

MOPAN METHODOLOGY REVIEW 2017-2018

*FINAL REPORT PREPARED FOR THE
MULTILATERAL ORGANISATION PERFORMANCE ASSESSMENT NETWORK*

ELLIOT STERN AND OLE WINCKLER ANDERSEN

MARCH 2018

ELESCO Consulting Limited UK

Contents

PART ONE: Introduction	1
1.1 Report Background	1
1.2 Understanding Methodology Review requirements	2
1.3 A framework: Methodology and its link to purpose and use.....	2
PART TWO: MOPAN PRIORITIES AND THE EXPERIENCE OF MULTILATERAL ORGANISATIONS	4
PART THREE: QUESTIONS RAISED IN TERMS OF REFERENCE FOR THIS REVIEW ...	7
3.1 Applicability of indicators	7
3.2 Consistency and calibration across assessments	9
3.3 Appropriateness and robustness of evidence lines.....	13
End Note	16

Part One: Introduction

1.1 Report Background

This report summarizes the findings of the MOPAN Methodology Review 2017-2018. Occurring relatively early in the lifecycle of MOPAN 3.0 assessment methodology, this was expected to be subject to a 'light-touch' review. We understood this to mean two things. First the review would only address a limited set of early questions that were raised following the initial 2015-2016 MOPAN assessment cycle. Second the reviewers were expected to supplement available information with prior experience they had of the methodological challenges that assessment systems like MOPAN face. To that extent this exercise is both a methodological review drawing on evidence and the views of stakeholders; and a peer review drawing on broader experience beyond MOPAN.

The report was first shared in draft form to with members of MOPAN and used as basis for the presentation and discussions of the MOPAN methodology at the MOPAN meeting in Copenhagen February 28- March 1st, 2018. Comments made by MOPAN members and by the Secretariat following the Copenhagen meeting have been taken into account when finalising this report.

The Copenhagen meeting focussed mainly on future options and recommendations. The presentation material for that meeting, which on the basis of this review highlights future methodology options that MOPAN should consider, is appended to this report as Annex 1.

To undertake this review the consultants conducted interviews and discussions with MOPAN members; MO staff who had experienced MOPAN; and members of the Secretariat. (Twenty-nine interviews/discussions took place.) A number of discussions also took place with IOD PARC in order to better understand the practical problems faced whilst implementing MOPAN assessments.

The consultants familiarised themselves with the MOPAN 3.0 Methodology Manual; a wide range of historical documents including the 2013 MOPAN evaluation; and current assessment procedures drawing on documentation provided by the MOPAN Secretariat and IOD PARC. In addition, the two review teams conducting the parallel Methodology and Strategic Reviews met together on three occasions to develop and share frameworks and perspectives.

The report is organised as follows:

- In the remainder of this Introduction we cover our understanding of the requirements for this review and how it has been framed
- Part Two summarises key issues related to this review that were salient in interviews conducted
- Part Three addresses the questions posed in the Terms of Reference (ToR) for the Methodology review

1.2 Understanding Methodology Review requirements

The ToR describes the rationale and objective of the methodology review. It should focus on challenges applying the assessment methodology (MOPAN 3.0) encountered in the 2015-16 cycle of assessments and in that light identifies three areas in which recommendations are sought:

- Applicability of indicators
- Consistency and calibration' across assessments
- Appropriateness and robustness of evidence lines

At the same time the ToR sets MOPAN in a broader perspective. It notes that MOPAN members 'share a common interest in enhancing the effectiveness of the multilateral system'; that the network 'aims to inform its Members' strategic engagement with multilateral organisations [...] and more broadly inform the international discourse on the multilateral system and its effectiveness.' It is also noted that 'MOPAN recognises that changes in the development landscape, new actors, evolving roles, reference frameworks, and financing structures require the methodological framework to keep up with the changes and challenges that the multilateral system is confronted with'. At a more operational level we understand that the two reviews are also expected to feed into upcoming decisions about the 2018 procurement process, and the renewal for the post-2019 OECD-MOPAN MOU.

Interviews with MOPAN members; staff of MOs that had undergone MOPAN assessments; and with the MOPAN Secretariat underlined these broader concerns. Joint discussions with the parallel Strategic Review also emphasised the forward-looking challenges that MOPAN faces. We were in particular struck by the idea of a 'MOPAN journey' – cited by several stakeholders interviewed - which suggested that MOPAN had already evolved; and that this was likely to continue. This 'journey' was made up of a number of elements: the rigour and systematisation of the assessment methodology; shared understanding among MOPAN members; and a willingness among some members at least to contemplate MOPAN moving beyond a focus exclusively on MO accountability to broader MO and donor concerns. Interviewees gave different degrees of emphasis to the original purpose of accountability and providing donors with feedback on their investments in the multilateral system; the need to enhance MO learning and improvement that resulted from MOPAN; and for some more fundamental longer-term questions about the multilateral systems' capacity to address global challenges as the SDGs and Agenda 2030 takes root.

All stakeholders interviewed continue to prioritise the 'accountability' purpose of MOPAN and agreed that nothing should be done to weaken MOPAN's commitment to that foundational purpose. But there were enough who saw other 'purposes' coming more into focus in the future for us to reflect on the implications this had for MOPAN methodology. This intersects with what we understand to be the scope and meaning of 'methodology'.

