UNEG
United Nations Evaluation Group

Stocktaking Study on the Utility of Peer Review

December 2021

Preface

I am pleased to preface this report summarizing the Stocktaking Study on the utility of Peer Reviews over the period 2014 – 2021 commissioned by the United Nations Evaluation Group (UNEG) Peer Review Working Group.

The peer review mechanism has evolved since it was created in 2004 as a joint initiative of UNEG and the OECD DAC Evaluation Network (EvalNet). An important milestone occurred in 2011 when a new Framework for Peer Review proposed a more streamlined approach to Peer Reviews shifting the overall responsibility and ownership of the mechanism to UNEG.

Fourteen Peer Reviews of evaluation functions of UN agencies have been successfully carried out since the last lessons-learned paper in 2013, all guided by a framework designed to be adaptable to evaluation functions of different sizes, adjusting the panel size and scope, focusing on specific issues identified jointly by the peer review panel and the evaluation function being reviewed.

The study confirms that each Peer Review has contributed in various ways to positive changes in the peer reviewed evaluation function. Peer reviews are seen as one of UNEG's most important contributions to joint learning in evaluation across the United Nations. Overall, the peer review modality can be seen to be an important vehicle for promoting the UNEG Norms and Standards and for strengthening evaluation functions across the UN system.

I am sure this rich and informative report, containing several important findings and recommendations will be a valuable resource. It will help UNEG continue to expand and refine its approaches to Peer Reviews, while setting new standards in how to make best use of the available expertise and experience both within the UNEG network as well as in the international evaluation system.

Oscar A. Garcia
UNEG Chair

Table of Contents

Acronyms

ECG	Evaluation Cooperation Group
ICAO	International Civil Aviation Organization
IFAD	International Fund for Agricultural Development
IOM	International Organization for Migration
ITC	International Trade Centre
JIU	UN Joint Inspection Unit
MOPAN	the Multilateral Organisation Performance Assessment Network
OIOS	Inspection and Evaluation Division of the Office of Internal Oversight Services
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations Children's Fund
UNITAR	United Nations Institute for Training and Research
UNODC	United Nations Office on Drugs and Crime
UNRWA	United Nations Relief and Works Agency for Palestine Refugees in the Near East
WFP	World Food Programme

Executive Summary

One of the primary aims of the United Nations Evaluation Group (UNEG) is to strengthen evaluation capacity within the United Nations system. The Group has sought to strengthen the practice and value of evaluations within the United Nations system with the development and promotion of the *UNEG Norms and Standards for Evaluation*, the development of specialised guidance on various areas of evaluation practice, including with a view – which is the focus of this study – to facilitate Peer Reviews of UN evaluation functions. Along with its Norms and Standards, the peer review modality is seen as one of UNEG's flagship initiatives.

The primary focus of this stocktaking study is to 'identify enabling and hindering factors to the utility of completed peer reviews' in order to identify lessons for future practice. It is intended to inform an updated UNEG Strategy on Peer Reviews. The study explores the utility of 12 Peer Reviews conducted between 2014 and 2021. All were undertaken using UNEG's 'second generation' guidance that was issued in 2014. Whilst they were included in the study, it is too early to fully assess the utility of the four Peer Reviews conducted in 2021, as they are too recent for action to be taken on recommendations management agreed to.

To undertake the stocktake, the consultant interviewed the individual who oversaw the evaluation function at the time of the Peer Review, the Chair of each Peer Review Panel and some OECD DAC representatives identified by EvalNet (which frequently participate as members of Peer Review Panels). The depth and quality of analysis possible was limited by the fact that some of the Peer Reviews took place eight years ago and interviewees did not have detailed recall of processes and outcomes that took place so long ago.

The study outlines the evolution of the UNEG peer review modality and sets it in the context of other ways to assess evaluation functions in the international system. This includes those by the Inspection and Evaluation Division of the Office of Internal Oversight Services (OIOS) for divisions of the UN Secretariat, the UN's Joint Inspection Unit (JIU), the Multilateral Organisation Performance Assessment Network (MOPAN) and the Evaluation Cooperation Group (ECG).

After outlining the conduct of a typical peer review, the study seeks to identify the contribution each peer review made to the evaluation functions of the peer reviewed organisations. It summarises the contribution each peer review made to the independence, credibility and utility of the evaluation function and its products. The study then looks at strengths and weaknesses of various stages of the peer review process and identifies how the latter can be addressed. The most important of these are set out below in the recommendations.

Key Findings

This study concludes that Peer Reviews do help to lift the quality and independence of evaluation functions and they do help to increase the credibility, utility and independence of evaluation products. They, therefore, do in aggregate reduce the need for donors to conduct evaluations of specific interventions that they fund. In aggregate, peer reviews help to give donors greater confidence in the performance evidence that agencies they fund produce.

There is unambiguous evidence that peer reviews have added to the credibility and utility of individual evaluation functions and their products. There is also evidence that peer reviews have had a positive impact on the independence of evaluation functions and their products. But this contribution is more varied, and impact is contingent on several factors.

All the evaluation functions reported improving the way they managed the evaluation function following a peer review. Even when an evaluation function was small and underfunded, peer reviews helped the function implement cost-neutral changes that brought them closer to embedding key UNEG standards (such as putting the organisation's evaluations in the public domain). Several evaluation functions went on to embed specific recommendations from a peer review report in their next evaluation policy.

As peer reviews are undertaken, they have a cumulative impact on the standards of evaluation in the UN system as a whole: they do enhance the professional standards of evaluation in UN agencies, as they aspire to do. The impact on the quality of evaluation and the quality of UN agency evaluation functions is significant. A 2013 study of 2005-2011 peer reviews reached the same conclusion. Thus, it could be said that Peer Reviews have consistently contributed to the strengthening of UN evaluation practice.

Peer reviews do sometimes contribute to the independence and, therefore, the institutional strength and robustness of the reviewed evaluation function. This is much more difficult task to accomplish and assessing the contribution of peer reviews to the independence and robustness of evaluation functions is more complex. Some evaluation functions have been able to use a peer review to leverage structural and financial independence – change that is transformational. The evaluation function and Panel chair have been able to use the peer review to show to the agency's management and governing bodies that the institution falls short of UNEG norms and standards and that it is in the organisation's best interests to give the evaluation function the independence and the finances it needs to deliver an effective evaluation function.

However, peer reviews rarely automatically deliver these kinds of changes. The head of the evaluation function and the panel chair need to purposefully position the peer review with the organization's management and key stakeholders. This needs to be done from the beginning to the end of the peer review process to prepare the ground for a positive reception of the findings and recommendations.

But it also must be recognised that agencies are not equally responsive to peer reviews. Agencies' overarching governance systems have a strong influence on how responsive management is. Agencies with active Executive Boards, a more direct relationship to their funders, and to whom the management is directly accountable tend to be keener to respond to peer review recommendations. It appears that agencies that are part of the UN Secretariat – with ultimate accountability through the Secretary General – may have more diluted incentives to strengthen their evaluation functions.

It is also the case that change is sometimes simply not possible even when all conditions align, and when the panel and evaluation function head have done all the 'right things'. Management may be resistant to change or scrutiny. Champions in management, who supported the peer review initially may move on.

Even when a peer review has had no impact on the independence and robustness of an evaluation function peer reviews are still worthwhile. They still provide a public analysis of how the evaluation function falls short of agreed UN standards, and what is needed to strengthen the function. They can contribute to building

pressure for change over the medium term. Peer review analyses serve as a benchmark for a next MOPAN or UNEG assessment.

Indeed, UNEG can receive help from the 'tailwind' created by MOPAN assessments. As these become increasingly mainstream and as agencies are repeatedly assessed, managers may be motivated to act on peer review findings and recommendations – including the more difficult ones relating, for example, to independence. MOPAN scores are disaggregated, and it is evident to both managers and governing bodies what elements of institutional performance drag its overall score down. Thus, managers have an incentive to ensure their next MOPAN assessment is not affected by a low-scoring evaluation function.

The report notes that there is scope for UNEG to engage more strongly with both EvalNet and (this is more tentative), the MOPAN secretariat and potentially its Member States.

Recommendations

The ToR ask this report to make recommendations to UNEG to enhance the utility of peer reviews. The report makes two recommendations under two themes, with nested actions. These are aimed at:

1.   Preserving the reputational integrity of peer reviews; and

2.   Strengthening systems, processes and partnerships to maximise utility.

The consultant is not a member of the Peer Review Working Group and therefore, several of the suggestions made are **for consideration only**. Unless otherwise specified, all the recommendations are for consideration and action by the Peer Review Working Group and potentially the broader UNEG membership. Some require collaboration with EvalNet.

1.   *Preserving the Reputational Integrity of the Peer Reviews*

1.1   Ensuring the integrity of Peer Reviews is the most important responsibility of the Peer Review Working Group. It needs to ensure there is no conflict of interest in the choice of a panel chair, panel members, and the consultant. It needs to ensure that the instrument continues to be seen as a robust assessment.

1.1 To preserve the credibility of peer reviews, the Peer Review Working Group needs to always ensure there is a visible degree of independence between the panel and the reviewed evaluation function

1.2   If UNEG allows the concept of a peer review to be devalued, UN agency managers will be able to dismiss peer reviews as a form of self-assessment that can be ignored.

1.2 To safeguard the respect and influence of peer reviews, the Peer Review Working Group should consider ensuring that the term 'peer review' refers only to exercises that involve peers (not consultants) in face-to-face engagement: the term 'peer review' should be not conflated with other modalities. (Peer reviews can of course be conducted virtually; and the inclusion of representatives of OECD DAC representatives is a separate issue).

2. Strengthening systems, processes and partnerships to maximise utility

The actions under this recommendation list specific measures that could be adopted to help maximise the utility of future peer reviews.

2.1   If peer reviews are begun only following explicit commitment from the organisation's management or governing bodies, there is likely to be greater follow up.

2.1 To secure a measure of 'buy in' from agency management, UNEG could consider starting a peer review only following an official request from senior management or governing bodies; and approval could be conditional on the agency agreeing to provide a comprehensive management response within a set period.

2.2   Delays can undermine the efficacy of peer reviews.

2.2 The Peer Review Working Group should seek to introduce systems that reduce the risk that, after a peer review has been agreed in principle, it is delayed.

2.3   The participation of OECD DAC panel members adds to the perceived credibility of peer reviews.

2.3 UNEG should continue to seek the membership of OECD DAC representatives on peer review panels as a priority.

2.4   The peer reviewed evaluation function and the panel can help build ownership of the peer review recommendations in the conduct of the assessment. The Peer Review Working Group may want to issue guidance on this.

2.4 Panels should seek to build engagement with the peer reviewed agency's management and governing bodies prior to, during, and after the mission. Reference Groups are a useful medium to help the panel engage with management in a structured way.

2.5   The Peer Review Working Group could add a step to peer reviews that may help encourage agencies to take seriously their commitments made in a management response, UNEG could consider:

2.5 Commissioning from either the peer review consultant or the reviewed agency a report on progress in implementing commitments made in the peer review management response 18 to 24 months after the peer review.

2.6   It would also be good practice to ensure that, where relevant, peer reviews assess progress made since a preceding review:

2.6 The Peer Review Working Group could add to its guidance the expectation that peer reviews will always report on the implementation of management response commitments of earlier reviews.

2.7   To continually learn from the conduct of peer reviews, the Peer Review Working Group could reinstate the request for participants to produce some form of a 'learning note' after the review. This could take the form of a simple 'Survey Monkey' questionnaire to the head of the evaluation function

and the panel chair. This would help ensure the Peer Review Working Group receives feedback and is able to adjust the instrument accordingly.

2.7 To help close the feedback loop, the Peer Review Working Group could consider introducing an uncomplicated way of to capture lessons from each peer review.

Introduction

UNEG's mission is to 'promote, strengthen and advocate for a robust, influential, independent and credible evaluation function throughout the UN system for decision-making, accountability and learning'.[1] The group has looked to strengthen the practice and value of evaluations within the United Nations system with the development and promotion of the UNEG *Norms and Standards for Evaluation*, the development of specialised guidance and, the focus of this study, facilitating peer reviews of UN evaluation functions, amongst others. Along with its Norms and Standards, peer reviews are seen as one of UNEG's flagship initiatives.

The primary focus of this stocktaking study is to 'identify enabling and hindering factors to the utility[2] of peer reviews' to identify lessons for future practice. This study also aims to identify the contribution peer reviews have made to various stakeholders and to explore how they can become more 'fit-for-purpose and useful to key stakeholders'. This stocktaking exercise is intended to inform an updated UNEG Strategy on Peer Reviews (see Annex 4 for the ToR).

The initial ToR intended that the study should cover 'second generation' peer reviews which were completed between 2014 and 2020. UNEG's peer review framework was revised in 2013 and peer reviews that used the current approach began in 2014. The first 2014 peer review (UN Women) was not included, and the study was expanded to include four additional peer reviews completed in 2021, taking the total to 12. A limitation of including the additional 2021 four peer reviews is that it is too early to fully identify their utility: it is too early to assess whether agreed recommendations have been implemented, even when management responses have been produced.

The study included only peer reviews that UNEG has categorised as such, and that are either published or are due to be published on its official website.

