference for bilateral assessments as part of a general "built-in notion ... that 'bilateral is best.'" He has been at the forefront of those policy-makers calling for the enhancement of MOPAN's performance and reach as a means to fixing what he argues is a currently "dysfunctional ... series of ill-coordinated individual reviews."⁴⁸

Yet we are likely to see bilateral assessments continue. Existing collective assessments of multilaterals do not answer donors' specific questions on alignment with their national goals. When each donor country entertains different priorities and goals, assessing an organizations' alignment with donor interests through a collective exercise would be tricky, at best. Furthermore, multilateral assessments may not keep up with the 'flavour of the month' or particular groupings of donor countries. Key questions to explore include: how can we manage a crowded field of overlapping assessments? And should assessments of multilateral effectiveness be comparative?

One method to harmonize assessments would be to concentrate on a common framework, such as MOPAN, which would conduct assessments of organizational effectiveness for all donor states. Donor states would then use these as the basis for their own assessments and conduct their own independent analysis, as necessary, of

⁴⁴ Proudlock, K., Ramalingam, B., & Sandison, P. (2009). Improving humanitarian impact assessment: Bridging theory and practice. 8th Review of Humanitarian Action: Performance, Impact and Innovation: 8.

⁴⁵ Narayan, D., Belshaw, D., Calderisi, R., & Sugden, C. (2001). *Voices of the Poor. Faith in Development*, World Bank and Regnum Books International, 39-48, retrieved 1/23/2016 from <http://goo.gl/WPn8CZ>.

⁴⁶ Wathne, C., & Hedger, E. (2010). *What Does an Effective Multilateral Donor Look Like?*. London: Overseas Development Institute. Retrieved 1/31/2016 from <http://www.odi.org/publications/4800-effective-multilateral-donor>.

⁴⁷ OECD. (2015). *Multilateral Aid 2015: Better Partnerships for a Post-2015 World*: 79.

⁴⁸ Ibid.: 82-83.

organizational relevance to national priorities. The advantage of this approach is that it would streamline and standardize assessments of organizational effectiveness, minimizing the reporting pressure on multilaterals. It would also mean that all donors would operate from a similar baseline of which organizations were effective and in what areas. A limitation of this approach is that it would focus solely on individual organizational effectiveness, not on the system as a whole. Such assessments would need to be carried out regularly to ensure they were not dated.

Another potential is that assessments could be carried out holistically across a particular sector such as agriculture, health or development. In this case assessments would focus on how a group of institutions collectively delivered global public goods in one area. This would be a more systematic approach to assessment but would require new frameworks. Thirdly, the multilateral system could be evaluated on how it is contributing towards the Sustainable Development Goals. Assessments would focus on each goal, which institutions were assisting with realizing it, and how effective they were at doing so.

How can we balance robust universal assessments with the need to take organizational differences into account?

There are significant differences between organizations – from the size of their bureaucracy, the scope of their mandate, the age of their creation, and the functions they hold.⁴⁹ In particular, practitioners highlighted how difficult it is to compare normative organizations and operational organizations. Normative organizations have a role in setting global standards (such as World Health Organization, International Labour Organization, United Nations High Commission for Refugees) and it is difficult to capture their success at performing this element of their work. It is much easier to measure success of organizations that primarily deliver a service or programme to states, such as the World Food Programme and UNICEF.

Thus an open question for academics and assessors of MOs is: How can we design and implement robust, universal performance assessments given the significant differences between international organizations? Should multilateral assessments use the same criteria to assess a wide range of multilaterals? And how should we differentiate between operational and normative (standard setting) organizations? In sum, there is a tension between standardized indicators which are often favoured by donors, and those that are flexible and specific to individual organizations, which are preferred by multilaterals.

This leads to whether we should be attempting to conduct standardized and comprehensive assessments at all. Some would suggest we should opt for more tailored assessments which are specific to organizations. Others would support a common, standardised framework and criteria for assessing multilaterals but would argue these assessments should not be comparative exercises. It is an open question whether we should rank international organizations against each other given significant differences in their mandates, governing and bureaucratic structures (see for an example of this the DFID Multilateral Aid Review).

How are multilateral assessments used?