1.3 A framework: Methodology and its link to purpose and use

Methodology is sometimes confused with methods and techniques: how to construct scales, sample respondents, organise and analyse data. Whilst all of these activities are included, methodology

encompasses a more extended process that begins with *purpose* of evidence gathering and analysis and ends with users and the *use* to which evidence is put once available and by whom. After the questions that need answering have been identified, methodology helps specify information coverage and form judged as sufficient to provide answers; clarifies what information is available and how it can be collected; and suggests the kind of analysis, synthesis and quality assurance that will be needed in order

to transform information into evidence and draw credible and robust conclusions. This is very much the process that the MOPAN 3.0 Methodology Manual describes.

The table below highlights the connections between the overarching purposes of MOPAN, the primary approaches associated with different purposes and the kinds of methods that this implies. The table suggests that as MOPAN matures in addition to accountability to donors, MOPAN will also face demands – and to an extent has already faced demands - to expand its learning and improvement and system’s development role in the overall multilateral system.

<i>MOPAN’s Purposes</i>	<i>Primary Approach</i>	<i>Typical methods</i>
MO Accountability & Donor Priorities	Inspection and compliance checking	Indicators and ranking; framework setting for MOs
MO Learning & Improvement	Feedback and engagement	Dialogue, ‘positives’ e.g. ‘good practice’, capacity development
Multilateral system Development – MOs and other actors	Networking across MOs and sectors	Facilitating networks, Peer-review, policy dialogue

Whilst MO accountability and MO learning¹ both feature in MOPAN’s current ambitions, work at the level of the multilateral system does not, even though it was raised by a minority of MOPAN members.

It is also worth noting that although learning and improvement does not feature in the ToR for this review, it has been given much greater prominence in MOPAN 3.0 compared with the antecedent ‘Common Approach’. The MOPAN 3.0 Manual takes the intention to learn seriously both as a process during an assessment and at reporting/post report stage; and does so both for MOs and for MOPAN members. The intentions and ways forward to encourage learning through engagement in MOPAN 3.0, follow weaknesses in MO learning identified in MOPAN’s 2013 evaluation.

Given that the so called ‘Learning and Engagement’ strategy is still in an early stage it has been difficult to judge how far it is working. Comments by MOPAN members suggest that it *is* making a difference for

¹ See ‘Learning and Engagement’ Section 12 of Methodology Manual

members themselves. Members knowledge and engagement with MOPAN appears high and members have a greater awareness of issues and future choices than is often the case in similar networks.

Comments by MOs recently or currently being assessed are less clear-cut. At the assessment report stage there was limited evidence of donors engaging with MOs, for example as part of Board discussions. Amongst those currently or recently undergoing an assessment some do regard it as an

opportunity to understand, reflect and learn. But others do not, being more concerned about demands on their time or the fit of the MOPAN framework to their own specific circumstances.

The above table is also intended to frame the discussion that follows about the three main questions posed in the ToR for this review. All these questions are in the top row of the table and concern MO accountability and informing donor priority setting and strategy. Greater emphasis on rows two and three in future would raise different methodological challenges and have considerable implications for skills MOPAN needs and where they should be located.

For example both learning/improvement and multilateral system development would probably require an investment in facilitation and networking skills, as likely to be located in the Secretariat as in subcontractors and definitely requiring more diverse skills and processes.

PART TWO: MOPAN PRIORITIES AND THE EXPERIENCE OF MULTILATERAL ORGANISATIONS

An important part of this exercise was to understand in some detail the priorities of MOPAN members in order to contextualise and frame questions posed in the ToR. It was also important to gain as much insight as possible of the experience and feedback from those who had already experienced assessments under MOPAN 3.0.

What follows is a summary of the main issues that surfaced in interviews with MOPAN members; MO interviewees and with the MOPAN Secretariat.

1. *MOPAN's purpose:* There seems to be a narrow and a broad view of MOPAN's purpose. The narrow view accepts an organisational effectiveness focus and limited coverage of results. The broader view regards results as important – a topic that should be deepened in future. This is consistent with a view that KPIs 1-8 are stronger than KPIs 9-12
2. *Who MOPAN is for:* There is a strong view that MOPAN is intended for donors and also for MOs i.e. to improve their effectiveness after feedback. However, there is also acceptance that in the contemporary development landscape global as well as 'partner countries' and other global development actors should also be expected to benefit from MOPAN
3. *Accountability and Learning:* Members see MOPAN as mainly fulfilling an 'accountability' function for donors. There is also a wish to strengthen the learning function of MOPAN for MOs

in particular. Notions of learning range from ‘behaviour change’ and enforcement (e.g. via funding decisions and discussions at Board level) through to collaborative dialogue and continuous improvement.