> The Peer Reviews included in the study are:
> - World Food Programme (WFP) 2014
> - United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA) 2015
> - International Trade Centre (ITC) 2016
> - United Nations Office on Drugs and Crime (UNODC) 2016
> - United Nations Children's Fund (UNICEF) 2017
> - International Fund for Agricultural Development (IFAD) 2019
> - United Nations Educational, Scientific and Cultural Organization (UNESCO) 2019
> - International Civil Aviation Organization (ICAO) 2020
> - World Food Programme (WFP) 2021
> - International Organization for Migration (IOM) 2021
> - United Nations High Commissioner for Refugees (UNHCR) 2021
> - United Nations Institute for Training and Research (UNITAR) 2021

---

[1] See UNEG, 2019, UNEG Principles of Working Together. P.2

[2] Utility is one of the central objectives for evaluation at the United Nations. It is defined in the Norms and Standards as being: 'manifest through its use in making relevant and timely contributions to organizational learning, informed decision-making processes and accountability for results', Norm 2, *UNEG Norms and Standards for Evaluation*.

Methodology

To undertake the stocktake, the consultant interviewed the individual who was in charge of the evaluation function at the time of the peer review (some of whom have since retired or moved to other positions), and the chair of each peer review panel. The then head of the evaluation function or the panel chair was interviewed for all peer reviews. Both were interviewed for 10 of the 12 peer reviews. Some other panel members were also interviewed. Whilst a total of 26 interviews were held, interviewees frequently had held multiple positions: i.e., some had hosted a peer review and been a panel chair or panel member. OECD DAC representatives from EvalNet were also interviewed – some of whom had taken part in peer reviews, but primarily to understand how OECD DAC members perceive the utility of peer reviews. See Annex 1 for a list of persons interviewed.

The depth and quality of analysis possible was limited by the fact that some of the peer reviews took place eight years ago and interviewees did not have detailed recall of processes and outcomes that took place so long ago – and that were often of a relatively routine nature. As such, in some ways it would be right to depict the authority of the research evidence presented as somewhat tentative/light.

The study also draws on a range of documents to identify the utility of peer reviews. These include peer review reports, management responses, Evaluation Policies and MOPAN assessments (See Annex 2 for a list of documents.)

The study also draws on the author's experience of supporting four of the target peer reviews as a consultant to the panel (UNESCO, WFP, IOM, and UNHCR) and of leading or supporting MOPAN assessments (UNESCO, UNODC, UNFPA). Where the report draws on an observation or recommendation from the author's experience, this is clearly indicated in the text.

This report represents the view of the consultant and does not necessarily represent the views of the Peer Review Working Group or UNEG.

Background to the OECD-DAC/ UNEG Peer Review Mechanism

The push for greater accountability in development aid since around 2000 led to the expansion of evaluation functions in multilateral and bilateral development organisations and in UN funds, programmes and independent agencies. Drawing on its experience of facilitating peer reviews of bilateral development agencies, the OECD-DAC evaluation network (EvalNet) and UNEG established a joint task force to pilot and support the development of a peer review mechanism for UN evaluation functions in 2004. UNDP piloted an approach to assess its own evaluation function, followed by UNICEF in 2006.

Building on the lessons from these two pilots, EvalNet and UNEG formalised a joint task force to support professional peer reviews of UN evaluation functions.[3] The task force developed guidance, specifying the focus, scope, roles, and responsibilities. The assessment framework used was flexible enough to apply to a wide range of agencies in terms of both size and mission (e.g., spanning the development and humanitarian

---

[3] DAC/UNEG Joint Task Force on Professional Peer Reviews of Evaluation Functions in Multilateral Organizations - Framework for Professional Peer Reviews.

spectrum). In this first phase OECD DAC members contributed to most of the costs of conducting peer reviews. The aims of the mechanism were to:

- Build greater knowledge and use of evaluation systems by the management, governing bodies and other stakeholders of entities reviewed.

- Evaluate the evaluators' through a professional peer assessment against recognized international standards.

- Share good practice, experience and mutual learning, building greater internal capacity and external confidence in multilateral evaluation systems and thus ultimately reducing demands for special outside assessments of performance.

Peer reviews were expected to examine the independence, credibility, and utility of evaluation systems and their products and to apply the question:

'Are the agency's evaluation function and its products: independent; credible; and useful for learning and accountability purposes, as assessed by a panel of professional evaluation peers against international standards and the evidence base.'

The mechanism was set up as a voluntary rather than a mandatory mechanism. It aimed to help evaluation offices learn from the experiences of their peers and help shape best practice in evaluation.

Second Generation Peer Reviews

In 2011, UNEG updated the peer review framework, integrating lessons learned following the completion of several peer reviews.[4] The update delivered a more streamlined framework and integrated the 2005 UNEG Norms and Standards as the normative framework for peer reviews of UN evaluation functions. All peer reviews now assess the level of adoption of UNEG Norms and Standards in line with the commitments made by UN evaluation entities as UNEG members.

The 'second generation' approach introduced the choice of applying a 'reduced framework' for small evaluation functions to 'keep costs relatively low and commensurate to the value to be gained from the peer review'.[5] The scope of assessment for this reduced modality remained the same as for other peer reviews. Of the peer review conducted between 2013 and 2021, only UNESCO and ICAO deployed the 'reduced framework' but both did so differently.[6]

The Peer Review Working Group is currently considering further modalities to make it easier for smaller evaluation functions to undergo peer reviews. These are: 'UNEG Peer Validated Self-Assessment' and

---

[4] UNEG Framework for Professional Peer Reviews of the Evaluation Function of UN organizations.
[5] The reduced approach included: a reduced number of panel members (2-4) without the recruitment of a consultant; and limited travel to one mission combining both the fact finding and peer exchanges phases. The idea of a differentiated approach for reviewing small or highly specialized agencies was mentioned in the first previous framework but not fully developed.
[6] The peer review panels of both UNESCO and ICAO panel both had a consultant and a representative from EvalNet.

'Expert Validated Self-Assessment'. UNITAR used the 'UNEG Peer Validated Self-Assessment' modality for its 2021 peer review.[7] The 'Expert Validated Self-Assessment' modality has not yet been used.

When the revised framework was introduced in 2013 the roles of EvalNet and UNEG in the conduct of peer reviews changed: UNEG took primary responsibility and ownership over the mechanism.[8] This had implications for the funding and management of peer reviews. Whilst the cost of peer reviews had been fully covered by individual EvalNet members taking part in peer panels, the cost of second-generation peer reviews has been borne mostly by the reviewed offices, by contributions from UNEG (particularly for consultants), or the panel chair's agency. Further, whilst panels were chaired by an EvalNet member, this responsibility passed to a member of UNEG.

It is probably right to say that as of 2021 there is no longer a functioning joint task force. However, except under the newest 'reduced' review modality, representatives of EvalNet members still participate in all peer reviews. Their participation is highly valued, as set out later in this report. UNEG representatives are invited to take part in EvalNet meetings and provide EvalNet members with an overview of peer review activity.

OECD DAC/ UNEG Peer Reviews in the Context of Other Assessments of UN Evaluation Functions

There are two complementary assessments within the UN that focus specifically on the performance of UN evaluation functions. The first is led by the Inspection and Evaluation Division of the Office of Internal Oversight Services (OIOS) for divisions of the UN Secretariat, and the second by the UN's Joint Inspection Unit (JIU), which has a mandate to look at the whole UN. External to the UN, donors organise assessments of multilateral agencies that include a particular focus on evaluation functions – Multilateral Organisation Performance Assessment Network (MOPAN) assessments. Finally, multilateral banks have formed the Evaluation Cooperation Group and created a peer review mechanism that mirrors that of UNEG for their members. Some UNEG members have also commissioned independent evaluations of their functions (e.g. WFO, FAO). The following paragraphs provide a short description of the different assessments:

UN Assessments

OIOS biennial assessments: these look at the evaluation function of programmes and departments of the UN secretariat, most of which are also UNEG members. The review consists of a scorecard assessment of 17 indicators on the capacity, quality and utility of the evaluation function of the departments and programmes. They describe and assess the status of evaluation in the United Nations Secretariat and identify key issues emerging from a sample of evaluations. The OIOS assessment is conducted biannually and reported to the Committee of Programme Coordination (CPC), the main subsidiary organ of the ECOSOC and the General Assembly for planning, programming and coordination. It thus provides a systematic and

---

[7] UNITAR used a consultant but did not have an EvalNet representative.
[8] UNEG, 2018, Background note on the peer review mechanism, p.5

mandated assessment of some of the UN's evaluation functions. OIOS has also conducted some inspections/reviews of evaluation functions.[9]

JIU assessments: these cover the UN secretariat and UN system entities. The JIU has a mandate to assist intergovernmental bodies in "carrying out their responsibilities for external evaluation of programmes and activities", and can "…advise organizations on their methods for internal evaluation and periodically assess these methods".[10] JIU assessments can cover the evaluation function of a specific UN agency or include it as part of a broader assessment of a single entity or in a thematic/system-wide study. Assessments conducted by the JIU thus vary in approach and frequency. Its reviews do not cover UNEG members which are not part of the JIU.

In 2013-14, the JIU conducted a comprehensive and much-quoted analysis of the evaluation functions of UN system entities covered by its mandate. The relative performance of corporate evaluation functions was assessed against a 'maturity matrix' that identified five areas and 66 indicators to benchmark against established standards endorsed by UNEG, JIU and development partners.[11] Nine of the 12 peer reviews cite this 2014 JIU report, mostly to illustrate how the evaluation function evolved since it was benchmarked in 2014. The JIU report also made a statement on the level of funding evaluation functions needed to operate effectively that has been cited in several peer review reports when recommendations on funding levels have been made.[12]

The Multilateral Organisation Performance Assessment Network and its relationship with evaluation functions and peer reviews

The primary donor-led assessment that includes a focus on the evaluation functions of UNEG members (as well as other multilaterals) is the Multilateral Organization Performance Assessment Network (MOPAN). It was established in 2002 and has revised its assessment framework several times. MOPAN identifies its mission as '…to support its members in assessing the effectiveness of the multilateral organisations that receive development and humanitarian funding..'[13]. Assessments 'identify strengths and areas for improvement' and 'Findings are used for discussions with the organisations and with their partners, and as ways to further build the organisations' capacity to be effective. Network members also use assessment findings as a source of input for strategic decision-making about their ways of engaging with the organisations, and as an information source when undertaking individual reviews'[14].

MOPAN's intention is that assessments are carried out every three to four years and that donors use these rather than conducting their own bilateral assessments of effectiveness. The network, which has a secretariat

---

[9] For example, of ITC in 2015 and UNWOMEN in 2020.

[10] A/RES/31/192, Article 5 -JIU Statute

[11] The areas assessed were: a) the enabling environment; b) relevance, responsiveness, efficiency and adaptability; c) independence / impartiality; d) quality; e) utility; and f) relevance and readiness to support United Nations Organization and system-wide reforms. Many elements have been carried over into UNEG's current self-assessment maturity framework.

[12] '…, the general view is that the range of funding between 0.5% and 3% of organizational expenditure is worth consideration depending on the mandate of the organization, the size of the organization, types of evaluations, and role of the function in institutionalization and support to strengthening decentralized evaluation, national capacities for evaluation, and partnerships in evaluation.' "Analysis of the Evaluation Function in the United Nations System." JIU/REP/2014/6. P. 29.

[13] See https://www.mopanonline.org/about/ourmission/

[14] https://www.mopanonline.org/about/whatismopan/

in OECD DAC, is composed of 20 donors who collectively agree on which multilaterals they will assess in each annual cycle of assessments. The importance that donors have given to the MOPAN network, and perceived relevance and visibility of assessments has waxed and waned in recent years. The network appears to have become more relevant in the last five years, and is likely to remain so. It has commissioned an ongoing evaluation of the influence of all assessments conducted between 2015 and 2021. This should be available in early 2022. The findings – and particularly whether reference is made of peer reviews – may be interesting for Peer Review Working Group members.

Many UNEG members have experienced at least one MOPAN assessment, and some have been through the process up to five times. MOPAN assessments are based on a standard assessment framework, interviews with the staff of multilateral organizations, an extensive stakeholder survey, and a detailed document review.

MOPAN assessments analyse four dimensions of **organizational effectiveness** (strategic management, operational management, relationship management and performance management) and provide an assessment of '**results**' (with a focus on the relevance of partners, delivery efficiency, and sustainability). The current iteration of the MOPAN methodology (3.1) embeds an assessment of an organisation's evaluation function as a facet of performance management, under 'Evidence-based planning and programming applied' key performance indicator (KPI). The methodology contains many detailed sub-questions that examine: structural independence, evaluation policy, coverage, quality of evaluation reports and stakeholder participation in the evaluation process (MOPAN annexes set out findings against each of these in detail). A summary of MOPAN's approach can be found on its website.

There is not a direct match between the frameworks used by UNEG and MOPAN for assessing the performance of UN organisations' evaluation functions. But there is a large degree of overlap particularly with respect to measuring independence. If anything, MOPAN's standards of independence are more stringent than those of UNEG.[15] Given their primary focus, peer reviews deliver a much more detailed assessment of the independence, credibility and utility of an evaluation function than MOPAN reports. In two ways there is, however, a symbiotic relationship between peer reviews and MOPAN assessments:

First, MOPAN assessments and UNEG peer reviews tend to ground their assessments of the evaluation function in the findings of preceding UNEG or MOPAN assessments. For example, the 2019 UNODC MOPAN assessment refers to findings of the 2016 UNODC peer review in its report. The MOPAN assessment amplified the findings made in the peer review and highlighted the fact that there had been no improvement in the independence of the UNODC evaluation function after the peer review. Likewise, the

---

[15] To score highly on independence, the MOPAN assessment requires that the head of the evaluation function reports to governing bodies rather than management. This is rarely the case for UNEG members. The MOPAN Micro indicator 8.1 asks the following questions, to assess whether "A corporate independent evaluation function exists ". Each 'Element' is scored from 1 to 4.