Another challenge is how to ensure that assessments are used by the multilateral and inform organizational learning. Practitioners' responses varied, and many emphasised the importance of ensuring that the organization supported the entire process of conducting an assessment and was not just focused on a final report. There was a concern that too often organizational assessments are ignored. One workshop participant mentioned that assessments' impact on organizational performance may relate not to their accuracy or robustness, but to who was involved in carrying them out (i.e. the higher the political capital of individuals involved the greater the chances for change). Scholars, such as Lipson, have also argued that better measurement of performance may not lead to better performance.⁵⁰ Meanwhile Moynihan in the context of

⁴⁹ Hall, N. (2004). Money or Mandate? Why International Organizations Engage with the Climate Regime. *Global Environmental Change*, 15(2): 79-97.

⁵⁰ Lipson, M. (2010). Performance under ambiguity: international organization performance in UN peacekeeping. *The Review of International Organizations*, 5(3), 249-284.

national politics has found that when performance information is used adversarially to distribute funding, it is not useful, since it is seldom definitive or uncontroversial. The use of performance assessment results in a more collegial manner, through what Moynihan calls “learning forums,” (routines where performance information is deliberately examined), and is more likely to enable learning and drive change in organizational culture.⁵¹ Assessments could also track their own impact, and follow-up on what changes (if any), are made in multilaterals as a result.⁵² MOPAN, for instance, surveys its members on whether they find the reports useful.

Many political dimensions can enable or interfere with evaluations and assessments. Donor countries may, for instance, conflate their own national priorities with organizational effectiveness and then use this as a basis to remove financing from organizations perceived to be ineffective. Practitioners also highlighted the need to include and inform a broad range of stakeholders through the process of assessment. Assessors should consider: who do we need to include in the conversation to ensure we see an improvement in results? Ideally management and the Executive Board will clearly delineate their roles and responsibilities in terms of governance and organizational reform. In fact, some proposed that assessments should evaluate the governance by the Executive Board in addition to evaluating the bureaucracy. Executive Boards play a critical role in the administration of multilaterals, but their governance is rarely considered in assessments.

Future Research Agenda

Here we outline several areas that should be the focus for future research by IR scholars and also policy evaluators. We emphasise the importance of developing robust assessments of effectiveness, while also ensuring that the assessments effectively inform organizational learning.

Evaluating the Governing Board of Multilaterals

IR scholars could look to develop metrics of effective governance by Executive Boards. They could draw on the well-established literature on why states use international organizations, and when delegation is effective. Principal-agent theory could shed light on when and why states may delegate mandates which are difficult for an international organization to implement. This scholarship could consider: what are measurable indicators of effective delegation to a multilateral? One element could include the clear definition of tasks which do not conflict with other existing organizational responsibilities. Policy evaluators can rarely examine questions of effective Board governance, given their work is most often commissioned by the Board (or a donor state). It is thus difficult to criticise donors, and generally there is no scope for this in their evaluations. It would be particularly useful for scholars to take up these questions.

Developing Indicators and Theories of Organizational Effectiveness

As noted in this working paper IR scholars and assessors of multilateral effectiveness use a wide range of indicators. IR scholars could focus on refining universal indicators, as well as organization specific criteria. A particularly useful contribution would be identifying the causal links between internal processes, outputs and outcomes. Research should test when and why a particular dimension of organizational effectiveness leads to impact on the ground. Scholars could also develop typologies to understand differences between normative and service organizations.

Ensuring assessments of multilateral effectiveness make a positive difference

Scholars could also explore when multilateral assessment triggers organizational learning and reform. It would be useful for donors, organizations and assessors to know more about the political contingencies which influence the impact of a particular assessment. Scholars could conduct a comparative study of bilateral and/or multilateral assessments and explore if, and in what ways they influenced multilateral and donor’s decisions.

⁵¹ Moynihan, D. P. (2005). Goal-based learning and the future of performance management. *Public Administration Review*, 65(2), 203-216.

⁵² Note that the MOPAN 2014 Annual report does this, see MOPAN. (2014). 2014 Annual Report. Retrieved 1/23/2016 from <http://www.mopanonline.org/publications/MOPAN%20Annual%20Report%202014.pdf>, 19-21.

There is an emerging literature on this subject which IR could build on and contribute to. Evaluators, such as Michael Quinn Patton, have examined the feedback loops during the evaluation cycle which enable learning to happen, and looked at what enables a ‘culture of evaluative thinking’ in which the organization staff and management comes to value the “questioning, internalize it, get better at it, and ultimately do it themselves.”⁵³

Conclusion

Donors, recipient countries, and citizens look to international institutions to provide critical public goods. Furthermore our international institutions are facing a changing global context: there are new global actors from global multi-stakeholder partnerships to non-traditional donors such as China, India and Brazil. Multilaterals are also under pressure to deepen the involvement of recipient governments in the planning and implementation of their work. Our global institutions must also demonstrate if and how they are contributing to the sustainable development goals, established in 2015. It is crucial to develop methodologically robust indicators, methods and theories of effectiveness.