4. *MOPAN is ‘on a journey’*: Even amongst those cautious of changing MOPAN 3.0 in the short term, there is a recognition of the relative ‘youth’ of MOPAN and a greater willingness to consider further developments in the medium term. For example investing more in MO learning; or expanding the numbers of MOs that MOPAN would assess. The notion of MOPAN being ‘on a journey’ was widely shared amongst MOPAN members.
5. *Importance of MO engagement*: It was clear that a successful assessment requires strong engagement by the MO concerned in the MOPAN process. Such engagement both allows for the assessments to be calibrated – e.g. by providing appropriate documents and interpretations of MO systems and activities; and for establishing learning dialogue which can encourages new thinking also among MO managements.
6. *Country engagement*: Consistent with the ToR there is a general recognition that country involvement (i.e. potential beneficiary countries) is not currently adequate. This is mainly framed in terms of improving evidence about the ‘results’ of MO activities.
7. *Country data sources*: The limitations of the ‘partner country survey’ were widely referred to. There seemed to be unresolved differences about what other sources of evidence were acceptable: e.g. country embassies and donor coordination arrangements; country own data; and case studies were amongst possibilities raised.
8. *Adaptation to the SDG-agenda*: Several interviewees mentioned that the MOPAN approach and methodology should reflect the SDG-agenda and how the MOs have adapted to this new agenda. This could e.g. comprise a revision and update of KPI 5 on Relationship Management and more focus on ‘clusters’ of MOs.
9. *The meaning of ‘results’ is not always clear*: These are sometimes understood as systems in place expected to contribute to results; and sometimes as actual results. There is also uncertainty among MOs not primarily engaged in development activities as to why MOPAN emphasises ‘development results’, which does not match the mandate or activities of MOs such as WHO, ILO and UNESCO.
10. *Customisation and flexibility*: There is a widespread wish to minimise ‘flexibility’. The standardised character of the MOPAN indicators is seen as a strength, contributing to the credibility of MOPAN in Capitals, but it was also recognised that the number of indicators and sub-indicators was very high. There was some acceptance of differentiation for different classes of MO – in particular for UN normative (specialised) agencies and IFIs. There was less reference to vertical funds.
11. *Lengthy ‘production’ time*: There is widespread view that the MOPAN process is onerous and lengthy – the 2-year ‘production time’ was mentioned consistently. This was seen as a problem both for MOPAN members and for MOs. E.g. MOs having to wait months between first contact by MOPAN and follow-up meetings or interviews. Solutions varied: from fewer MOs in a cycle

(e.g. 6 MOs in 1 year rather than 12 over 2 years); through to 'slimming down' the assessment process.; or if necessary putting more resources into MOPAN. A multi-year workplan that identified MOs to be assessed over future cycles was also seen as a way of reducing elapse times.

12. *Differentiated 'products'*: the possibility of new MOPAN products was mentioned by some interviewees. Such products included synthesis across assessments; results-oriented case studies; and sectoral or thematic clusters. Any differentiation of MOPAN 'products' would raise the question of workload and how this might be resourced, including how to make present activities more efficient.
13. *Frequency of assessments*: There seems little consensus about how often assessments should be repeated. The prime driver of assessment frequency is the policy needs of donors for budget approval. Some suggested an assessment every 2 or 3 years whilst others mentioned 4, 6 or even 8 years cycles. Sources of variance derive both from the different member funding cycles; and different agency funding – e.g. voluntary contributions, budget periods and replenishment rounds. It was suggested that having a longer-term 'workplan' for MOPAN could allow members to align their own planning and funding cycles with MOPAN assessments.
14. *Expansion of scope*: Despite concerns for workload, members appear to be willing to contemplate an expanded coverage by MOPAN in future i.e. more rather than fewer organisations. EU agencies were offered as an example of MOs that could be considered as part of such an expansion.
15. *Data collection and analysis*: It seems to be generally agreed that MOPAN reviews could do more to use existing documentation and other available data sources e.g. existing bilateral reports. In addition there is scope to undertake further analysis of existing data.
16. *Methodological understanding*: There is considerable diversity of background knowledge amongst MOPAN members. This inevitably shapes expectations of what an indicator system can be expected to achieve. Some for example see current indicators as a 'rough and ready' guide only; others use words like rigour, objectivity and science more freely. The use of indicators as a basis for comparison between MOs was generally not seen as appropriate, although a few members seem to consider this use as possible.
17. *Member reporting and organisational arrangements*: There is considerable difference in home Ministry structures – some are from UN only departments others have oversight of all multilaterals; whilst others have responsibilities for both multilateral and bilateral aid. These differences shape their understanding and priorities.
18. *Confidence in MOPAN and Secretariat*: There was overall confidence in MOPAN and in the Secretariat. MOPAN was seen as having undergone a process of 'professionalisation' in recent years. Members seemed receptive to further Secretariat 'institution-building' and leadership.

Most of the issues that surfaced through interviews have been taken on board in this report even though not all were specifically prioritised in the ToR for the Methodology Review. They provide a useful

context both when assessing aspects of methodology; and when considering possible future developments.

PART THREE: QUESTIONS RAISED IN TERMS OF REFERENCE FOR THIS REVIEW

This part of the report addresses each of the main questions raised in the ToR whilst as far as possible also keeping in mind the different purposes that MOPAN members may wish to emphasise in future. This is informed by the results of interviews with MOPAN members and MO informants (mainly ‘Institutional Leads’); by a careful reading of the MOPAN 3.0 Methodology Guide; and prior experience in similar settings.