Element 1: The evaluation function is independent from other management functions such as planning and managing development assistance (operational independence)

Element 2: The Head of evaluation reports directly to the Governing Body of the organisation (Structural independence)

Element 3: The evaluation office has full discretion in deciding the evaluation programme

Element 4: A separate budget line (approved by the Governing Body) ensures budgetary independence

Element 5: The central evaluation programme is fully funded by core funds

Element 6: Evaluations are submitted directly for consideration at the appropriate level of decision- making pertaining to the subject of evaluation

2021 IOM UNEG peer review refers extensively to the MOPAN 2019 assessment of IOM's evaluation function. The 2021 peer review explicitly assessed whether the challenges highlighted in the MOPAN assessment had been addressed.[16]

Second, the entire MOPAN assessment of 'results' necessarily rests on evaluative evidence. The results section of the MOPAN report, with its KPIs on the relevance of partners, delivery efficiency, and sustainability, relies on evidence gleaned from the organisation's evaluations. Therefore, MOPAN members/donors have a direct incentive in ensuring that organisations' evaluation functions produce adequate and credible evaluations to inform results assessments. This is a linkage that UNEG can consider exploring. Peer reviews help strengthen evaluation functions both in their own right but also so that they provide credible evidence of performance (the accountability function). In its narrative, UNEG could highlight the relevance of its peer review work to MOPAN donors.

The Evaluation Cooperation Group (ECG)

There is also a 'sister' group to UNEG, the Evaluation Cooperation Group (ECG). It plays a comparable role to UNEG but in relation to its multilateral development bank (MDBs) members rather than UN agencies. The ECG was established in 1996 by the heads of the evaluation departments of MDBs 'in response to a call in 1996 of an OECD DAC Task Force for harmonization of evaluation methodologies, performance indicators, and criteria'. The ECG's mandate focuses on 'harmonization of evaluation principles, standards and practices and facilitating the involvement of borrowing member countries in evaluation'[17]. Like UNEG it created a peer review mechanism and has a review framework.

Several UNEG members have served on panels for ECG peer reviews. One interviewee who has been involved in both UNEG and ECG peer reviews considered that they are similar but that UNEG peer reviews give more emphasis to learning than ECG ones, whilst ECG reports give more emphasis to assessing the quality of evaluation products. This interviewee noted that if UNEG and ECG combined as one unit, peer reviews would be able to draw on a larger pool of potential panel chairs (identifying chairs has sometimes been a constraint for UNEG, and ECG can only draw from 10 evaluation function heads). ECG peer reviews rely heavily on consultants, as do most UNEG peer reviews.

Organising and Conducting a Peer Review

To provide context for the next section, which reviews the conduct and utility of the 12 peer reviews, this section outlines what a typical peer review process entails – and indicates some of the variations. Readers who are familiar with peer reviews can skip this section.

---

[16] 2019 MOPAN 2017-18 Performance Assessments: International Organization for Migration.

[17] Its membership comprises the African Development Bank, the Asian Development Bank, the European Bank for Reconstruction and Development, the Inter-American Development Bank and the World Bank. Evaluation Cooperation Group, 2009, 'Review Framework for the Evaluation Function in Multilateral Development Banks', p.1. This document also sets out ECG's peer review framework. The ECG currently has ten members and three permanent observers. IFAD is a member and GEF are observers, as are UNEG and the OECD.

Preparation

The head of an evaluation function contacts the chair of the Peer Review Working Group and expresses interest in receiving a peer review. The evaluation function may want a stocktake of their evaluation practice; they may also want input into a new Evaluation Policy. Some agencies evaluation policies stipulate that a UNEG peer review will be undertaken to inform the subsequent policies. The Peer Review Working Group and the evaluation function together draft ToR and agree the skill sets/experience needed on the panel. Who identifies panel members varies, with some agencies finding members themselves. UNEG asks EvalNet to find an OECD DAC panel member, preferably from among the reviewed UN agency's main donors (the 'reduced modality' does not include an OECD DAC member among the Panel). The Panel chair is responsible for recruiting a consultant to support the panel (if used). Consultants have primarily been paid by the reviewed agency, the agency of the panel chair, or by UNEG. Smaller evaluation functions have three panel members, larger ones have had up to six (consultants are not considered members of the panel). The process of finding a chair and panel members can take anything from a couple of months to – when volunteers are either not forthcoming or available – up to a year.

The Peer Review Mission

The panel chair and evaluation function agree ToR, thereby launching the peer review. The evaluation function will then complete a detailed self-assessment (currently framed around 48 questions) and share it with the chair.

If a consultant is used the consultant will be responsible for drafting a preliminary assessment of the evaluation function broadly in terms of its independence, credibility and utility: the purpose of this is to orientate the panel for its HQ meetings. The consultant will draw on interviews with (mostly) technical staff (either a mission or remote meetings), analyse documentary evidence, and will submit the Preliminary Assessment a few weeks before the mission. Sometimes the evaluation function reviews and provides comments on the Preliminary Assessment and sometimes it is not shared (these documents can vary from 25 to 100 pages, depending on the complexity of the evaluation function).

With the help of the consultant, the panel chair identifies who the panel wishes to meet and a mission schedule with between 15 and 30 meetings is agreed. Missions seek to meet senior management, sometimes including managers of regional units, evaluation staff, and often representatives of the agency's governing bodies. The panel also generally seeks to schedule a 'learning exchange event' whereby the evaluation function and panel members explore issues of concern to the receiving agency outside the parameters of the peer review.

Follow up

At the end of the mission the panel will generally feed preliminary findings back to the evaluation function – and sometimes also verbally to a Reference Group if one has been established for the peer review. Depending on the peer review, the consultant will work under the guidance of the panel to produce a consolidated report. Panel chairs and members have sometimes drafted the reports themselves – and have had to when a consultant is not recruited - but this is less common.

The report then goes through a review and agreement process that differs by agency. In the majority of cases the draft report is agreed between the panel and the evaluation function. When a Reference Group has been established it can also provide comments on the report. Reports vary in length and detail, depending on the size and complexity of the evaluation function, but all explicitly address the independence, credibility and utility of the evaluation function. All reports bring together key findings and make a series of recommendations to the evaluation unit, management and governing bodies.

After the report is agreed the Panel chair sends it officially to the organisation – it has also sometimes been shared with the organisation by the UNEG Chair. Some evaluation functions submit the peer review report officially to their governing bodies – giving the report the status of an officially endorsed document. Likewise, some panel chairs present the report officially in a verbal presentation to the governing bodies and/or Member States.

The evaluation function is then involved in facilitating the creation of a management response, if one is produced. There is no further official follow-up to a peer review.

The guidance for a Peer Review asks agencies to write up a 'learning note' and share it with UNEG after a Peer Review is completed. Only a small number have been completed and this request has largely been overlooked – especially recently.[18]

Characteristics of the Peer Review

The following table sets out some basic features related to the evaluation functions and the nature of the peer reviews.

Agency and Year of Peer Review	Year	Evaluation function comparative size (approx.)[19]	Framework applied (as reported by orgs)	Panel size	No. of Consultants	Number of missions
WFP	2014	Large	Full	3	2	2
UNRWA	2015	Small	Reduced	3	1	1
ITC	2016	Small	Reduced	3	1	1
UNODC	2016	Small	Reduced	3	1	2
UNICEF	2017	Full	5	1	1	
IFAD	2019	Large	Full	3	2	1
UNESCO	2019	Medium	Reduced	3	1	1

[18] The 2018 study refers to a few Lessons Learned notes that had been completed to early 2018 but they could not be located on UNEG's document repository and were not reviewed for this study. The 2018 study notes that five out of a potential 11 learning notes were produced between 2011 and 2018 (p.9).

[19] This is a rough categorisation. The number of permanent professional posts at the time of the peer review has been used as a proxy for size: 'small' functions are defined as having 1-5, 'medium' 6-12, 'large' 12+ permanent professional posts.

Agency and Year of Peer Review	Year	Evaluation function comparative size (approx.)[19]	Framework applied (as reported by orgs)	Panel size	No. of Consultants	Number of missions
ICAO	2020	Small	Reduced	3	0	1
WFP	2021	Large	Full	6	1	1 (remote)
IOM	2021	Small	Full	3	1	1 (remote)
UNHCR	2021	Medium	Full	6	1	1 (remote)
UNITAR	2021	Small	Peer Validated Self-Assessment	2 (no OECD DAC representative)	1	1

Prior Analyses of the Utility of Peer Reviews

While there has never been a focused study of the utility of peer reviews, UNEG and the Peer Review Working Group commissioned two studies that contribute some analysis: a 2013 lessons learned study and a 2018 'background note'.[20]

The 2013 study of seven peer reviews conducted between 2005-2011 aimed to examine whether (a) the reviews made a difference in the use made of evaluations, in particular by donors; and (b) whether these reviews had a positive impact on the evaluation departments of the agencies reviewed. The study concluded that the peer reviews had a 'direct positive impact on reviewed evaluation functions as well as on their organizations, at both levels of governance and of management'. It also concluded they strengthened the evaluation community - including bilateral donors as well as UN agencies. The peer review approach was judged to be contributing to quality evaluations and, through that, to improved organizational performance.

The study concluded it could not judge whether donors had started to rely on evaluations of peer reviewed evaluation departments: it could not say if they had reduced their own evaluations of UN agencies. However, it noted a positive effect on donor and stakeholders' perception of the value and credibility of evaluations undertaken by peer reviewed evaluation offices, as well as the accountability of their respective organizations. Based on this, the study judged that peer reviews constitute **'a necessary but insufficient building block for increased and full reliance on UN agency evaluations by donors and stakeholders'**. The report made a set of recommendations and UNEG produced a management response.

The authors of the 2018 Peer Review Working Group-commissioned 'Background Note' carried out an online survey in early 2018. It was sent to 47 UNEG members and 35 responded. Evaluation function heads were also asked what benefits they valued. They cited:

---

[20] UNEG, 2013, Report of Lessons Learned Study Peer Reviews of UNEG Evaluation Functions, UNEG, 2018, Background note on the UNEG peer review mechanism.

- Obtaining a comparison to norms and standards and best practice allowing them to identify areas for improvement

- Strengthening the credibility of the evaluation function within organizations

- Increasing the utilization of evaluation products by senior management

The 2018 study began to ask about impact – but only to a limited extent due its reliance on an online survey. It asked survey respondents 'To what extent were the recommendations issued through the Peer Review Implemented'. Eleven evaluation function heads responded. Seven said 'to a great extent, two said 'somewhat and two said 'not at all'. The responses do not, however, give any indication of the significance of the recommendations that were/were not implemented.

The 2018 survey report does make two observations on whether changes resulted from the Peer Review. It states that:

- Responses indicated that evaluation policies and methods were strengthened and that more emphasis was placed on management responses and recommendation follow-up.

- In the case where recommendations were not implemented, the factors noted included the non-binding nature of recommendations and the absence of a follow-up mechanism.

Importantly, the study notes that 'The follow up and uptake in the formal decision making of offices peer reviewed, and impact of peer reviews in the organizations reviewed would require more in-depth analysis'[21]. This is what the current study has been commissioned to do.

The influence of 2014-2021 Peer Reviews on the independence, credibility and utility of evaluation functions

This section of the report gives an overview of some of the key features of each of the 12 peer reviews and seeks to summarise the contribution each peer review made to the evaluation function. As indicated above all peer reviews are required to assess the independence, credibility and utility of the evaluation function and its products. Peer reviews are intended to give recommendations to strengthen these aspects of an evaluation function. It is logical, therefore, to assess whether peer reviews have an impact on independence, credibility and utility. This section seeks to identify the impact each peer review has had in terms of two parameters:

a. The extent to which the peer review helped the evaluation function strengthen the credibility and utility of evaluations and the function as a whole; and

b. The extent to which the peer review helped strengthen the structural and financial independence of the evaluation function.

---

[21] UNEG, 2018, Background note on the UNEG peer review mechanism, p.9.

c. The summaries and descriptions of key peer review features are, by necessity, reductions and cannot tell the full story. They reflect information from interviews of panel members and staff of the reviewed evaluation functions, and draw on secondary sources (the peer review, MOPAN assessments etc.).

The assessments relating to the four peer reviews conducted in 2021 are more tentative than the others and ratings were not given to them. The full utility and impact of the 2021 peer reviews cannot be judged because they were carried out very recently and there has not been adequate time to implement recommendations, even when a management response has been agreed. The four 2021 peer reviews have a blue backdrop in the table below, to highlight this.

All peer reviews used the 'standard methodology' except UNITAR which used the 'Peer Validated Self-Assessment': there was not an OECD DAC member on the panel.

Explanatory note on the red, amber, green and grey 'gradings' in the table

Red, amber and green (RAG) ratings have been given to each of the two sets of criteria (with the colour indicated by an initial to aid accessibility). Grey' shaded boxes against some of the 2021 peer reviews indicate 'not yet known' as discussions on the agency's response to the peer review were ongoing at the time of writing.

