Reference document:

Origin and background of the DAC Evaluation Criteria

This note is made available to workshop participants as a background document, particularly for workshop participants who were not present at the 21st meeting of the DAC Evaluation Network. It is an excerpt of a discussion paper prepared by Hans Lundgren for the 21st meeting with some updated information.

Informal Workshop
15 March 2017
In the context of a broader debate about the future of development evaluation, a discussion has begun on whether it is time to re-think the so called “big five” – the DAC evaluation criteria of relevance, effectiveness, efficiency, impact and sustainability. These criteria underpin most evaluations in international development and have broad practical use far beyond the membership of the network. This note outlines the background, the changing development context.

1. Origin and background

1. The origin of the evaluation criteria goes back to the Principles for Evaluation of Development Assistance which was developed by a predecessor body to the Network – the DAC Expert Group on Evaluation. They were agreed to by the members of the Expert Group following a two year period of elaboration and discussions, and subsequently endorsed by the DAC High Level Meeting in 1991. They were a landmark achievement setting out important principles for members related to: the definition and purpose of evaluation, the institutional structure for managing evaluations, the principles of impartiality, independence and credibility, evaluation management processes, and partnerships.

2. Moreover, the document included principles for design and implementation of evaluations and recommended a manageable framework for the definition of questions in the terms of reference. It defined basic groups of evaluation issues such as rationale, objective achievement, impacts and effects to be able to ensure that evaluators could assess overall results, sustainability and possible lessons learned.

3. The principles are still in use as a benchmark in peer review processes of DAC members, notably in the assessments related to evaluation policies, institutional set-up, and use of evaluations.

4. The Network has complemented this early work with developing norms and standards on various topics over the years. These include: guidance for evaluating humanitarian assistance in complex emergencies; the glossary of key terms in evaluation and results based management; evaluation systems and use: a working tool for peer reviews and assessments; guidance for managing joint evaluations; guidance on evaluating peacebuilding activities in settings of conflict and fragility; and quality standards for evaluation.

5. The DAC Evaluation Criteria draws on the principles and also on the definitions agreed in the glossary of key terms in evaluation and results based management. The glossary was developed in a three-year consultative process and approved in 2002 and have since been published in 17 languages, and is used in training courses and practical work.

6. The purpose of the evaluation criteria is to help evaluation managers reflect on and structure the questions of the evaluation. The criteria has become widely applied in evaluation, even more so than originally expected, both for decentralised and centralised evaluations and for studies done by partners or implementing agencies.
7. Through successive reviews of evaluation systems, the most recent in 2010 and 2016, the widespread use of DAC norms and standards by members in policies and evaluations have been confirmed. Overall, DAC norms and standards play a central role in the development of evaluation policies and guidance. Most members base their evaluations on DAC terminology, principles, quality standards and criteria, using them as a reference framework for further elaboration in ministry or agency specific guidance.

8. The DAC Quality Standards were developed through a three-year consultative process with members and partners and agreed to by the network in 2010 and subsequently endorsed by the DAC. On criteria, the quality standards states:

   Selection and application of evaluation criteria. The evaluation applies the agreed DAC criteria for evaluating development assistance: relevance, efficiency, effectiveness, impact and sustainability. The application of these and any additional criteria depends on the evaluation questions and the objectives of the evaluation. If a particular criterion is not applied and/or any additional criteria added, this is explained in the evaluation report. All criteria applied are defined in unambiguous terms.\(^1\)

9. In short, the quality standards recommend the use of the criteria but also clarified that additional or fewer criteria can be used depending on the purpose of the evaluation. It does not recommend a mechanical use of the criteria.

1.1. Humanitarian assistance

10. In the guidance on humanitarian assistance, developed by the expert group on evaluation, it was noted that while the five evaluation criteria are broadly appropriate, their applicability in contexts of complex emergencies could be increased through specific elaboration of some criteria or adding sub criteria. For example, sustainability of humanitarian operations would rather be assessed through connectedness to assure that activities of short term emergency nature are carried out in a context which takes longer term and interconnected problems into account. The guidance on humanitarian assistance also suggested adding complementary criteria such as coverage (of populations and different groups) and coherence (referring to policy coherence), and co-ordination (the latter could be considered under the criteria of effectiveness but was singled out due to the multiplicity of actors in an emergency response).

11. In 2016, ALNAP issued a new guide titled “Evaluation of Humanitarian Action Guide”. This comprehensive document of 429 pages involved consultations and drafting over a six year period with a pilot guide issued in 2013. It covered many practical steps to guide the evaluation process. On the issue of evaluation questions, it provides guidance on the need to focus and make the exercise useful. Criteria are described as useful tools once evaluation questions have been decided. It notes: “Although the OECD-DAC criteria are not perfect, there are good reasons to use them to check your evaluation questions:
   • using standard criteria makes meta-evaluation much easier
   • standard criteria are likely to capture common weaknesses in humanitarian action
   • evaluation that use standard criteria are easier for evaluation managers and other evaluators to work with”.\(^2\)

12. It further elaborates on sample questions linked to each of the evaluation criteria included in the DAC 1998 guidance on evaluation of humanitarian aid and uses the term “appropriateness” rather than relevance.

---

\(^1\) Quality Standards for Development Evaluation, OECD, 2010, pg.9

1.2. Peacebuilding

13. In the DAC Guidelines on evaluating peacebuilding activities in settings of conflict and fragility (2012), it is noted that the five OECD DAC criteria for evaluating development assistance are usually considered, though it may be more manageable to focus the evaluation on looking at a few criteria in-depth, depending on the evaluation purpose and intended use. Additional considerations that may be particularly relevant to situations of conflict and fragility, namely coherence and co-ordination, could also be subject to examination. Moreover, an entire section of the guidance discusses the use of the criteria and evaluation questions related to peacebuilding activities.