3.1 Applicability of indicators

Question: To what extent, and in what ways, could MOPAN adapt its indicator framework to ensure optimal applicability to a diverse range of organisations, including but not limited to UN Funds, Programmes, UN Specialised Agencies, Multilateral Development Banks, Vertical Funds, Trustee Entities, other normative mandate/functions?

Rationale: To ensure that the indicator framework is fit-for-purpose.

Standardisation and consistency is one of the perceived strengths of the MOPAN 3.0 system of MO assessment. At the same time one of the persistent complaints of those interviewed from MOs that have been assessed, is the inappropriateness of MOPAN indicators for certain kinds of MOs. The most commonly cited problems are for ‘normative’ MOs such as ILO, WHO, UNESCO etc. Indicators are seen as not dealing with the standard setting work of these MOs in particular. Similar issues or challenges have been identified for other types of MOs – for example vertical funds and GAVI.

Most mentions of the (in)applicability of indicators referred to results-oriented KPIs (9-12) rather than the four ‘Management’ strands – Strategic, Operational, Relationship and Performance. However, in the view of some MO interviewees there are similar problems in other areas such as funding and budgeting which fall under the ambit of Management KPIs.

At present MO differences are handled by relatively minor modifications to micro indicators (MIs) whilst the overarching KPI remains unchanged.² We agree that for most aspects of organisational effectiveness commonly accepted good practice applies to all types of organisations and KPIs. Minor modifications of MIs appear to be sufficient.

² It is not clear the extent to which these modifications are consistently applied and at whose initiative. Some interviewees gave the impression that MI modifications occur only after vigorous lobbying by the MO concerned rather than as a matter of course.

With regard to 'Results' - and potentially the preconditions for results - more differentiation in relation to types of MOs should be considered. This currently pertains to the analysis of pre-existing evaluations and might also have implications for a future country survey or other methods (e.g. case studies) that would assess results in greater depth. (The general question of Results is discussed more fully below in relation to 'country' data and perspectives. This section concentrates on a narrower aspect of results; that pertaining to the 'applicability of indicators'.)

The way that this differentiation might be structured has to be built on a credible typology of MOs and their activities. Typologies of MOs discussed in policy circles tend to follow an administrative classification. The ToR for this Review distinguishes between: UN Funds, UN Specialised Agencies, MDBs, Vertical Funds etc. We would argue however that the underpinnings of MO typologies need to be more nuanced, cutting across administrative categories.

As a heuristic device we have applied a 'task' perspective³ – one that regard all organisations as engaged in a number of primary tasks. MOs appear to be engaged with at least 5 such 'tasks':

- Service and Programme delivery
- Finance
- Capacity building (e.g. of policy and governance systems)
- Standard setting
- Coordination

Whilst most MOs will undertake more than one task, many will mainly focus on one 'core' task – a UN Fund on Service and Programme delivery; GAVI on Coordination and Capacity Building; and ILO on Standard Setting. We would suggest that the assessment of Results would be more appropriate if these core tasks and associated up-stream activities that enabled them, were considered in greater depth. This could be considered as a variant of 'risk based' approach, concentrating efforts on those aspects of an MOs mandate and function that constitute its comparative advantage. (This would also ensure that the MOPAN system was adequately 'calibrated' to match different MO activities and tasks, see below.)

Such an approach could be seen as complementing rather than replacing standardised KPIs which would continue to be applied to all MOs. In essence we are putting forward the idea of a *modular approach* in which *core* modules that relate to performance management are universal but at least one additional *task-specific* module is available for any MO where the core module risks oversimplifying an assessment. Although this logic could be applied most obviously in relation to results, it would also need to be considered in relation to those activities that are upstream from results as well.

The extent to which this approach should be considered is partly related to how 'Results' are understood in MOPAN and how much emphasis there will be in future, as MOPAN evolves. This is discussed further below. The rationale for Question A is 'To ensure that the indicator framework is fit-for-purpose'. Insofar as the main purpose of MOPAN is accountability of MOs; and this is narrowly understood to be the existence of management processes and procedures rather than results then the indicator framework can be said to be 'fit-for-purpose'. Strictly speaking existing KPIs and MIs approach results from a 'results chain' logic, i.e. confirming that management and operational systems assumed to be necessary for results to be achieved are in place. They do not attempt to describe, capture or

³ Gibson Burrell and Gareth Morgan (2017) Sociological Paradigms and Organisational Analysis

measure actual results. If actual results were considered more important it would be more difficult to argue that the existing system is fit for purpose.

It was also pointed out that when MOPAN 3.0 was being finalised the SDGs had not yet been agreed; in advance of the 2017-18 assessment cycle, Members agreed to amend MI1.3 to incorporate Agenda 2030, though in this regard indicators are not considered 'fit-for-purpose'.

3.2 Consistency and calibration across assessments

Question: How can MOPAN ensure that the score/rating that it awards to organisations is accurate and well-calibrated, considering different interpretations of an indicator, and the need for coherence across assessments?