The RAG ratings given show the following:

Impact on Credibility & Utility of the evaluation function (upper box)	
R	No changes (can be due to lack of resources) [no peer reviews have a red rating]
A	Some developments/changes made in the evaluation function following the peer review
G	Significant changes made in the evaluation function following the peer review
Impact on the Independence (both structural and financial) of the evaluation function (lower box)	
R	No impact - no substantive change in the independence of evaluation function
A	Medium level impact - some substantive changes in the independence of evaluation function but that do not amount to a step change/transformative change
G	High level of impact: marks a step-change/transformative improvement in the independence of the evaluation function

Impact of peer reviews on the professionalism and independence of evaluation functions

	Summary & Context and conduct of Peer Review	**Impact on evaluation function in terms of:** **Credibility & Utility (upper box)** **Independence (lower box)**	
WFP **(2014)**	**Summary**: *High utility and impact. The peer review was seen to have 'helped evaluation in the organisation.'* The peer review was mandated by WFP's existing Evaluation Policy. The evaluation function's objectives were: • to get WFP to establish a much stronger policy on resourcing. • for the peer review to inform the establishment of a decentralized evaluation function (none existed). The evaluation function established a Reference Group for the peer review. The panel met representatives of the Executive Board during the mission and following it the Chair briefed Executive Board on key findings. The evaluation head ensured the report was widely shared and endorsed by management and the Executive Board. This was WFP's 2nd peer review.	The peer review mandated an entirely new decentralized evaluation system where none existed. It fed directly into a new Evaluation Policy and Strategy, and broad range of initiatives that lifted standards across the board.	G
		The evaluation unit obtained commitment to a new resourcing formula to allocate 0.8% of expenditure for evaluation: this was seen as a major achievement. The evaluation unit's wish to establish a decentralized evaluation function was endorsed in the face of middle-management reluctance: the peer review was pivotal in this and built on the OIOS 2014 study that criticised a UN-wide weakness in decentralized evaluation systems. The 2021 WFP peer review assessed implementation of the 15 recommendations made in 2014: all except one had been implemented.[22]	G
UNRWA **(2015)**	**Summary**: *The peer review had little to no sustained impact on the evaluation function as a whole.* The peer review was intended to familiarise management with UNEG standards with what a robust evaluation function needed to look like (especially relative to UNRWA's expenditure). It was UNRWA's first peer review. Prior to the peer review the UNRWA evaluation function had little financial independence. Its structural independence was and continues to be adequate. One member of senior management was a champion for the peer review and was keen to ensure UNRWA operated within UNEG standards but left UNRWA before the panel's HQ mission. The case for UNRWA undergoing a peer review was helped by discussions at the time in UNEG on	The evaluation function implemented as many cost-neutral Peer Review recommendations it was able to – given human and financial resources. However, it still lacks the funds to implement some recommendations – such as conducting strategic evaluations with its own resources.	A
		A management response was produced but there was no follow up action from UNRWA's management at that time. UNRWA's Oversight Committee repeated the Peer Review recommendations, but this too did not lead to action. The funding and size of function remained largely unchanged in the two years following the peer review – undermining its ability to commission or undertake independent evaluations. The UNRWA MOPAN report and a recent Medium Term Strategy evaluation both noted that little of	R

---

[22] WFP UNEG/OECD DAC (2021) Peer Review of the Evaluation Function at the World Food Programme, Annex 3.

	Summary & Context and conduct of Peer Review	Impact on evaluation function in terms of: Credibility & Utility (upper box) Independence (lower box)	
	making peer reviews mandatory. UNRWA's management was persuaded to hold a peer review by the likelihood that they would be mandatory. The Peer Review Panel met senior management and Member States from the Executive Committee. It had adequate engagement during the mission. The report was presented to management	what the peer review had recommended had been implemented.[23]	(red)
ITC (2016)	**Summary**: *High utility and high impact, and led to sustained changes in a small evaluation function.* The peer review facilitated financial independence with the allocation of a regular budget for the evaluation function for the first time; it led to some partial changes that enhanced the independence of the evaluation function. The evaluation function requested a peer review to: familiarise management with UNEG standards particularly around operational and financial independence; and to obtain an external perspective on the function. The Executive Director at the time was 'fully on board'. Prior to the peer review the agency lacked a meaningful evaluation culture and management was unfamiliar with the principle of independence (this was its 1st peer review). The evaluation function appreciated the report, and considered it was high quality. The participation of the OECD DAC member of the panel was said to be very important in encouraging management to engage fully (the member represented a major donor).	The peer review helped inform the agency's first Evaluation Policy and integrate UNEG norms and standards in an appropriate way.	G
		The peer review report recommended that the position of the head of the evaluation function report directly to the Executive Director (enhancing independence) and the seniority of the head of evaluation be increased. It also recommended increasing the predictability of resources. The organisation's management response accepted all the peer review recommendations except one relating to the reporting line of the head of the evaluation function (though a 'dotted' reporting line to the Executive Director was agreed). In line with a peer review recommendation – but around two years after the review – the evaluation function was allocated a regular budget for the first time. Management now largely respects the principle of independence of the evaluation function (though some ambiguity around reporting lines persists). The evaluation function considers that the peer review 'enhanced the status of evaluation in ITC'.	G
UNODC (2016)	**Summary:** *Little to no impact on the evaluation function as a whole.*	The evaluation function appreciated the tools/methodology and advice of the peer review. The evaluation function noted that it has used the report in numerous ways, and appreciated the insight generated by the peer review. It has	A

---

[23] See MOPAN, 2019, '2017-18 MOPAN Assessment of UNRWA' and UNRWA 2021, Evaluation of UNRWA's Medium Term Strategy.

	Summary & Context and conduct of Peer Review	Impact on evaluation function in terms of: Credibility & Utility (upper box) Independence (lower box)	
	This was the first peer review for a relatively fragile evaluation function. Several years before the peer review UNODC downgraded the evaluation function. The function was reinstated with a regular budget after the engagement of Member States. A peer review was a way of taking stock of the re-established evaluation function and building familiarity within the organisation with UNEG standards. There was a delay of around 9-10 months in initiating the mission due to difficulties identifying an OECD DAC panel member. Senior management were said to have seen the peer review as a self-assessment rather than a formal external exercise. This perception was based on the fact that UNEG is a network rather than a professional body. The final report was not discussed by Member States in a Board-level venue, and management engagement with the report was not strong.	introduced cost-neutral changes – those that were feasible within its existing resources. But finance and staff have remained key constraints. The Evaluation Policy has not been revised since the peer review.	O
		A management response was not produced. It was explained that a request for a management response would have been taken more seriously by UNODC's management if it had come from a body with the authority of a professional body. The peer review concluded that the evaluation section was not sufficiently operationally, structurally or behaviourally independent. Budgetary independence is an ongoing issue, and the lack of budgetary autonomy had negative impacts on the independence of the function. A MOPAN assessment, conducted two years later, concluded that very little had changed in the governance of the function: the independence and resourcing challenges the evaluation function faced remain.[24]	R
UNICEF (July 2017)	**Summary**: *Significant, but delayed, influence on both the management of the evaluation function and financial independence*. The peer review was mandated in UNICEF's 2013 Evaluation Policy. The peer review was seen to provide external scrutiny of implementation of the Evaluation Policy; and of trying to ensure there was management buy-in for implementation. This was the agency's 2nd peer review. UNICEF has an influential Executive Board – to which management is responsive. Difficulties securing a panel chair and members led to significant delays to the start of the peer review. The delay meant that the head of the evaluation function changed immediately after the report was issued – leaving a gap in follow up.	Some significant changes were implemented following the peer review, but there was a delay in implementation due to a change in leadership of the evaluation function. The peer review made specific recommendations on the future shape of the evaluation function, including on how the evaluation policy needed to better define the accountability framework for evaluation and on its professionalization.	G
		There was initial management reluctance to agree many of the peer review recommendations, but management eventually agreed to most significant ones and they were fully implemented (except one relating to financing of the decentralized evaluation function). Management agreed to the establishment of an independent and decentralised evaluation function and to giving the head of evaluation greater authority and resources. The panel chair and the OECD DAC panel member both engaged with management and the Executive Board after the peer review to encourage follow up.	G

---

[24] UNODC MOPAN assessment, 2019, p.128.

	Summary & Context and conduct of Peer Review	Impact on evaluation function in terms of: Credibility & Utility (upper box) Independence (lower box)	
		The OECD DAC panel member thus played an important role.	 (green)
IFAD (2019)	**Summary:** *A 'health check' peer review of an evaluation function that already met UNEG standards.* The 2019 peer review used a combination of UNEG and ECG standards (IFAD is an IFI and a UNEG member). The 2019 peer review constituted a 'health check' (as was the WFP 2021 review, see below) and was IFAD's second peer review. Its 2009 peer review led to revisions in the Evaluation Policy and resulted in much greater independence of the evaluation function: it established its functional independence, strengthened the role of a Board-level Evaluation Committee and established an independent budget process. The 2019 peer review therefore assessed an organisation in which the major independence and governance issues had already been tackled. The Executive Board is routinely very engaged with evaluation and the peer review was presented to the Executive Board. The report was seen as useful and the evaluation function is currently still implementing the recommendations. Ratings of 'medium' impact have been given in both boxes because the peer review resulted in modest changes to existing strong practices.	The peer review recommendations and findings helped to inform IFAD's new Evaluation Policy. The majority of peer review recommendations were related to strengthening its internal systems – e.g. simplifying internal processes, revising its products mix, and updating its evaluation manual.	A
		IFAD already had one of the most independent evaluation functions: the Executive Board appoints the head of evaluation. It is also one of the best resourced evaluation functions: it receives 0.5% of organisational expenditure. This peer review therefore did not need to make any significant governance-related recommendations. The most significant recommendations related to revising the ToR of the existing Evaluation Committee and enabling the evaluation function to use its budget allocation more flexibly.	A
UNESCO (2019)	**Summary**: *A reasonably impactful first peer review: it encouraged the function to develop stronger decentralised evaluation systems and steered the organisation to give the central evaluation function more financial independence.* This first peer review was requested to help inform a new Evaluation Policy, and followed a positive assessment of the evaluation function in a preceding MOPAN assessment.	The peer review is reported to be influencing the design of the forthcoming 2022 Evaluation Policy. Specifically, the new Evaluation Policy will include a new comprehensive framework for decentralized evaluation – which the evaluation function previously lacked and which the peer review recommended was developed.	A
		UNESCO's management response was delayed and some follow up actions were not taken due to a change in leadership and personnel changes. The peer review has contributed to the financial independence of the evaluation function – although	G

	Summary & Context and conduct of Peer Review	Impact on evaluation function in terms of: Credibility & Utility (upper box) Independence (lower box)	
	The independence of the function was relatively strong prior to the peer review but the way finance was accessed limited the function's ability to select evaluands. The decentralized evaluation system was also somewhat under-developed. Internally, the peer review was seen to broadly endorse the existing approach and did require radical change. The panel engaged with Executive Board members during the mission but did not present to the Board after the review.

Implementation of agreed recommendations faltered slightly as the Head of UNESCO's Internal Oversight Service (in which evaluation lies) moved post shortly after the peer review and the leadership of the evaluation function changed at the same time. However, two years after the peer review, the main recommendation relating to financial independence were implemented, allowing the evaluation function to directly manage resources needed to commission/manage evaluations. | two years after the review. The peer review recommended that the evaluation function have full discretion on how to programme funds made available for evaluation be spent. These changes will be introduced in January 2022 and will allow the evaluation function to commission cross-cutting and/or strategic evaluations without requiring spending clearance from programme sectors.

A panel recommendation to enhance scrutiny of independent supervisory committee was also agreed by management. | |
| ICAO (2020) | **Summary**: *The peer review had a moderate degree of impact on a very small evaluation function: it helped inform management of UNEG standards and did not deliver substantive changes in governance of the function – noting that the scope for this was limited.*

ICAO's evaluation function is very small - with one specialist working within an audit/oversight entity. The panel applied the 'reduced framework' approach and did not use a consultant.

The evaluation function perceive the report as very helpful in providing recommendations that framed ICAOs subsequent Evaluation Policy. The panel met the Executive Director and the final report was presented to the Governing Board. The evaluation function staff were somewhat concerned that the full scope of a peer review was applied to ICAO despite its size. The Panel rather than a consultant wrote the report. This was ICAO's first peer review. | The peer review directly helped with the development of a new Evaluation Policy: it is now in line with UNEG norms and standards and the function has integrated changes recommended by the review.