This paper examined the growth of multilateral assessments. It firstly looked at those conducted by practitioners on behalf of and/or by states and international organizations and secondly examined the international relations scholarship on effectiveness of global institutions. It examined the purposes and scope of a range of multilateral assessments – from MOPAN to DFID. These assessments have aimed to: inform donor strategies and in some cases financing; share good practices; provide systematic information on development results; and improve performance of multilaterals. However, these assessments do not cite any IR scholarship, hence the need to build understanding between these two communities.

Meanwhile IR scholars have examined four dimensions of effectiveness. They have looked at: if and how multilaterals command stakeholder support; whether they have legitimate and effective governance by their executive board; their internal management and bureaucratic processes, and finally examined their organizational results on the ground. This paper pointed out how these strands of scholarship would benefit from further engagement with practitioners to inform their own understanding and measures of effectiveness.

There are after all many challenges that academics and practitioners face in measuring multilateral effectiveness. These include accurately and robustly measure effectiveness, and in particular link organizational outputs to impact. In addition, practitioners are asking how they can balance donors’, recipients’ and other stakeholders assessments of effectiveness, and how to balance universal assessments with the need to take organizational differences into account. Ultimately, it is also imperative to examine if and how assessments are used by the organizations and their governing boards. Future IR research could contribute to answering these questions. They could do so by: developing indicators and theories of organizational effectiveness, examining when organizational assessments make a positive difference and also evaluating the governing boards of multilaterals. IR researchers would also benefit from greater dialogue with the public administration scholarship on performance in national bureaucracies.

Overall this working paper has argued that IR scholars and assessors of MOs would benefit from greater dialogue. We would encourage scholars and practitioners to take up these questions, and continue to collaborate through workshops, conferences and/or an online network of multilateral assessors which could be instigated by MOPAN. After all it is crucial to assess and enable multilaterals to operate effectively given the scale and scope of global challenges.

⁵³ Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford Press: 63.

Appendix

Overview over Existing Assessments

***MOPAN Organizational Effectiveness Assessments (multilateral)*⁵⁴**

Coordinated by 19 donor countries, MOPAN assesses the effectiveness of multilateral organisations, exploring whether these organisations have the systems, processes and behaviours needed to be effective over time, and the results achieved by multilateral organisations, particularly at country level. The goal is to make available quality, comprehensive, qualitative as well as more quantitative data on development effectiveness of MOs that can be used at once by donors to hold MOs accountable, and by the MOs themselves to promote reform and enhance donor-MO interaction.

Between 2009 and 2014, MOPAN worked under a “Common Approach” to multilateral organisation assessment, built around four strategic dimensions of organisational effectiveness (strategic management, operational management, relationship management, knowledge management and development results). From 2015, MOPAN started implementing the MOPAN 3.0 assessment framework, which built on the existing indicator framework but added emphasis on MOs’ effectiveness in achieving results at country level. Data is collected through a survey (tailored to the operations of the different types of multilateral organisations), consultations and a document review process.

***QuODA (multilateral)*⁵⁵**

The Quality of Official Development Assistance (QuODA) measures donors’ performance on 31 indicators of aid quality to which donors have made commitments. The indicators are grouped into four dimensions associated with effective aid: maximizing efficiency, fostering institutions, reducing the burden on partner countries, and transparency and learning. These indicators can be interpreted as measures of “high-quality” aid, and were selected based on what donors, aid agencies and academics have concluded are important elements of aid effectiveness.

Highlighting that donors each have specific strengths and weaknesses, rarely excelling in all dimensions, QuODA reports make it clear that their purpose is simply to allow donor agencies (both bilateral and multilateral) to review where they stand against their peers on critical dimensions of development effectiveness. The main purpose is to serve as tools for starting dialogue and promoting learning.

***Peer Review of Evaluation Functions (multilateral)*⁵⁶**

The Peer Reviews of the Evaluation Functions of MOs are carried out jointly by the UN Evaluations Group and the OECD Development Assistance Committee (DAC), based on a “Framework for Professional Peer Reviews of the Evaluation Function of UN Organizations” (2011). The peer reviews focus on three core issues of evaluation function performance: independence, credibility and utility. Their main purpose is to “share good practice, experience and mutual learning, building greater internal capacity and external confidence in UN evaluation systems and thus ultimately reducing demands for external assessments of organizations’ performance and effectiveness”.