Rationale: To improve consistency across assessments. One of the key challenges in the 2015-16 cycle was consistency and calibration across the 12 assessments. MOPAN does not compare organisations; however, the shared indicator framework and scoring/rating system allow for comparisons to be made. Thus, consistency across the application of the rating system is important.

An assessment system should be *consistent* for a number of reasons, such as:

- Strengthening MOPAN's status and credibility
- Providing 'assessors' with a replicable toolkit
- Making it easier to quality assure MOPAN outputs
- Providing a shared language for strategic decisions by donors
- Encouraging common pathways for MO improvement

We understand 'calibrate' to mean how one ensures that indicators are adjusted to whatever units of measurement are appropriate. In MOPAN 3.0 there is an elaborate manual which both specifies the scoring and rating system – i.e. the composition of KPIs and MIs; and their aggregation. This is backed up by consistent staff selection and training; management and Secretariat 'validation'; and a high degree of transparency.

The preconditions for consistency and calibration therefore appear to be in place. There is no reason to believe that provided quality control measures are applied, scores and ratings are not accurate and well calibrated. It is not possible however on the basis of this 'light-touch' review to confirm that these preconditions – ranking, validation, etc. – are being fully implemented. It may be useful for MOPAN and its contractors to consider whether existing Quality Assurance arrangements are adequate.

The area where calibration can be questioned is in relation to different types of MOs. This would partly be covered by the kind of modular approach suggested above, for normative MOs particularly in relation to results. However we have been told that there are also Management MIs that do not match all MO systems. At present MIs are adapted as necessary when standard wordings pose difficulties. Once there is sufficient experience of MOPAN 3.0, it may be useful to conduct a review of these kinds of difficulties and prepare a more standard subset of alternative MIs that can be drawn on when needed.

Consistency and calibration are important criteria to apply to MOPAN and are operationalised in a way that accords with well-recognised evaluation good practice. (For example the system is well-documented; the process of implementation is transparent; assumptions built around a Theory of Change are made explicit; validation and triangulation methods are in place, etc.). At the same time there are potential dangers in assumptions made by some MOPAN members although the Methodology Manual is appropriately cautious. Two examples of this are:

- i. assumptions of objectivity and scientific precision; and
- ii. assumptions that consistency allows for comparability

- i. *Objectivity and precision:* However consistent and well calibrated, an indicator system such as that applied in MOPAN rests on reasoned human judgement that has to be applied to assess the relevance, strength and significance of available evidence. Some indicators are close to being 'facts', e.g. the existence of a gender equality policy (2.1a) or the allocation of resources as planned (4.2); but in most instances judgements have to be made. This is recognised in the scoring systems at the MI level; and in the triangulation and validation safeguards that MOPAN has put in place.

At the same time, it is true that KPIs that describe activities or functions located at MO Headquarters level and which are formally documented require less judgement than delivery, results or country-based activities. Some MOPAN members appeared more concerned about objectivity than others. Some even used the word 'scientific' to describe their expectations of MOPAN assessments.

One hypothesis that was suggested by interviews was that those who were most concerned about objectivity were also those who favoured restricting KPIs to those that minimised the need for judgement. This underlines the importance of continuing transparency and dialogue about the way judgements are made and safeguarded to maintain MOPAN member confidence in the way these judgements are made.

- ii. *Consistency and comparability:* As the initial version of the MOPAN 3.0 Methodology Manual made clear, MOPAN KPIs and the basis for judgement is 'criterion-referenced'. In the general world of education tests, personnel tests and many product tests there are three approaches:
 - *Criterion-referenced tests* which judge performance against a specified criterion. Thus in MOPAN MOs are being judged as to whether they meet best practice criteria as agreed internationally across a number of performance areas – such as strategy formation, risk management, HR policies, etc.
 - *Normative tests* which judge performance against a 'normal' distribution of all of those being tested. A school exam that allocates students into higher or lower grades is the classical case of a normative test. Individual students are compared with others taking the same exam.

- *Developmental tests*⁴ which judge a specific subject of assessment (e.g. a student) at a moment in time against a prior level of attainment achieved by the same subject previously using a similar test. Tests of student improvement over time are classic examples of a developmental test.

The decision to follow a criterion-referenced logic was partly determined by the kinds of data that could be expected in an MO setting.

‘The justification for this type of approach lies mainly in its appropriateness for the MOPAN assessment process – a criteria-referenced basis for judgement is suitable when: criteria are multi-dimensional, there is a mix of both qualitative and quantitative data, and it is not possible to calculate a simple sum of the data points. These conditions apply to the design of MOPAN 3.0.’ (MOPAN 3.0 Manual 2015-2016)

Criterion-referenced tests as noted are mainly qualitative. This also relates to the kinds of scaling and ranking that is possible and the analyses that are possible. There are four kinds of scales:

- *Nominal scales* which are not quantitative but simply label something as being different from something else – e.g. A UN Fund rather than an IFI; a house owner rather than a rental tenant.
- *Ordinal scales* are concerned with the *order* of entries along a scale in terms of more or less. (A 5-point scale Likert questionnaire scale is ordinal) without being able to say by how much different entries differ.
- *Interval scales* work when one can quantify the interval between entries but as there is no ‘zero point’. We know the difference between 30 and 40 degrees Celsius and a high and low IQ test but there is no ‘no temperature’ or ‘zero IQ’ entries.
- *Ratio scales* on the other hand do have a fixed zero point – budgeted expenditure could notionally be zero as could the number of managers in a regional office.