The evaluation unit valued the report's findings related to the enabling environment and creating awareness of evaluation in the organisation. Evaluations are now in the public domain. | A |
| | | A management response agreed to all recommendations except one which was agreed partially. There is still no independent budget for evaluation – but this was explained as a consequence of ICAO's budgeting framework. The review was seen as helpful in making management and governing bodies aware of what evaluation standards are – and how other organisations manage their evaluation functions: 'They heard how evaluation is handled by other organisations.' | A |

	Summary & Context and conduct of Peer Review	Impact on evaluation function in terms of: Credibility & Utility (upper box) Independence (lower box)	
WFP 2021	**Summary**: *A third peer review of a well-established evaluation function that gave the function the equivalent of a 'health check'.* The peer review was mandated in WFP's Evaluation Policy. It was not expected to address any fundamental governance issues or challenges. The peer review was not transformational and could be equated to an institutional 'health check' or stocktake. The implementation of recommendations made in the 2014 peer review meant the function had already addressed more difficult challenges. The 2014 peer review had helped it secure adequate resources and obtain a mandate to build a robust decentralized evaluation function. Among other issues, the evaluation unit was keen to receive the panel's views on how the decentralized evaluation function was evolving and how it was 'meshing' with the rest of the evaluation function. The peer review engaged with senior management through WFP's existing 'Evaluation Function Steering Group', which served as a Reference Group. This group was created following the 2014 peer review. The panel met with Executive Board members during the mission and briefed them afterwards informally and formally. The entire mission was conducted remotely due to COVID-19.	WFP produced a particularly detailed self-assessment. In appropriate sections it included a self-assessment of the decentralized evaluation function completed by regional evaluation officers. The evaluation office found the self-assessment valuable – and it helped it identify strengths and weaknesses before the panel's mission. The panel validated the self-assessment. The review was seen to have helped staff view the evaluation function in a more balanced way and brought 'further insight'. Partly as a result of completing the self-assessment, the evaluation unit felt it was already aware of many of the issues the panel identified. The evaluation unit is in the process of developing a new Evaluation Policy with the intention of ensuring it reflects peer review recommendations.	
		The peer review made relatively minor recommendations on governance-related issues as its operational and financial independence were already strong (e.g. the independence of function was fully respected; and funding had doubled since 2015). At the time of writing this report, WFP had mapped out a comprehensive and collaborative process to deliver an agreed management response, but it was still under preparation.	
IOM 2021	**Summary**: *It is too early to judge the utility of IOM's first peer review, as IOM's management response was issued in November 2021. Some key recommendations were agreed whilst responses to others relating to independence and financing are more tentative – noting that discussions will take place. The report and management response will be considered by IOM's governing body, and a new Evaluation Policy will be developed to embed the peer review findings.*	The management response indicates full agreement to a series of significant changes that should, over time, enhance the credibility and utility of evaluation in IOM. Some initiatives had already been launched as of December 2021. IOM is developing a new Evaluation Policy with the aim of presenting it to governing bodies in spring 2022. The evaluation function intends to integrate into the revised Evaluation Policy the recommendations made in the peer review report.	
		IOM's management response to the peer review indicates that it is less clear whether there will be	

	Summary & Context and conduct of Peer Review	Impact on evaluation function in terms of: Credibility & Utility (upper box) Independence (lower box)	
	The peer review was requested by IOM's Deputy Director General to follow up on a MOPAN assessment that was critical of the resourcing and status of the evaluation function, and to benchmark the evaluation function with UNEG norms and standards. IOM joined the UN only in 2016 and this was the organisation's first peer review. The unit is small (three permanent officers with combined responsibility for monitoring as well as evaluation). The peer review was delayed as it took many months to secure a Panel Chair. This delay had practical consequences: the peer review had been scheduled to feed into IOM deliberations on internal governance – with direct bearing on the evaluation function. By the time the peer review report was ready in May 2021 IOM's management had already reached a number of provisional decisions. The peer review report was presented to the IOM Governing Body (SCPF) in October. The OECD DAC panel member engaged with her government's representative to the Governing Body to encourage demand for a substantive discussion on the peer review and the management response. The mission was conducted remotely due to COVID-19.	progress on issues related to the structural independence and financial independence of the evaluation function. Contrary to a peer review recommendation, the evaluation function will be moved out of the (more independent) Department of the Inspector General and into an administrative department – albeit one that reports directly to the Director General. And it is likely to continue to hold responsibility for monitoring. There is commitment to discuss and review financing but no concrete commitment.	
UNHCR 2021	**Summary**: *as with the IOM peer review, it is too early to assess the degree of utility of the peer review, particularly on issues that require management authorisation. At the time of writing management was still considering its response.* The peer review was mandated in the agency's 2016 Evaluation Policy, which established a standalone evaluation function in UNHCR for the first time. The peer review is intended to inform the agency's next policy and was requested to take stock of how the function had evolved since 2016 and to help it chart its future strategy. The evaluation function also wanted management to hear an external perspective on how the evaluation function is	The peer review findings and recommendations will inform UNHCR's next Evaluation Policy and was seen as useful. The report contained recommendations that relate the how the decentralized evaluation system is structured.	
		At the time of writing the organisation was still defining its management response and discussions were ongoing – particularly on recommendations related to resource allocation.	

	Summary & Context and conduct of Peer Review	Impact on evaluation function in terms of: Credibility & Utility (upper box) Independence (lower box)	
	resourced: in terms of both volume and predictability. A recent MOPAN assessment had been critical of the independence of the evaluation function. The panel was unusual insofar as, in addition to an OECD DAC representative, it also had two representatives from humanitarian bodies (the 'Active Learning Network for Accountability and Performance in Humanitarian Action [ALNAP] and the International Committee of the Red Cross in Geneva. [ICRC]), a representative from OIOS, and a second representative of a major donor. The panel met both senior management and representatives of UNHCR's governing bodies – its Executive Committee. UNHCR established a Reference Group for the peer review. At the time of writing, the final peer report had not yet been discussed by governing bodies but a request for it to be discussed was made. The mission was conducted remotely due to COVID-19		
UNITAR 2021	**Summary**: *First peer review using the new 'Peer Validated Self-Assessment' modality – a smaller team without a representative of OECD DAC.* UNITAR has a very small evaluation function in a multi-function unit (two staff with three functions). UNITAR was motivated to request a peer review to inform its first standalone Evaluation Policy (its previous policy covered three functions). UNITAR did not have an OECD DAC member on the Panel, but a consultant supported the panel. The peer review report has been reviewed by the Evaluation Advisory Board and was presented to its Executive Board.	The panel's recommendations focus on developing the evaluation function – with a range of specific measures that should enhance the function. The evaluation function valued the learning opportunity highly and valued the self-assessment exercise. The panel reviewed UNITAR's suggested changes to its prior Monitoring and Evaluation Policy Framework. This has now been replaced by the agency's first standalone Evaluation Policy, which was informed by the peer review panel's contributions.	
		The peer review report does not make recommendations related to governance. However, the Panel discussed separating monitoring responsibility from the evaluation function and this proposal was discussed by the Evaluation Advisory Board. UNITAR has sent the management response to UNEG and started to implement recommendations. The evaluation unit felt that the	

	Summary & Context and conduct of Peer Review	Impact on evaluation function in terms of: Credibility & Utility (upper box) Independence (lower box)	
		peer review accorded the function greater recognition within UNITAR and valued this.	

The box below summarises and comments on the distribution of the ratings given to each peer review.

Number of peer reviews		Comment
Impact on Credibility & Utility of the evaluation function		
R - 0	No change all	All evaluation functions reported strengthening their management of their evaluation functions following a peer review – even if modestly. Hence, no agencies recorded 'no change'.
A- 5	Some developments/changes made in the evaluation function following the peer review	Most (5 of 8) introduced some changes, even if the peer review delivered nothing in terms of strengthening the independence of the function. Note that two of the five in this category did not make 'significant changes' because their practices were already strong (IFAD, UNESCO).
G- 3	Significant changes made in the evaluation function following the peer review	Three used the peer review to facilitate significant change – including mandating entirely new decentralized evaluation function.

Impact on the Independence (both structural and financial) of the evaluation function

R - 2	No impact - no substantive change in the independence of evaluation function	Two peer reviews can be considered to have delivered no improvements in independence (UNRWA, UNODC).
A - 2	Medium level impact - some substantive changes in the independence of evaluation function but changes that do not amount to a step change/transformative change	Two peer reviews had some impact on the degree of independence of the evaluation function (IFAD and ICAO). IFAD's degree of independence was already strong prior to the peer review.
G - 4	High level of impact: marks a step-change/transformative improvement in the independence of the evaluation function	Four of the eight peer reviews can be considered to have had a transformative effect on the independence of the evaluation function (WFP 2014, ITC, UNICEF, UNESCO).

The next section of the report explores what features are associated with peer reviews that have high and low degrees of impact on credibility, utility and independence.

What factors enhance and reduce the utility of a peer review?

It is possible to identify several lessons and trends from this overview of peer reviews in each agency. The paragraphs below highlight some overarching findings.

The consultant was asked to comment on whether features such as the peer review modality used or the quality of the report would have affected the impact of the peer review. There is no evidence that this was the case. The size and complexity of the evaluation function informed whether the full or reduced framework was applied and how many panel members served. The majority of peer reviews were conducted using a 'hybrid' modality, with one panel mission, with a representative of EvalNet and addressed the full scope of a peer review (i.e. following the UNEG independence, credibility and utility assessment). There was also variation in how the 'reduced modality' was applied.

It is not possible to state whether the quality of the peer review report has any bearing on the impact of the peer review. The peer review reports vary widely in how they treat the evaluation function. This is to be expected as they aim to review evaluation functions that range from one officer (ICAO) to nearly 30 (WFP). Additionally, what constitutes a 'quality' report will be subjective. An evaluation function may view a report as weak if its officers had hoped for it to say something that it does not. This said, all the reviewed reports met basic quality requirements – they were well written, well-structured and presented reasoned judgements and gave recommendations that appeared actionable. The extent to which their recommendations reflected an understanding of the evaluation function's organisational environment and constraints is extremely hard to judge after the fact: a reviewer is not party to all the contextual factors the panel will have been exposed to. Hence, there is limited utility in assessing this element of the quality of reports.

Clearly, though, the quality of the report and its recommendations will have a direct bearing on the impact. A strong panel and consultant with good comparative experience and a strategic approach will give clear and helpful recommendations. A weak panel will likely produce a weak report. None of the interviewees attributed a peer review's impact to the quality of the report, however. Therefore, assuming reports are of adequate quality, the factors that determine the influence of a peer review lie elsewhere. This is what the following sections explore.

The maturity of the agency's evaluation function

Peer reviews can have a transformative effect on an evaluation function: they can help an evaluation function make a step change in its adherence to UNEG norms and standards. But peer reviews do not have a transformative effect on mature evaluation functions that have already established institutional adherence to UNEG standards of independence, credibility and utility. The two WFP peer reviews illustrate this well. The 2014 peer review mandated transformative changes: it helped the evaluation function to secure significant and predictable resources and to establish an entirely new decentralized evaluation system. The 2021 peer review (WFP's third) thus provided more of a health check of already strong standards and systems: it contributed to course corrections rather than transformative change. Likewise, IFAD's 2019 peer review was of a mature evaluation function which had already established its independence.

The limits of a peer review in facilitating improvements in independence

It is clear that some evaluation functions were in need of transformative change – and the heads of the evaluation function invited peer reviews in the hope they would help them drive these changes. Some were successful and some were not. If an evaluation function's management is not committed to ensuring an independent and credible evaluation function, a peer review cannot change that. The added value of a peer review in these contexts will therefore be limited. As one interviewee noted 'the politics of an organisation can be very strong: if an evaluation function is vulnerable a peer review can't help'; and 'The tool itself is not strong enough to drive change; if you have a favourable environment, it can have an impact'. Despite hopes that the peer reviews would lead their organisations to implement UNEG standards, the UNODC and UNRWA peer reviews had little to no effect on their financial independence. The stance of the respective agencies' management (at the time) towards their evaluation functions and their attitude to the UNEG peer review (in the case of UNODC) explain the lack of impact.

However, even when an agency does not implement peer review recommendations a review still delivers value. This value is in making an agency's management and governing bodies aware of UNEG standards and of how their agency falls short. Whilst this awareness may not lead to any action immediately after a peer review, it serves as a benchmark that successive studies (e.g. MOPAN assessments) can draw against. It helps to make weak adherence to UNEG norms and standards more visible and therefore exerts some pressure for change over the long term.

Key stages of a peer review: Analysis of Factors that influence impact

The section below explores the strengths and weaknesses related to different stages of a peer review and seeks to identify which practices can be strengthened. It identifies good as well as poor practices. The observations below were mainly made by interviewees. Where they are from the author, it was indicated clearly in parenthesis. Each of the processes highlighted in the figure below are treated separately:
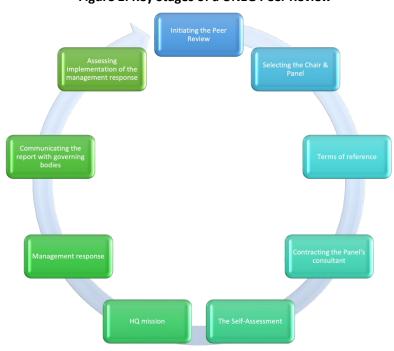
Figure 1. Key Stages of a UNEG Peer Review

Initiating the Peer Review

Timing and purpose: There are advantages to scheduling a peer review to inform a new Evaluation Policy

Several evaluation functions requested a peer review specifically to inform their next Evaluation Policy – basing the timing of the peer review on this need. This includes WFP 2014 and 2021, UNESCO 2020, UNHCR, IOM AND UNITAR in 2021. Interviewees considered that the peer review helped identify key elements they needed to consider in their new Evaluation Policy, and developed the new policy on the back of the peer review report. One head of an evaluation function regretted not doing so, considering it a missed opportunity.

Building engagement with agency management and governing bodies prior to a peer review

Cultivating management and governing bodies' buy in prior to a peer review helps to enhance follow-up, and ownership of peer review findings and recommendations. Some interviewees considered that either an agency's governing bodies or management should formally write to UNEG to request a peer review – it is not clear how many agencies do this (UNHCR and IOM did). One noted 'When the Executive Board is not interested the peer review is treated as routine and the quality of engagement suffers'.

> UNEG could insist that a peer review is only initiated following a request from senior management or governing bodies, and its response could set expectations. At a minimum these should be that the agency produces a management response, and that it is published with the peer review report.

Selecting the Chair & Panel Members

Difficulties finding members and the importance of keeping to schedule

Peer reviews are based on a voluntary system and there are a limited number of potential panel chairs: many appropriate chairs are already over-stretched managing an agency's evaluation function. It is not a surprise therefore that the Peer Review Working Group has had difficulty identifying panel chairs and members.