⁵⁴ <http://www.mopanonline.org/commonapproach/>

⁵⁵ Birdsall, N., & Kharas, H. (2014). The Quality of Official Development Assistance (QuODA). *Center for Global Development, Washington, DC*. Retrieved 1/23/2016 from http://www.cgdev.org/sites/default/files/QUODA_final_revised_september.pdf.

⁵⁶ UNEG. (2011). UNEG Framework for Professional Peer Reviews of the Evaluation Function of UN organizations. (UNEG/REF(2011)). Retrieved 1/23/2016 from <http://www.uneval.org/document/detail/945>.

The Peer Reviews are neither full-fledged evaluations nor organization-wide assessments – they are more in-depth assessments that follow rigorous methodology to apply key principles of evaluation while drawing on peer review mechanisms to facilitate cross-organization dialogue and stimulate organizations to change, achieve goals and meet standards.

Development Effectiveness Review (multilateral)⁵⁷

A joint meta-evaluation/meta-synthesis approach developed by the members of the DAC Evaluation Network to complement the peer reviews of evaluation functions and the institutional assessments by MOPAN. The overarching purpose is to overcome the perceived information gap between the information on institutional performance and development effectiveness made available by multilaterals' own independent evaluation departments and management boards, and the information demands of a host of bilateral donors and other stakeholders.

The reviews provide an independent and evidence-based assessment of the development effectiveness of the humanitarian and development programming, based on six main criteria: achievement of development objectives and expected results; cross-cutting themes (environmental sustainability and gender equality); sustainability of results/benefits; relevance of interventions; efficiency; and, using evaluation and monitoring to improve development effectiveness.

COMPAS (multilateral)⁵⁸

The Common Performance Assessment System was designed in 2005 as a framework through which the multilateral development banks (MDBs) could track their capacities to manage for development results. Much like the peer reviews of the UNEG/DAC, COMPAS has become an established and recognized report for constructive dialogue within and among MDBs and partners to sharpen their focus on managing for development results (MfDR).

COMPAS is a self-reporting exercise focused on measuring the MDBs' capacities to apply and improve operational processes toward achieving results on the ground. It measures MfDR capacity and progress consistently through the analysis of key performance indicators consolidated into a concise and convenient format, and comparable within each MDB. COMPAS is not designed to make direct comparisons across institutions, although the matrix format of the report does provide opportunities for the MDBs to learn from one another.

Publish what you Fund (multilateral)⁵⁹

An assessment of donor behaviour on aid transparency, covering 68 aid agencies, including 50 bilateral donors, 17 multilaterals and the Bill and Melinda Gates Foundation. The principle behind this assessment is simple: aid transparency matters for improving governance and accountability, increasing the effectiveness of aid and lifting as many people out of poverty as possible.

The purpose is to stimulate aid transparency among reporting agencies, through continuous assessment along 39 indicators, drawing on a number of existing data sources.⁶⁰ The study assesses

⁵⁷ <http://www.oecd.org/dac/evaluation/evaluatingmultilateraleffectiveness.htm>

⁵⁸ <http://www.mfdr.org/COMPAS/index.html>

⁵⁹ <http://www.publishwhatyoufund.org/resources/assessment/>; Dinham, M. (2011). Study of AusAID's Approach to Assessing Multilateral Effectiveness: A study commissioned by the Independent Review of Aid Effectiveness to assist in their overall analysis of the effectiveness and efficiency of the Australian aid program. *ALNAP*. Retrieved 1/31/2016 from <http://www.alnap.org/resource/9851>: 56.

⁶⁰ Including the OECD DAC Creditor Reporting System; the Paris Declaration Monitoring Survey; the HIPC Capacity Building Program; the OECD Predictability Survey; the EU AidWatch 2010 Survey; and information from the International Aid Transparency Initiative (IATI).

donors in three main categories: donors' overall commitment to aid transparency; transparency of aid to recipient government; and transparency of aid to civil society, providing a general ranking.

Monitoring reports for the Global Partnership (multilateral)⁶¹

These quarterly reports published by the Global Partnership for Effective Development Cooperation track behaviour change in development co-operation, with a focus on the strengthening of results for developing countries and effective co-operation organizational practices, based on the Busan Declaration. A voluntary exercise for bilateral and multilateral donor agencies, the reports' stated purpose is to support accountability based on Busan standards, highlight opportunities and challenges for further progress, stimulate inclusive dialogue on improving the effectiveness of development co-operation, and promote agreement on actions to use the Busan principles successfully.