These differences matter because of what can be done statistically with data that conforms to different scale characteristics. For example, one can only calculate averages (means) and standard deviations with Interval and Ratio data. One cannot do so for nominal and ordinal scales. Most KPIs are built on ordinal elements – even if some rely on ‘presence’ or ‘absence’ of something – e.g. a gender equality policy – which is itself ‘nominal’. The scoring, rating and aggregation process is based on ‘more’ or ‘less’ (Highly Satisfactory through to Highly Unsatisfactory). It ends with a rating scale that is about order even if these are allocated numerical ratings. The use of averages for each MO’s rating when combining MIs is acceptable but where it does not work is in terms of comparison between MOs. As the MOPAN 3.0 Methodology Manual states: ‘MOPAN 3.0 assessments are discrete; organisations are not compared.’ At a population level (i.e. all MOs ever assessed by MOPAN) it may be possible to say: ‘There has been a discernible improvement in the quality of Operational Management across all MOs assessed in 2020 compared with 2015’ based for example on a ‘traffic light’ system. But it is not possible to say that MO X is better or worse than MO Y given that each MO will be exemplifying its own distinctive mix of characteristics. These may be structural e.g. large and small, normative or not, new or well-established.

⁴ Whilst there are no examples of ‘developmental’ testing in MOPAN we would argue that this might be something to consider in future. (See below)

They may also be contextual operating in contexts that also shape performance e.g. peaceful or post-conflict; or well-funded or subject to funding reductions.

This is not to suggest that comparisons are impossible only that it would not be methodologically sound to do so on the basis of 'raw' MOPAN KPIs and assessment scales. It would be possible by identifying clusters of MOs with shared and overlapping characteristics to compare MOs within these subsets or clusters using well-recognised comparative and configurational methods. This would move away from reliance on the kinds of scaling and ranking that characterises MOPAN 3.0 at present. (It might for example rely on configurational comparative or Bayesian methods. To do so would constitute a new 'product-line' for MOPAN).

The statement in the ToR for this Review is ambiguous about comparison:

'MOPAN does not compare organisations, however, the shared indicator framework and scoring/rating system allow for comparisons to be made.'

The first part of this sentence is uncontentious whilst the second is only correct at an aggregate level of analysis as suggested above – i.e. all MOs at a point in time have improved their management systems. MOPAN as presently constituted cannot be used to compare MOs with each other.

Reference was made earlier to 'developmental' tests. As MOPAN beds down similar reviews will be undertaken of the same MO offering opportunities to track development and improvements over time. This is anticipated in the Methodology Manual:

'This approach allows us to situate organisations along the continuum of the **'performance journey'**; organisational maturity having emerged from the Interim Document Reviews thus far as a key theme.'
(p43)

However the existing scoring and rating system does not give prominence to developmental progress. There is scope to raise the profile of this part of the report in future revisions of the MOPAN process. For example scoring and rating could be based on additional KPIs that sought to capture overall 'maturity'; or existing KPIs could be scored in terms of improvement over time. This would need to be seen as part of a larger package of possible reforms of the present MOPAN system.

For example, looking forward it is also clear that several other issues will be raised by an established cycle of regular (i.e. repeated) assessments. In particular:

- The *frequency of assessments* appears not to have yet been nailed down. As indicated in the summary of interview material some favoured shorter and some longer cycles – ranging from every 2 years through to every 5 or even 8 years. These differences are driven as much by national policy-making considerations – when national aid strategies are reviewed; and by MO replenishment cycles as well as by assumptions about assessment needs per se.
- The *scope of repeat assessments* is also not yet thought through. This will also have implications for frequency as a narrower scope would make it less likely that 'assessment fatigue' sets in among MOs. In other assessment systems once an organisation or institution has been assessed

to a certain standard a) frequency of assessments is reduced and b) further assessments focus on 'areas of potential improvement'.⁵

- The *possibilities for self-reporting* will also need to be investigated once MOs have gone through more than one assessment and regular cycles of assessment have been established. In this model a simplified self-reporting format is adopted in those areas where an organisation has been judged as performing well with effective systems in place etc. (A move in this direction has taken place in the OECD's own In-Depth Evaluation – IDE - system in relation to Policy Committees judged positively in previous IDE rounds.)

The importance of looking forward methodologically in this way is heightened by upcoming MOU and Procurement events that MOPAN faces. Several of the implications of this review, can be translated into new products or further elaboration of existing products. This could imply additional resources and to an extent this may be unavoidable. However this could also be accompanied by innovations to save resources and reduce current workload by streamlining the existing system in terms of frequency, scope and self-reporting.

3.3 Appropriateness and robustness of evidence lines

Questions

Q1: In what ways could the lines of evidence that MOPAN uses be enhanced in terms of relevance, rigour, and validity?

Q2: Are there biases in the methodology due to sampling, combination of data sources, consultations, roles of stakeholders etc.?