Several interviewees indicated that finding appropriate panel chairs and other members had delayed the start of a peer review. Difficulty finding peer review chairs, panel members, or sometimes OECD DAC representatives meant peer reviews could be delayed by up to 12 months. Interviewees highlighted that if a Peer Review is undertaken later than anticipated it can reduce its potential impact. Delays have led to negative consequences in two ways:

- 'Windows of opportunity' can be missed: A delay can mean that a 'champion' from the agency's senior management team who would otherwise have been in post leaves, rendering the peer review without meaningful management engagement (e.g. UNRWA). One panel member noted

of their delayed peer review that 'It had an effect on the impact the Peer Review– we lost the window of opportunity – the timing was wrong.'

- A delay can affect the degree of follow up taken after a peer review: For example, if, on the original schedule, the head of an evaluation function would have been in post but is due to move soon after, a delay can make a big difference to follow up (e.g. UNICEF). The peer review can end up 'hanging', with weak follow up. It is important to ensure that the panel is put together on time, but also that there is a buffer between the peer review and any planned changes in the leadership of the evaluation function.

Finally delays, or the potential of delay, can lead some agencies to choose an alternative modality such as a review or evaluation.

> The Peer Review Working Group should consider how best to avoid delays relating to finding panel members: delays can undermine the efficacy of peer reviews.

Qualities of the panel Chair and members

Interviewees identified some principles for panel composition that can make a peer review more impactful. Several interviewees noted that there is an implicit understanding that to be respected by an agency's management, the panel chair needs to be the head of a larger evaluation function. This does limit the 'pool' from which chairs can be sought for different peer reviews. Chairs are also said to need to be able to be able to 'navigate the waters politically' to engage effectively with an evaluation function's management and governing bodies. In terms of the broader panel, interviewees noted that a panel with a 'good balance of people' can have a lot of influence with management and governing bodies – it does not rely solely on the quality of the chair.

Ensuring the independence of the panel chair and of panel members from the reviewed evaluation function

Several interviewers stated that it is crucial that the 'brand' of peer reviews is protected from any suspicion that peer reviews amount to 'friends reviewing friends'. Ensuring the integrity of Peer Reviews is perhaps the most important responsibility of the Peer Review Working Group. It needs to ensure there is no conflict of interest in the choice of a Panel Chair, Panel members, and the Consultant.

Some evaluation function heads reported having to come to informal arrangements with another evaluation function head to act as a Panel Chair. This was sometimes said to be a way of ensuring a quality review was done; or that it was done within the needed timeframe. Given the small pool and the need to find suitably experienced panel chairs, this is understandable. But the Peer Review Working Group needs to ensure that there is never a conflict of interest among any of the parties involved.

One UNEG member noted that his organisation's governing bodies insisted that the evaluation function was assessed by an external evaluation rather than a peer review as it considered the former more robust.

UNODC's management did not engage fully with the peer review or issue a management response, ostensibly, because it did not consider UNEG to be an accredited professional body.

> To preserve the credibility of peer reviews, there needs to be a visible degree of independence between the panel and the reviewed evaluation function.

The contribution of an EvalNet/OECD DAC member on the panel – and to the efficacy of the review

Panel chairs and other UNEG members were unanimous in attributing importance to the fact that one of the panel members represented an OECD DAC country. They noted that the membership of a representative of a major donor country made the agency's management take the entire peer review more seriously. The link between UNEG and OECD DAC is seen as, 'really powerful: it strengthens the credibility of Peer Reviews'. The 2018 survey of evaluation function heads made a similar observation.[25]

In two cases where it appeared the agency's management was not engaging as expected with the outcome of a peer review, the OECD DAC member used their links to their representative on the agency's governing bodies to raise issues of concern. OECD DAC panel members sought to mobilise their national representatives when they felt that an agency's management was not treating peer review seriously.

However, as interviewees noted, any role OECD DAC members play is on their own initiative and informal: there is not a formal expectation that they will mobilise either their own government or other governments; nor is there an expectation that they will formally report back to their own organisation (and specifically its multilateral funding department) on the outcome of a peer review.

> The presence of OECD DAC panel members adds to the perceived credibility of peer reviews; they can also be influential when an agency fails to act on a peer review – though there is no stated requirement for them to exert their influence.

---

[25] The 2018 report of the survey states that 'A commonly identified strength was the involvement of an OECD-DAC panelist and its value in strengthening the credibility and independence of the review'.

The Peer Review Terms of Reference

The quality of the TOR

Several Panel Chairs said that they considered that the agreed ToR were too broad – and that they combined 'micro' issues with more strategic issues. As one noted, 'We need to ensure that ToR are realistic'. In the course of undertaking the peer review, panels tended to focus down on the issues they considered most strategic – but this selectivity could sometimes leave the evaluation function feeling 'short-changed' - as some issues listed in the ToR were not addressed. Another comment was that ToR need to be better aligned to the situation, status and context of the organisation.

The following issue was not mentioned by interviewees but is my observation: one way of narrowing down the scope of ToR is to make the request for the panel (or the consultant) to quality assure a sample of evaluations contingent on the organisation not commissioning independent quality assurance themselves. Large evaluation functions tend to commission their own quality assurances. For example, the ToR for the peer reviews of UNHCR, WFP, UNESCO and IOM required a quality assurance exercise but all either had already embedded quality assurance systems or had had recently commissioned them (e.g. IOM). In these cases it was appropriate for the panel to review the quality assurance systems rather than a sample of evaluations themselves.

Contracting a consultant

Consultants are contracted for all peer reviews. They can make the difference between a strategic and a weak peer review. One interviewee explicitly asked the author to 'bring the role of the consultant into the sunshine'.

A consultant with strong skills will help the panel by taking the analysis to a certain level. The consultant's analysis should help the panel grasp the strategic issues facing the evaluation function thereby giving them a platform from which to develop their ideas and recommendations. As one chair noted, their consultant 'helped the panel really understand, and the panel used what the consultant delivered and improved on it – so there was synergy'. Another stated 'you need someone who has vision, not a nuts and bolts person'. A couple of panel chairs complained that their consultants were too focused on 'nuts and bolts'. Knowledgeable and strategic consultants are therefore a key requirement for effective peer reviews.

The role of the consultant is all the more important because peer review chairs and panels engage voluntarily whilst juggling professional duties (in the case of the chair, leading an evaluation function). As one chair noted, if the consultant could not have operated at a strategic level, the chair would have had to invest a lot more time in the peer review – time the chair did not have. Where consultants provide a 'full service' (e.g., draft Preliminary Assessments and then reflect the panel's views in final reports), this a) minimises the demands on the Chair/other Panel members b) gives assurance to the Chair that they can take on the chair role.

> Consultants with strong strategic and analytical skills help to make the continued voluntary peer review model feasible, productive and sustainable.

The Self-Assessment

Several peer reviewed evaluation functions noted that conducting the detailed self-assessment tool was useful. The tool, updated and expanded in November 2020, asks agencies to assess themselves against 48 organizational and performance criteria. These are drawn from assessment criteria and benchmarks 'used by UN entities such as the Joint Inspection Unit' and on earlier peer reviews.[26]

One noted that it is 'very, very useful' as it provides a tool for self-reflection and is flexible for a wide range of organisations. Some said that completing it meant that they could almost anticipate the findings of a peer review as they had already identified their own strengths and weaknesses.

The depth and breadth of agency's self-assessments reflects their size and complexity. For example, WFP used the self-assessment to reflect on its decentralized evaluation function and asked the organisation's regional evaluation staff to complete the relevant sections independently. This enabled the panel to focus on the decentralized evaluation system's strengths and challenges, and was helpful for both WFP and the panel.

The HQ Mission

Panels, and the chair in particular, need to build a relationship with an agency's management and Executive Board/representatives to develop buy-in before and during a Peer Review. This can help ensure that the panel's report is treated seriously, and that its recommendations are endorsed and supported. As one Panel chair stated: 'we need to build space for frank and open debates with the right people at the right time.' Peer reviews differed in the extent to which they engaged in this way.

Two evaluation functions, WFP and UNHCR, established or used existing groups as Reference Groups for the peer review. These gave the panels opportunities to discuss their preliminary findings with senior staff, test and 'socialise' their potential recommendations. Other evaluation functions that invite a peer review may wish to create a Reference Group.

A panel also has a stronger chance of 'landing' the peer review recommendations if it presents the report to the agency's governing bodies after it is agreed – in informal or formal sessions. This is not a universal expectation but would represent good practice.

> To build ownership of the peer review results, the panel needs to make a point of having high quality discussions with management and governing body representatives during and after the mission.
>
> Further, Reference Groups are a useful medium to help the panel engage with management in a structured way.

---

[26] See the UNEG Peer Review Working Group, 2020, Self-Assessment Maturity Matrix for UN Evaluation Functions, Draft updated 11 November 2020, for the complete matrix of criteria.

Management Response

Whilst a Peer Review is not an evaluation, the development and agreement of a management response helps ensure there is a basic degree of accountability for delivery of what was agreed. A robust management response helps ensure that management has deliberated on the report's recommendations; and where it agrees recommendations, that it has bought into them and can be held to account. Some agencies have not produced a management response – or it is much delayed.

> If the production of a management response is an obligation on the peer reviewed agency, there is a higher likelihood that the peer review has utility and impact.

Assessing implementation of the management response commitments

There is currently no consistent reporting on whether commitments made in a peer review management response have been implemented.

One suggestion made by interviewees was that the original peer review consultant could (on their original contract) review implementation of agreed recommendations 18 months after the management response was issued, and share this with UNEG. Alternatively, the peer reviewed agency or the Peer Review Working Group could be asked to produce an '18-24 month on' report (reporting could be earlier than 18-24 months).

The WFP 2014 and IFAD 2009 peer reviews did assess the extent to which the management response commitments were implemented.[27] This would appear to be best practice but will only be relevant for evaluation functions that have had a previous peer review. It would make sense for Terms of Reference to add the requirement that, if there has been a previous Peer Review, the panel/consultant review the implementation of agreed actions from the previous Peer Review. (*This observation is the author's and was not raised by interviewees*).

> UNEG could consider:
>
> - Commissioning from either the peer review consultant or the reviewed agency a report on progress in implementing commitments made in the peer review management response – 18 to 24 months after the peer review.
>
> - Adding to its guidance the expectation that peer reviews will always report on the implementation of management response commitments of previous reviews.

---

[27] WFP 2021 peer review, Annex 3 and IFAD 2019 peer review, p.3)

Broader observations on peer reviews as an instrument

The final part of this section explores some other features of peer reviews that have not yet been discussed.

The added value of peer reviews

The value added of the instrument

Many interviewees expressed very strong appreciation for the role played by peer reviews. They noted that peer reviews are one of UNEG's most important contributions, along with its contribution to norms and standards, knowledge and learning. It was seen to be one of the 'greatest value added from UNEG'.

The special added value of the peer modality

Interviewees highlighted the special value of the review being conducted by peers who themselves face similar challenges. As one noted, 'we have been in the evaluation function's shoes, we know how difficult it is to push an agenda with management – a peer review does not parachute someone it to preach'. Others also noted peers can be effective at engaging with the evaluation function's senior management to convey the need to meet UNEG standards.

But at the same time, interviewees noted that peers still make robust and independent assessments. Panel members make judgements and give recommendations that the evaluation function does not always necessarily welcome.

The contribution to learning and knowledge sharing

Interviewees stated repeatedly that peer reviews make a significant but intangible contribution to the learning of all panel members, including OECD DAC panel members, and to the reviewed evaluation function. This contribution was mentioned repeatedly in interviews. Peer reviews were cited as 'a mutually beneficial approach' that helps spread good practice. It also shows junior members of an evaluation function of the value of engaging with UNEG.

The primacy of the independence and credibility of peer reviews – protecting the brand: just one Peer Review instrument?

As mentioned above, several interviewees considered that one of UNEG's highest priorities should be to safeguard the credibility of peer reviews by preserving their independence (in relation to identifying panel members). Interviews also highlighted a second but related concern: that, in the view of interviewees, UNEG risks watering down the reputational integrity of peer reviews by calling other review instruments peer reviews. The strength of some interviewees' views on this subject should not be under-estimated. They were concerned that by calling instruments that are not fully-fledged peer reviews 'peer reviews' the entire system will be devalued.

Interviewees considered it fine if another instrument is needed but they should not be called peer reviews.

If UNEG allows the concept of a peer review to be devalued, UN agency managers will be able to dismiss peer reviews as a form of self-assessment that can be ignored.

> To safeguard the respect and influence of peer reviews, the Peer Review Working Group should consider ensuring that the term 'peer review' continues to refer only to genuine peer reviews that involve 'live' reviews by peers (whether these peer reviews include representatives of OECD DAC is a separate issue). The term 'peer review' should not be conflated with other modalities such as 'Expert Validated Self-Assessments'.

Substantive panels are sometimes needed for small evaluation functions

Whilst the peer review is considering adopting less intensive methodologies, UNEG should not assume that a small evaluation function requires a less intensive peer review. It should also factor in the operational expenditure of the agency. IOM is a case in point. At the time of its peer review it had three combined M&E staff but the organisation had an operational budget of $2B. It was appropriate to have a substantive Panel. The governance intricacies – and trying to map a route to establish a healthier evaluation function – are also no less complex in agencies that have smaller evaluation functions. *(This is the author's observation and was not raised by interviewees).*

The role of the Peer Review Working Group

Few interviewees mentioned the role of the Peer Review Working Group except in relation to its help establishing the panel. Its role and degree of engagement with panels has possibly changed over the last few years.