The framework does not monitor development outcomes, focusing instead on the following indicator framework:

1. Development co-operation is focused on results that meet developing countries' priorities
2. Civil society operates within an environment which maximises its engagement in and contribution to development
3. Engagement and contribution of the private sector to development
4. Transparency: information on development co-operation is publicly available
5. Development co-operation is more predictable
6. Aid is on budgets which are subject to parliamentary scrutiny
7. Mutual accountability among development co-operation actors is strengthened through inclusive reviews
8. Gender equality and women's empowerment
9. Effective institutions: developing countries' systems are strengthened and used
10. Aid is untied

UNDCF's mutual accountability survey (multilateral)⁶²

Based on calls for greater mutual accountability in development cooperation coming out of the 2008 Doha Review Conference on Financing for Development, the 2010 MDG Summit, and the 2011 Busan Partnership for Effective Development Cooperation, the UN Development Cooperation Forum (UNDCF) has used its mutual accountability surveys as a hub for candid and inclusive exchange among aid donors and recipients. The survey's stated purpose is to review progress made in implementing mutual accountability, identifying how to implement enabling policies and results frameworks, identifying key challenges or barriers to mutual accountability, and promoting inclusive dialogue.

The survey covers six broad areas: national aid/partnership policies, mutual accountability coordination forums in recipient countries, quality/transparency of information on development cooperation flows, support for capacity development at country level, impact of national mutual accountability processes, and recipient country's overall evaluation of mutual accountability.

⁶¹ <http://effectivecooperation.org/about/global-monitoring-framework/>

⁶² http://www.un.org/en/ecosoc/newfunct/DCF_account.shtml

Australian Multilateral Assessment (bilateral)⁶³

The Australian Multilateral Assessment (AMA) covers 42 MOs and is designed to:

- provide a firm base of information about the effectiveness and relevance of multilateral organisations, from the perspective of the Australian aid program;
- inform decisions on funding allocations;
- design a rating system that can be used on an annual basis to inform decisions on subsequent funding allocations and policy engagement.

The AMA considered multilateral organisations against an assessment framework that includes seven components. The first three components relate to results and relevance (delivering results, alignment with Australia's priorities, contribution to the wider multilateral development system), and the other four components relate to organisational behavior (strategic management and performance, cost and value consciousness, partnership behavior, transparency and accountability).

The AMA acknowledges that organisations are assessed against the development-related aspects of their mandate, which might not fully reflect the contribution of organisations with mandates stretching well beyond development, such as the United Nations Educational, Scientific and Cultural Organization (UNESCO).

UK Multilateral Aid Review (bilateral)⁶⁴

The UK MAR arose out of desire of the British Government to assess the value for money offered by all the multilateral organisations that receive significant core aid funding. The value for money calculation is the core but not the only purpose of the Review, which also seeks to provide evidence basis for the UK to work with the multilateral organisations and other partners to develop a focused reform agenda, and to identifying areas of weakness, as well as strength, across the whole multilateral system, and to help monitor progress.

The MAR groups its nine indicators into two indices: (i) the contribution to UK development objectives, which primarily assesses what organisations do (critical role in meeting development objectives, attention to cross-cutting issues, focus on poor countries, contribution to results); and (ii) organisational strengths, which primarily assesses how well they are organised and managed (Strategic and performance management, Financial resource management, Cost and value consciousness, Partnership behaviour, Transparency and accountability, Likelihood of positive change). These composite indices are then used to determine a value for money score for each multilateral: *poor* value for money, *adequate* value for money, *good* value for money, and *very good* value for money.

Denmark Multilateral Review (bilateral)⁶⁵

Denmark stopped carrying out separate bilateral assessments in 2010 in light of a policy decision that MOPAN should be one of its key instruments for assessing MO development effectiveness. Prior to

⁶³ Dinham, M. (2011). Study of AusAID's Approach to Assessing Multilateral Effectiveness: A study commissioned by the Independent Review of Aid Effectiveness to assist in their overall analysis of the effectiveness and efficiency of the Australian aid program. *ALNAP*. Retrieved 1/31/2016 from <http://www.alnap.org/resource/9851>.

⁶⁴ DFID. (2011). Multilateral Aid Review: Ensuring maximum value for money for UK aid through multilateral organisations. Retrieved 1/23/2016 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67583/multilateral_aid_review.pdf.