Q3: What are the benefits and drawbacks of weighting the evidence streams?

Q4: Are there better, feasible ways for MOPAN to assess partner/stakeholder perspectives at the country level?

Q5: How can country-level data be applied more explicitly and used more systematically in the assessments?

Rationale: To understand the strengths and weaknesses of the four lines of evidence (document review, survey, interviews and consultations); whether (and how) they should be weighted in the analysis; and how to better use country level data. This will be undertaken through the lens of ensuring suitability for achieving MOPAN's intended objectives.

Through the partner survey component, MOPAN is interested in assessing whether organisations' strategies, policies and intended business practices have permeated their ways of working. MOPAN would like to understand the robustness and utility of this information as it is currently gathered, and reflect on how different organisations' delivery at global, regional, sub-regional, national, and sub-national levels can be better assessed.

⁵ We understand that this has already been raised in relation to the World Bank but anticipate this will become a more general issue in future.

Currently data is collected from select countries, regional and global partners using an online partner survey, and documentary evidence from those countries is sourced when available. The data, for the most part, is used to triangulate overall findings, rather than informing a specific country-level component in the assessments. Feedback from stakeholders has pointed to the limited number of countries covered, the value of the data collected, etc.

MOPAN depends on four lines of evidence: Document Reviews, a Survey of country partners, Interviews and Consultations at the Headquarters of MOs. In terms of resources committed and positioning within the MOPAN process, by far the strongest line of evidence is the Document Review. Interviews and consultations do reportedly extend and even change initial ratings, but mainly by putting forward additional documents that reviewers had not initially included in their first scan.

The initial MOPAN 3.0 Methodology Manual described interviews with senior HQ staff as a 'secondary line of evidence'. Interviews are intended to 'deepen and interrogate'; 'confirm and clarify', 'update' and 'contextualise' what documentary reviews first identify.

Consultations mainly take the form of group-interviews with technical and administrative staff. They are described in the Manual as focusing on 'contextualisation/ triangulation/validation/ deepening'. From interviewee descriptions of the 2015-2016 assessment process quite large numbers of junior staff are spoken to in groups, each lasting one hour in the course of one or two days.

There are risks in assuming that senior staff are always the most knowledgeable about MO functioning and therefore consistently allocating more (interview) time to senior management and less ('consultation') time. This risk has been acknowledged in the 2nd MOPAN 3.0 cycle where there has been an effort to ensure a more appropriate balance between management and technical staff.

The Survey stream directed at MO stakeholders was widely recognised as the weakest line of evidence. Weaknesses in the 2015-2016 assessments included:

- Difficulties finding suitable respondents
- Low response rates amongst those who are sent questionnaires
- A mismatch between countries chosen for surveys and the MOs included in a particular assessment cycle

These difficulties are seen by some as connected to the abandonment of Country Lead/Facilitator role, currently reduced to a liaison support; and the reduced involvement of donors in-country compared with the Common Approach era.

Few criticisms have been voiced as to questionnaire content. At present the content of questionnaires is designed to match Micro-Indicators and key management processes, for example regarding financial management; the design of interventions; relationship management etc. There are two problems with such content:

- First, detailed questions across a range of MO planning and implementation areas makes it likely that few respondents will be able to answer everything. This could result in unanswered

questions or lead to an increase in non-response after those approached read the questionnaire and decide that most does not apply to them.

- Second, the survey takes a procedural/results chain approach. At present this appears to be acceptable to MOPAN members. Any future efforts to strengthen MOPAN as a means to improve MO performance would require greater understanding of actual results rather than intentions alone. A survey on its own is unlikely to be the best way to gather evidence of results in these circumstances, and at present only formal documentation is used to assess the results-focused KPIs 9 through 12. Past evaluations (i.e. part of the Documentary Review exercise) when they are available have been seen as a more reliable source of results evidence. However gaining an accurate picture of results (rather than evidence that upstream elements of a 'results chain' have permeated through to the national level) has proved problematic in many evaluations.

In response to the questions posed in this part of the ToR:

- Within the terms of judgement-based indicators, the MOPAN system is already *rigorous* – i.e. reliable, replicable and trustworthy. *Validity* takes different forms. *Construct validity* depends on the match between what MOPAN sets out to describe and the way this is represented by chosen indicators and scales. MOPAN has built its assessment system around a results chain logic and 'theory of change'. From the limited time this review has had to scrutinise the system, the assessment appears to be true to this conceptualisation of organisational effectiveness. It also produces '*internally*' valid assessments of MOs, i.e. accurate depictions of the MOs assessed. These assessments are not however *externally valid*⁶ which is why MO assessments cannot be compared.
- *Relevance* as already noted, pertains to purpose. The MOPAN indicator system is *relevant* to a narrow understanding of MO accountability, expressed as 'systems in place that support results' but not to a wider understanding that measured or described actual results.
- The four evidence streams already carry different *weights* in the overall assessment. This is in part by design i.e. it is intended that other evidence streams 'support' the Document Review. In part it is because of the weakness of other evidence streams, especially the Country Partner Survey.
- It is reasonable to place greater emphasis on documentary evidence provided this does not discriminate in favour of highly proceduralised MOs at the expense of those that are less procedure-intensive. It was noted by several interviewees that some newer MOs such as GAVI were likely to have fewer documented processes and policies than longer established MOs. It is also reasonable as long as MOPAN's focus on accountability continues, stopping short of assessing actual results and achievements. All assessments of relevance depends on assumed priorities and purpose.
- Notwithstanding any reasonable weighting of evidence towards documentary sources, even in a supplementary role other evidence streams need to be adequate. We do not yet know whether measures put in place to strengthen survey responses have made a difference in the 2017-18 cycle.