The head of one evaluation function that recently invited a peer review considered that the Peer Review Working Group gave 'a lot of very good advice' and is 'very valuable'.

The Peer Review Working Group has also helped to cover the costs of some peer reviews and this has been appreciated. Some interviewees consider that the Peer Review Working Group could go further in committing UNEG finances to covering the costs of peer reviews.

It has facilitated the exchange of letters with respect to some recent peer reviews. The Peer Review Working Group also arranged for the head of UNEG to send the UNHCR peer review report officially to the organisation's management (the IOM final report was sent in the name of the Vice-Chair of UNEG, as the Chair was absent). Official UNEG 'endorsement' was seen as helpful in these instances, but this is not a universal view. One interviewee considered that the authority of a peer review derives from the chair and the panel – and that it is not the role of UNEG to 'weigh in'.

Its performance in constituting peer review panels has been uneven – as noted in the text above. Criticisms were made by some interviewees that the Peer Review Working Group took a very long time to put the panel together. One interviewee considered that the Working Group had not taken a systematic approach

to recording and then carrying forward a list of individuals who had offered to participate in a panel – leading to delays.

When asked what the Peer Review Working Group should prioritise going forwards, several stated that its priority should be to safeguard the integrity of the peer review instrument and the association with OECD DAC.

Given that the Peer Review Working Group is comprised of volunteers, if it is to provide more support to peer reviews in future it may need to be supported by part-time, paid support (who could provide a secretariat function, e.g. on helping develop systems to ensure panels are staffed). This is something it should consider.

Findings and recommendations

The report now reviews the extent to which peer reviews do and can facilitate change in the credibility, utility and independence of UN evaluation functions. It then brings together recommendations generated by the analysis.

In terms of the study's overarching conclusion: there is clear evidence that peer reviews have added to the credibility and utility of individual evaluation functions and their products. There is also clear evidence that peer reviews have facilitated positive impact on the independence of evaluation functions and their products. But this contribution is more varied and impact is contingent on several factors.

The following paragraphs explore these assertions in more detail.

What impact do peer reviews have?

With high confidence we can state that each peer review has contributed to the credibility and utility of the reviewed evaluation function and its products. All the evaluation functions reported improving the way they managed the evaluation function following a peer review. Even when an evaluation function was small and underfunded, the peer review helped the function introduce cost-neutral changes that brought them closer to embedding key UNEG standards (such as putting the organisation's evaluations in the public domain). We saw that several evaluation functions went on to embed specific recommendations from a peer review report in their next evaluation policy.

The impact of individual peer reviews on the standards of individual agencies can therefore be considered significant. Logically, as peer reviews are undertaken they have a cumulative impact on the standards of evaluation in the UN system as a whole: they do enhance the professional standards of evaluation in UN agencies, as they aspire to do. As indicated in an earlier section, the 2013 study of 2005-2011 peer reviews reached the same conclusion. Thus, it could be said that peer reviews have consistently contributed to the strengthening of UN evaluation practice.

Peer reviews do sometimes contribute to the independence and therefore the institutional strength and robustness of the reviewed evaluation function. As we have seen this is much more difficult to do and

assessing the contribution of peer reviews to the independence and robustness of evaluation functions is more complex.

Some evaluation functions have been able to use a peer review to leverage much greater structural and financial independence – change that is transformational. The evaluation function and panel chair have been able to use the peer review to demonstrate to the agency's management and governing bodies that the institution falls short of UNEG norms and standards and that it is in the organisation's best interests to give the evaluation function the independence and the finances it needs to deliver an effective evaluation function.

A completed peer review will never automatically deliver these kinds of changes, however. The case studies show that to have an impact on independence, the head of the evaluation function and the panel chair need to purposefully position the peer review with the organization's management and key stakeholders. This needs to be done from the beginning to the end of the peer review process to prepare the ground for a positive reception of the findings and recommendations. This takes preparation – it almost requires the head of the evaluation function and the chair to have an engagement strategy.

But it also has to be recognised that agencies are not equally responsive to peer reviews. Agencies' overarching governance systems have a strong influence on how responsive management is. Agencies with active Executive Boards, a more direct relationship to their funders, and to whom the management is accountable tend to be keener to respond to peer review recommendations – their managers have a more direct incentive to ensure their stakeholders have confidence in their effectiveness. Some were said to be motivated by the desire to keep pace or exceed the standards of their sister agencies.

Agencies that are part of the UN Secretariat – with ultimate accountability through the Secretary General – have more diluted incentives to strengthen their evaluation functions. For example, contrast UNICEF and WFP (more responsive) with UNWRA and UNODC (less responsive). The former agencies are more concerned with how their performance looks in comparison to sister agencies, and tend to be more receptive to peer reviews' findings and recommendations.

But it is also the case that change is sometimes simply not possible even when all conditions align, and when the panel and evaluation function head have done 'all the right things'. Management may be antithetical to change or scrutiny. Champions in management, who supported the peer review initially may move on. Governing body members may be disinterested or distracted, or a financial crisis may hit an organisation – or all of the above.

Even when this is the case, and when a peer review has had no impact on the independence and robustness of an evaluation function, peer reviews are still worthwhile. They still provide a public statement of how the evaluation function falls short of agreed UN standards, and what needs to be done to strengthen the function. They can contribute to building pressure for change over the medium term. Peer review analyses serve as a benchmark for the next MOPAN or a subsequent UNEG assessment.

Indeed, UNEG can benefit from the 'tailwind' created by MOPAN assessments. As these become increasingly mainstream and as agencies are subjected to repeated assessments, managers may be motivated to act on peer review findings and recommendations – including the more difficult ones relating to independence. MOPAN reports disaggregate scores and it is evident to both managers and governing bodies what elements of institutional performance drag overall scores down. Thus managers have an incentive to

ensure their next MOPAN assessment is not affected by a low-scoring evaluation function. Indeed, one head of an evaluation function stated that whilst negotiating the peer review management response with management, they indicated as much: they were willing to implement peer review recommendations primarily insofar as they would secure better scores in the next MOPAN assessment.

Finally, peer reviews can encourage agencies to establish or strengthen internal mechanism that routinely support the independence and quality of the evaluation function. This can raise the likelihood that any improvements facilitated by the peer review are sustained over the long term. Several peer reviews recommended strengthening the attention that independent accountability bodies give to the evaluation function. The IFAD peer review recommended revising the role of an existing Evaluation Committee to ensure full oversight of evaluation and results reporting functions IFAD-wide (the 2009 peer review had built in oversight income the Evaluation Committee).[28] The UNESCO, IOM, and UNHCR peer reviews recommended strengthening the attention given to evaluation by existing independent oversight committees: they recommended the committees added a member with evaluation expertise and/or expanded its mandate to include the performance and resourcing of the evaluation function. When management agrees to these recommendations agencies gain a permanent authoritative 'champion' for the evaluation function. (*This is the author's observation and was not raised by interviewees*).

> The Peer Review Working Group may wish to recommend that panels, in their deliberations, routinely consider whether the agency has independent oversight bodies that, with focused attention, can help strengthen the evaluation function over the long term.

The utility of peer reviews for partners outside the UN system

The original theory of change for peer reviews was that they would strengthen UN evaluation functions to such an extent that multilaterals would not need to commission their own evaluations and could rely on evaluations conducted by the agency itself. It was beyond the scope of this study to see if this has happened in reality.[29] But the original theory of change was somewhat unrealistic in the first place. There is not an effective feedback mechanism between UNEG and EvalNet. A representative of the Peer Review Working Group sometimes attends EvalNet annual meetings and indicates what peer reviews have been completed or are ongoing, but the current relationship is not deeper than this.

And even if UNEG consistently updated EvalNet on the results of peer reviews this 'logic' would probably not have an effect. This is because EvalNet members represent their organisation's evaluation function not their multilateral funding department. It is difficult to ensure that messages will be fully understood and acted upon.

---

[28] IFAD 2019, p.8

[29] This would entail looking at whether each EvalNet donors reduced the number of evaluations they commissioned directly in each peer reviewed agencies.

EvalNet interviewees noted that, more broadly, its members are not well sighted on the significance of peer reviews. The Peer Review Working Group may wish to issue a simple, short (two page?) document outlining what peer reviews were conducted in the preceding year, and synthesising key themes or findings (as well as sharing full reports). Due to the turnover of EvalNet members, it would probably be worth regularly briefing members on the purpose of UNEG peer reviews; and how EvalNet members can use peer review findings (e.g., share them with their multilateral funding departments).

As noted earlier in the report there is a symbiotic relationship between MOPAN assessments and UNEG peer reviews. MOPAN assessments can add pressure on an agency to ensure that its evaluation function meet UNEG standards. Further, UNEG peer reviews help ensure that MOPAN assessments of institutional effectiveness and results are based on credible independent evaluations. The chair of the Peer Review Working Group may want to discuss with the MOPAN Secretariat in OECD DAC whether information sharing could be mutually beneficial. At the very least, the MOPAN Secretariat could be made aware of the role and value of UNEG in helping to ensure that – in an important subset of the MOPAN's target multilaterals– the quality of evidence improves.[30] It may be also advantageous for UNEG to brief MOPAN's membership (20 donors) on its work.

In aggregate then, peer reviews do help to lift the quality and independence of evaluation functions and they do help to increase the credibility, utility and independence of evaluation products. They therefore do in aggregate reduce the need for donors to conduct evaluations of specific interventions that they fund. Secondly, in aggregate they allow donors and their MOPAN reports to trust the performance evidence that these agencies produce.

Recommendations

The ToR asks this report to make recommendations to UNEG to enhance the utility of peer reviews. This section brings together the comments made through the body of the report and makes two recommendations under two themes, with nested actions. These are aimed at:

1.   Preserving the reputational integrity of peer reviews; and

2.   Strengthening systems, processes and partnerships to maximise utility

The consultant is not a member of the Peer Review Working Group and therefore, several of the suggestions made are **for consideration only**. Unless otherwise specified, all the recommendations are for consideration and action by the Peer Review Working Group and potentially the broader UNEG membership. Some require collaboration with EvalNet.

1.   Preserving the Reputational Integrity of the Peer Reviews

1.1   Ensuring the integrity of Peer Reviews is the most important responsibility of the Peer Review Working Group. It needs to ensure there is no conflict of interest in the choice of a panel chair, panel

---

[30] It is noticeable that two recent MOPAN studies make no mention of the contribution of peer reviews to evaluation in multilaterals. See MOPAN 2018 Strategic Review; MOPAN, 2021, Lessons in Multilateral Effectiveness Is this time different? UNDS Reform: Progress, challenges and opportunities.

members, and the consultant. It needs to ensure that the instrument continues to be seen as a robust assessment.

1.1 To preserve the credibility of peer reviews, the Peer Review Working Group needs to always ensure there is a visible degree of independence between the panel and the reviewed evaluation function

1.2   If UNEG allows the concept of a peer review to be devalued, UN agency managers will be able to dismiss peer reviews as a form of self-assessment that can be ignored.

1.2 To safeguard the respect and influence of peer reviews, the Peer Review Working Group should consider ensuring that the term 'peer review' refers only to exercises that involve peers (not consultants) in face-to-face engagement: the term 'peer review' should be not conflated with other modalities. (Peer reviews can of course be conducted virtually; and the inclusion of representatives of OECD DAC representatives is a separate issue).

2.   Strengthening systems, processes and partnerships to maximise utility

The actions under this recommendation list specific measures that could be adopted to help maximise the utility of future peer reviews.

2.1   If peer reviews are begun only following explicit commitment from the organisation's management or governing bodies, there is likely to be greater follow up.

2.1 To secure a measure of 'buy in' from agency management, UNEG could consider starting a peer review only following an official request from senior management or governing bodies; and approval could be conditional on the agency agreeing to provide a comprehensive management response within a set period.

2.2   Delays can undermine the efficacy of peer reviews.

2.2 The Peer Review Working Group should seek to introduce systems that reduce the risk that, after a peer review has been agreed in principle, it is delayed.

2.3   The participation of OECD DAC panel members adds to the perceived credibility of peer reviews.

2.3 UNEG should continue to seek the membership of OECD DAC representatives on peer review panels as a priority.

2.4   The peer reviewed evaluation function and the panel can help build ownership of the peer review recommendations in the conduct of the assessment. The Peer Review Working Group may want to issue guidance on this.

2.4 Panels should seek to build engagement with the peer reviewed agency's management and governing bodies prior to, during, and after the mission. Reference Groups are a useful medium to help the panel engage with management in a structured way.

2.5   The Peer Review Working Group could add a step to peer reviews that may help encourage agencies to take seriously their commitments made in a management response, UNEG could consider:

2.5 Commissioning from either the peer review consultant or the reviewed agency a report on progress in implementing commitments made in the peer review management response 18 to 24 months after the peer review.

2.6   It would also be good practice to ensure that, where relevant, peer reviews assess progress made since a preceding review:

2.6 The Peer Review Working Group could add to its guidance the expectation that peer reviews will always report on the implementation of management response commitments of earlier reviews.

2.7   To continually learn from the conduct of peer reviews, the Peer Review Working Group could reinstate the request for participants to produce some form of a 'learning note' after the review. This could take the form of a simple 'Survey Monkey' questionnaire to the head of the evaluation function and the panel chair. This would help ensure the Peer Review Working Group receives feedback and is able to adjust the instrument accordingly.