⁶⁵ Dinham, M. (2011). Study of AusAID's Approach to Assessing Multilateral Effectiveness: A study commissioned by the Independent Review of Aid Effectiveness to assist in their overall analysis of the effectiveness and efficiency of the Australian aid program. *ALNAP*. Retrieved 1/31/2016 from <http://www.alnap.org/resource/9851>: 58.

that, Denmark had collected yearly data on the performance of its three most important multilateral partners in each of 15 countries. DANIDA, the country's aid agency is now focused on building the performance of MOPAN and producing organisational strategies on each MO receiving more than 20m Danish krone (about A\$ 3.7m) each year, or less if they have a particular strategic importance. The strategies do not imply direct links between MO performance and funding, but concern with development results means that more and more MOs' performance against the DANIDA strategies can impact future funding decisions.

Linked to these organisational strategies, Denmark also has specific performance frameworks in place for certain UN agencies. These are agreed with the agencies and done jointly with other donors, and contain specific performance milestones linked to the agencies' own strategies and results frameworks. Unlike DFID, Denmark does not link MO performance against these milestones directly to funding allocations. Other Danish reporting processes include annual reports to Parliament and DANIDA, and DAC/UNEG peer reviews of MO evaluation functions.

Sweden Multilateral Review (bilateral)⁶⁶

Sweden has assessed MOs' development effectiveness since 2008, when a first review was conducted on 23 organizations to provide baseline judgements on all the organisations it supported. Follow up on that initial work is taken once every 3-4 years for each MO, with the timing where possible decided to coincide with forthcoming decisions on replenishment or budget rounds. The aim of the assessments is two-fold: to feed into a more strategic policy approach to each MO, by providing the basis for the organisational strategy and informed policy dialogue; and to inform upcoming decisions on funding.

The assessments are based around three broad criteria: relevance; organisational effectiveness; and development effectiveness. They draw on MOPAN reports to the extent possible for information on organisational effectiveness.

Norway Multilateral Review (bilateral)⁶⁷

Norway currently produces a strategy or 'profile' on each MO it supports, gathering information on their activities, results and performance from a range of existing sources such as MOPAN and the MOs' own reporting and policy documents. Norway does not supplement this material with any separate formal bilateral assessments of its own. These profiles help structure proposals on core funding to MOs made by the Government to Parliament.

MO profiles are developed on the basis of the following indicator framework:

1. Their ability to document results,
2. Their relevance in relation to Norway's political priorities,
3. Their systems for planning, budgeting and results reporting,
4. Their systems for internal audit and anti-corruption work,
5. Their contributions to national capacity and institutional development, and national ownership,

⁶⁶ Ibid.: 59.

⁶⁷ Dinham, M. (2011). Study of AusAID's Approach to Assessing Multilateral Effectiveness: A study commissioned by the Independent Review of Aid Effectiveness to assist in their overall analysis of the effectiveness and efficiency of the Australian aid program. *ALNAP*. Retrieved 1/31/2016 from <http://www.alnap.org/resource/9851:60-61>; <https://www.regjeringen.no/en/topics/foreign-affairs/the-un/review-multilateral-org/id737780/>.

6. Norway's opportunities to influence the organisation as a whole, and
7. Their willingness to take concrete steps to implement reform.

The profiles of the organisations each contain a facts and figures section, which gives details of Norway's financial contributions, a description of the organisation's mandate and focus areas, and an overview of results achieved in 2012. The last section describes Norwegian policy towards the organisation in question.

Germany Multilateral Review (bilateral)⁶⁸

Germany looks regularly at each MO it funds, drawing together information on their effectiveness from a variety of existing sources such as MOPAN, OECD/DAC sources, independent evaluations where these exist, the MOs' own reporting, COMPAS reports and so on. In drawing up these assessments Germany looks not only at the link between performance and financial allocations, but attempts to weigh performance against other criteria such as the relevance of the organisation, its comparative advantages and its particular mandate. Germany does not base allocation decisions on comparison between MOs, looking rather at a given organisation's direction of travel to assess history of and potential for implementing reforms to achieve more effective development cooperation.

⁶⁸ Dinham, M. (2011). Study of AusAID's Approach to Assessing Multilateral Effectiveness: A study commissioned by the Independent Review of Aid Effectiveness to assist in their overall analysis of the effectiveness and efficiency of the Australian aid program. *ALNAP*. Retrieved 1/31/2016 from <http://www.alnap.org/resource/9851>: 57.