⁶ External validity - whether findings are applicable to other populations or settings than where they were first originated – is the foundation for generalisability claims.

- Additional ways of increasing response rates could be explored based on how countries are sampled. For example one might begin by pre-screening MOs to identify where their country-based activities e.g. programmes, national or regional offices, recent or ongoing sites of humanitarian engagement. It would then be possible instead of starting with a set list of MOs and matching countries to them; to select sets of MOs in a cycle that had in common overlapping sets of countries.
- Concerns about possible bias as noted earlier are sometimes associated with concerns about objectivity. It was suggested for example that the reason members of MOPAN and indeed Institutional Leads now have a reduced role is because of concerns that their active participation would interfere with an 'objective' assessment.
- There are several alternative or additional ways that partner/stakeholder perspectives at the country level could be assessed. One would be to detach country-level work from MO specific cycles and undertake additional studies based in countries/sets of countries/regions that spanned several MOs. A variant of this would be to focus on overlapping domain clusters – in health, humanitarian, educational or financial sectors, again with a country/regional focus.
- In the longer run a Multilateral 'Observatory', i.e. an information point or network that gathers continuous/routine data on MO performance and acts as a learning resource for accumulating and sharing good practice could also be considered.
- These and similar methodological innovations would need to be carefully designed taking advantage of face-to-face approaches (fieldwork, interviews and peer-reviews); comparative and configurational methods; and indirect tools such as online monitoring, surveys and big data/machine learning approaches⁷.

End Note

This report has set out to place the MOPAN 3.0 methodology into a broader and forward-looking framework. This framework places answers to specific questions raised by MOPAN members into a broader context.

It also signposts areas for further development and strengthening that includes possible new MOPAN products and future required competencies and skills; and identifies issues that will need to be taken into consideration in upcoming procurement decisions and decisions regarding any future MOU with OECD. Options and recommendations in these areas are taken further in material prepared for the Copenhagen Workshop, appended to this report.

Among the areas that this report has identified are:

- i. *The indicator system and its use.* There seems to be a need to elaborate the results indicators; and how customisation of indicators is applied in practice. The review has also identified the need to adapt to the SDG-agenda and that the potential of a flexible/modular approach should be explored.

⁷ For example scanning and categorising publicly available documentary sources at national as well as MO level would be well-suited to machine learning approaches.

- ii. *Flexibility and the development of differentiated products.* Such products included synthesis across assessments; results-oriented case studies; and sectoral or thematic clusters. Any differentiation of MOPAN 'products' would raise the question of workload and how this might be resourced, including how to make present activities more efficient. Other options might include focussed assessments with shorter production times. All options will require methodological development to ensure their validity and coherence with other parts of the overall MOPAN 3.0 process.
- iii. *Production time and timing.* This could imply fewer ongoing assessments with shorter production time, but also having a longer-term 'workplan' for MOPAN, which could allow members to align their own planning and funding cycles with MOPAN assessments.
- iv. *Frequency of assessments.* There is little consensus about how often assessments should be repeated. The prime driver of assessment frequency is the policy needs of donors for budget approval. Sources of variance derive both from the different member funding cycles; and different agency funding – e.g. voluntary contributions, budget periods and replenishment rounds. A key question is whether the same degree of frequency should apply for all MOs.
- v. *Use of documentation and triangulation.* Current assessments do not make full use of 'other' existing documentation and analyses. Self-reporting within an agreed framework has also been suggested as an efficient alternative to a full external assessment in some circumstances. Triangulation and quality assurance may be even more important if MOPAN moves towards greater product flexibility.
- vi. *Partner-country surveys and country engagement.* Surveys and use of country information need reconsideration. There seemed to be unresolved differences about what other sources of evidence were acceptable: e.g. country embassies and donor coordination arrangements; country own data; and case studies were amongst possibilities raised. This review also noted the possibility of a Multilateral Organisation Observatory. These and similar options will need further investigation and detailed design-work.
- vii. *The role of the Secretariat.* Any change in products, more focus on learning and engagement; and on MO system development will influence the role and activities of the Secretariat. The composition and capacity of the Secretariat will have to be considered alongside the methodological developments identified in this report. It will also have implications for the division of labour with external consultants in future.

It is important to recognise the limitations of a 'light-touch' review with tight resource and time constraints undertaken early in the life-cycle of a new assessment system, such as MOPAN 3.0. We therefore regard this report more as a means of awareness raising rather than as a kind of definitive judgement.

We are nonetheless confident that the issues and ideas discussed are sufficient to stimulate discussion and following such discussions allow for more specific decisions to be identified by MOPAN members.