2.7 To help close the feedback loop, the Peer Review Working Group could consider introducing an uncomplicated way of to capture lessons from each peer review.

Annexes

Annex 1: Persons Met

Please note that several interviewees had served as Panel Chairs as well as being heads of evaluation functions that were peer reviewed.

UNEG Members

Adan Ruiz Villalba, Head of Evaluation, World Intellectual Property Organization

Andrea Cook, Director of Evaluation, WFP

Andrew Fyfe, Head, Evaluation Unit, United Nations Capital Development Fund (UNCDF)

Brook Boyer, Director, Division for Strategic Planning and Performance & Manager, Planning, Performance Monitoring and Evaluation Unit, UNITAR

Christophe Franzetti, Head of Evaluation, IOM

Claudia Ibarguen Head of Evaluation, UNESCO

Colin Kirk, Former Director of Evaluation, UNICEF

Dr. Elil Renganathan, Head, WHO Evaluation Office, World Health Organization (WHO)

Gugsa Farice, Evaluation Specialist, ICAO

Guy Thijs, Head of Evaluation, ILO

Helen Wedgewood, Former Director of Evaluation, WFP

Indran A. Naidoo, Chair, Director, IFAD (Former Director, of the Independent Evaluation Office, of the United Nations Development Programme (UNDP).\

Inga Sniukaite, Chief of Independent Evaluation Service, UN Women (email communication)

Katherina Kayser, Head of Independent Evaluation Section, UNODC (email communication)

Katinka Koke, Specialist, Planning, Performance Monitoring and Evaluation Unit, UNITAR

Lori Bell, Head of Evaluation, UNHCR.

Michael Spilsbury, Director Evaluation Office, UNEP

Miguel Jimenez-Pont, Head Monitoring and Evaluation Unit, ITC

Oscar A. Garcia, Director of the Independent Evaluation Office, UNDP. Former Director, IFAD

Robert Stryk, Regional Evaluation Head, UNICEF (Former Head of Evaluation, UNRWA)

Susanne Frueh, Former Director, Internal Oversight Service (IOS) UNESCO; Director for Internal Oversight Services, Organization for Security and Co-operation in Europe (OSCE)

Tuncay Efendioglu, Director of the EAO, ICAO

OECD DAC Representatives

Isabelle Mercier Director of Evaluation at Global Affairs Canada (and Vice-Chair of EvalNet)

Megan Grace Kennedy-Chouane, Head of EvalNet Secretariat, OECD DAC

Meike de Goede, Policy and Operations Evaluation Department, Ministry of Foreign Affairs, Netherlands

Per Øyvind Bastøe, Associate Professor/Evaluation Adviser (Former Chair of EvalNet)

Rob van Poelje, Policy and Operations Evaluation Department, Ministry of Foreign Affairs, Netherlands

Tom Hennessy, Head of Evaluation & Audit Unit, Department of Foreign Affairs, Ireland.

Annex 2: Bibliography

DAC/UNEG 2017 Joint Task Force on Professional Peer Reviews of Evaluation Functions in Multilateral Organizations - Framework for Professional Peer Reviews

Evaluation Cooperation Group, 2009, 'Review Framework for the Evaluation Function in Multilateral Development Banks'.

ICAO 2020, Professional Peer Review of the ICAO Evaluation Function

IFAD 2019, Final Draft Report of the External Peer Review of the Evaluation Function At IFAD

IOM 2021, UNEG Peer Review Report

ITC 2016, Professional Peer Review of the Evaluation Function & Management Response to the Peer Review

JIU, 2014 Analysis of the Evaluation Function in the United Nations System

MOPAN 2021, Lessons in Multilateral Effectiveness Is this time different? UNDS Reform: Progress, challenges and opportunities.

MOPAN, 2018 Strategic Review

MOPAN, 2019, '2017-18 MOPAN Assessment of UNRWA'

MOPAN, 2019' MOPAN 2017-18 Performance Assessments: International Organization for Migration'.

OECD DAC (n.d.) Evaluating multilateral effectiveness - work of the DAC Evaluation Network - OECD

OIOS, 2020, OIOS Inspection of UN Women Evaluation Function

UNEG 2011, The UNEG Framework for Professional Peer Reviews of the Evaluation Function of the UN Organizations.

UNEG, 2013, Report of Lessons Learned Study Peer Reviews of UNEG Evaluation Functions,

UNEG, 2016 Revised UNEG Guidelines for Professional Peer Reviews

UNEG, 2017 Norms and Standards for Evaluation.

UNEG, 2018, Background note on the UNEG peer review mechanism.

UNEG, 2019, UNEG Principles of Working Together

UNEG, 2020, Self-Assessment Maturity Matrix for UN Evaluation Functions, Draft updated 11 November 2020

UNESCO, 2020 UNEG Professional Peer Review of the UNESCO Evaluation Function

UNHCR 2021, Peer Review of the Evaluation Function at the Office of the United Nations High Commissioner for Refugees (UNHCR)

UNICEF 2017, DAC/UNEG Peer Review of the Evaluation Function of UNICEF & Management Response to the Peer Review

UNITAR, 2021, Professional Peer Review of the UNITAR Evaluation Function

UNODC, 2016 Professional Peer Review of the UNODC Evaluation Function

UNRWA, 2016 Professional Peer Review of the UNRWA Evaluation Function & Management Response to the Peer Review

UNRWA 2021, Evaluation of UNRWA's Medium Term Strategy

UNWOMEN Professional Peer Review of the Evaluation Function of UN Women Final Report of PR

WFP, 2007, Peer Review of the Evaluation Function at the UN World Food Programme

WFP, 2014 Peer Review of the Evaluation Function of the UN World Food Programme (2008-2013)

WFP, 2021 UNEG/OECD DAC Peer Review of the Evaluation Function at the World Food Programme

WFP 2021, WFP response to the recommendations in the summary report on the peer review of the evaluation function at the World Food Programme, Executive Board Annual Session 18 May 2021

UNEG, 2013, Report of Lessons Learned Study Peer Reviews of UNEG Evaluation Functions

UNEG, 2018, Background note on the UNEG peer review mechanism

Annex 3: Summary Terms of Reference for Stocktake

UNITED NATIONS EVALUATION GROUP (UNEG) PEER REVIEW WORKING GROUP

WORKSTREAM: UTILITY FUNCTION

Terms of reference for a Stock-taking Study on the Utility of Peer Review

Background

3.   The peer review mechanism is one among a broad spectrum of modalities that can be used to assess the functioning and results of evaluation functions in the international system. It was established as a voluntary mechanism, designed to help evaluation offices learn from the experiences of their peers, point to strengths and weaknesses, and help shape good practice in evaluation amongst bilateral and multilateral organizations. Initiated in 2004 through an OECD-DAC/UNEG joint Task Force, the first generation of Peer Reviews focused primarily on the three critical core criteria - independence, credibility and utility that, together, provide a perspective on the performance of an evaluation function. In 2011, a new Framework for Peer Review[31] proposed a more streamlined framework for the continuation of peer reviews shifting the overall responsibility and ownership over the mechanism to UNEG. The framework also was designed to be adaptable to evaluation functions of different sizes, adjusting the panel size and scope. The second generation of peer reviews also focused on specific issues identified jointly by the peer review panel and the evaluation function being reviewed. Since then, the UNEG Peer Review Working Group (PR WG) has produced various guidelines and, most recently, the self-assessment matrix has been revised and simplified to reflect the updated UNEG Norms and Standards (2016)[32].

4.   As stated in the 2011 UNEG Peer Review Framework, peer reviews: (i) **provide an independent and professional assessment** on the extent to which the UNEG Norms and Standards have been adopted by an evaluation function; (ii) **build greater knowledge and confidence of evaluation systems thereby leading to an enhanced use of evaluation** by management, governing bodies and other stakeholders; (iii) **provide a means to "evaluate the evaluators" through a professional and independent panel** which can lead to increased trust, confidence and accountability in UN evaluation functions; and (iv) **provide support and mutual learning** through the sharing of good practices, tools and experiences, and contribute to enhanced professionalization of UN evaluation functions.

5.   A first stock-taking exercise on peer review processes was carried out in 2013 which helped UNEG to develop further good practices and guidance on peer review.

6.   As part of its 2021 workplan and building on the experience gained since 2013 from the second generation of peer reviews, the UNEG PR WG intends to revisit the UNEG Framework for Peer Review and the related guidance. This **stock-taking exercise** will focus on the results/utility of peer reviews

---

[31] UNEG/REF (2011), The UNEG Framework for Professional Peer Reviews of the Evaluation Function of the UN Organizations. Available at: http://www.unevaluation.org/document/detail/945.

[32] Available at: www.unevaluation.org/2016-Norms-and-Standards.

completed to its key stakeholders, with the intention of informing an updated UNEG Strategy on Peer Reviews, to include an expanded array of peer review modalities, as well as revisions to the current Normative Framework and guidance documents.

Purpose and Scope of the Study

7.   The stock-taking study aims to understand the purposes that peer reviews have actually fulfilled, and how, and by whom, the peer reviews have been used in the agencies peer reviewed. The study will cover the period 2014-2020. The study will also examine whether, and in what way, peer reviews may serve a broader set of stakeholders beyond the agencies peer reviewed (e.g. UNEG, Evaluation Community; OECD-DAC members). It will identify enabling and hindering factors to the utility of peer reviews, in particular linking utility to peer review processes and drawing lessons for the future. The study will also explore how peer reviews can become more fit-for-purpose and useful to key stakeholders, by clarifying the roles they can play, as compared to other types of assessment modalities of evaluation functions in international organisations.

8.   The study will attempt to respond to the following key questions:

- How has the implementation of peer reviews varied over the period review in terms of: i) the Terms of Reference applied; ii) the composition of the Panel; iii) the structure and length of the report; iv) the process of the peer review; v) the consistency of quality standards applied; vi) the contribution and mandate of the Peer Review Working Group in supporting peer reviews?

- How has the quality of peer reviews varied in the judgement of those being peer reviewed/the UNEG Group charged with overseeing the peer reviews? How far have peer reviews met the core UNEG Norms and Standards of independence and credibility?

- How have the findings and recommendations of the peer review been communicated and disseminated to the relevant stakeholders?

- To what extent have recommendations from peer reviews been fully adopted? If recommendations have not been fully adopted, why not?

- Have recommendations from peer reviews played a role in changes to evaluation functions? If not, why not?

- How have peer reviews been used and by which stakeholders? If not, why not?

- To what extent have the second generation peer reviews contributed to improving the status, practices and effectiveness of the organization's evaluation function?

- What other changes have been directly or indirectly induced by the peer reviews whether within the evaluation functions or more broadly within the agencies (senior management and governing body)?

- What factors have facilitated or hindered the effectiveness of the peer reviews?

- What is the value-added of peer reviews as compared to other types of evaluations? Are there noticeable differences between the peer review modalities? What are the strengths and weaknesses of the various peer review modalities in terms of utility?

- What expectations do the UN's key evaluation stakeholders have when it comes to the UNEG peer review modality? What are some of the perceived strengths and weaknesses, opportunities and threats of UNEG – led peer reviews of UN evaluation functions in the context of UN reform?

- How should peer reviews be best managed and supported by UNEG and the Peer Review Working Group in view of i) UNEG's role as the accepted setter of evaluation norms across the United Nations system and ii) the Working Group's role as the custodian of the peer review modality?

Approach and Methodology

9.   Since 2013, 10 "second generation" peer reviews have been completed, including two of the GEF Evaluation Office. The study will incorporate all these peer reviews (with the exclusion of the GEF peer review of 2014) specifically: UN Women (2014); WFP (2014); UNRWA (2015); ITC (2016); UNODC (2016); UNICEF (July 2017); IFAD (2019); UNESCO (2019); ICAO (2020).

10.   For each peer review, the study will attempt to identify actions that have been taken and changes induced by the review whether based on the recommendations contained in the report or following exchanges during the peer review process. In addition, the study will seek to identify those actions that have not led to changes, as well as the factors that have both facilitated and hindered the uptake of the recommendations.

The study will be based on:

- Document reviews (peer reviews; management responses; reporting documents on follow-up actions to the peer reviews; lessons learned documents; UNEG peer review related references)

- A survey, focus group discussion and interviews of stakeholders of peer reviews using virtual platforms.

- Stakeholder analysis: Possible stakeholders are: Head and staff from the evaluation function; Senior Management and Board members, (if appropriate); Panel Members; OECD DAC members (when relevant), Oversight Advisory Committees (when relevant) and UNEG members.

11.   UNEG is looking to recruit a senior consultant to conduct this study. S/he will work in close collaboration with the UNEG PR WG, report to the coordinators of the WG workstream 3 on peer review utility. The consultant will design the details of the study, including confirming the completed peer reviews covered by the study, suitability of the approach given that some of the peer reviews have been completed long ago and the staff turn-over in UN agencies. The consultant will be supported by a junior analyst who will help with the collection of data, organization of meetings, interview notes and document library. The consultant will prepare a report which will draw lessons on the experience over the past 7 years and include an analysis of the uses and users of peer review, factors that have hindered or enabled use, and contribution of peer reviews to changes in evaluation functions. The report will also make recommendations to UNEG

to enhance the utility of peer reviews to be integrated in a UNEG strategy for peer reviews as well as in its Framework and guidance documents.

12.   The study is to be undertaken in accordance with the United Nations norms and standards for evaluation, and the UNEG Ethical Guidelines.

Learn more about UNEG