1.3. UNEG and ECG

14. UNEG issued a new standards and norms publication in 2016, updating the earlier norms from 2005. The new standards were produced by a working group of UNEG members. It contains a broad set of general and institutional norms and standards for evaluation. It includes four new institutional norms compared to the 2005 standards which are related to UN agreed principles, goals and targets, human rights and gender equality, national evaluation capacities and professionalism. The objective is that these norms should be reflected in the management and governance of evaluation functions.

15. On the conduct of evaluations, the issue of evaluation criteria is referred to in standard 4.3: “Terms of reference”, where it is stated that the terms of reference should include, inter alia…“the evaluation objectives with key evaluation questions and/or criteria”. Thus, specific evaluation criteria are not elaborated on in the UNEG standard on TOR.

16. However, in another standard on “Human rights-based approaches and gender mainstreaming strategy” (4.7) it is stated that “the evaluation design should include considerations to the extent to which the United Nations system’s commitment to human-rights based approaches and gender mainstreaming strategy was incorporated in the design of the evaluation subject”. The standard goes on to specify six requirements of the evaluation terms of reference which relates to incorporation human rights and gender equality aspects. These include: to spell out relevant human rights and equality instruments or polices that will guide the evaluation process; incorporate an assessment of relevant human rights and gender quality aspects through the selection of the evaluation criteria and questions.

17. In sum, the UNEG standards are quite specific on how to incorporate human rights approaches and gender in the design of the evaluation, but they are not specific on the evaluation criteria to consider in TORs.

18. The Evaluation Co-operation Group (ECG) of the international financial institutions has earlier developed good practice standards (GPS) for private sector operations and for public sector operations. While some of the MDBs made strong use of these, they were however, unevenly applied across the MDBs, with some only applying one of the two GPS sets. It seems that the ECG has ceased elaborating good practice standards and is now focusing on developing guidance for particular themes and types of evaluations. More emphasis is put on stocktaking exercises, like in the area of self-evaluation in MDBs, evaluation recommendations, governance and independence, and on exchange of practices and lessons.

2. The changing development landscape

19. The past thirty years have seen rapid social and economic development in many parts of the world, with diminishing numbers of poor, better health and education for many. Hans Rosling
documented these developments in his famous presentations demonstrating that the world has overall become a better place for many. At the same time, there are countries and regions where development does not seem to take root, governance is moving backwards, civil society space is reduced, conflicts are becoming ever more protracted and the number of refugees and forced displaced persons are reaching new global highs. The rise of extremist violence and its consequences is another important part of the changed landscape. Other factors influencing the development scene are related to digitalisation, robotisation, artificial intelligence and the new industrial revolution. At the same time, we are witnessing increased inequality in many societies. Demographic shifts, including urbanisation and changes in age distribution and are also underlying influencers of the planet’s population dynamics. Overall, it is a fragmented and increasingly diverse context.

20. Against such a background, the adoption by the UN of the SDGs in 2015 was an important landmark and provides a vision for a better world where no one will be left behind. The implementation of the 17 goals, 169 targets and the monitoring process with 232 indicators will clearly provide numerous challenges but also opportunities for more holistic and cross-sectoral approaches to development. Sustainable development is not only a cross-governmental priority but a cross-societal one, inclusive of civil society and the private sector. The latter seems increasingly interested in the opportunities arising from the new framework.

21. The IEG blog raised the question whether we have reached a "Copernican moment" with tremendously important shifts of perspectives which will require a re-set of evaluation. In the debate, including at a recent UNDP hosted conference where a panel discussion on evaluation criteria was held; several “candidates” for new criteria have been mentioned, such as leaving no one behind, inclusiveness, diversity, and human rights and gender. It has also been argued that the evolution in evaluation methods and approaches needs to be reflected in the criteria, such as complexity, and the use of new and rapidly evolving technology.

22. The ODI report “Global development trends and challenges” by Homi Karas and Andrew Rogerson (published in October 2017) brings out a different viewpoint on the SDGs. It argues that the SDGs may not prove capable of transforming the development industry by 2030 and there is no indication yet that they are helping to generate greater financing.

23. While different parties may hold different views and perceptions, the factual situation is that, at this point in time, the effects on policies and programmes of the 2030 Agenda are still to be proven as there are as yet no evaluation evidence available.

24. The SDGs have targets and indicators related to many areas, however perhaps less on the climate front, as the Paris Agreement was still being negotiated at the time of the adoption of the 2030 Agenda. It has been suggested that the Paris Agreement should be seen as the complement to the SDGs, at least for all signatories of both agreements. Nevertheless, the implications of the Paris Agreement seem less referred to in the current debate on re-thinking evaluation than the 2030 Agenda.

25. Moreover, the issue of current economic and development models leading to unsustainable consumption patterns has been raised in the debate on evaluation and the criteria. In the blog by IEG it was suggested that new evaluation tools are needed to unpack intrinsic impacts on consumption patterns. How this aspect would influence a re-think of the criteria would require further clarification.

26. The development co-operation scene has also seen the emergence of new actors, including major new players, new small and mid-sized providers, private sector agents, foundations, and the
growth of south-south co-operation. Global multi-stakeholder partnerships are also influencing the global policy scene.

27. Development evaluation will need to evolve and adapt to these broader developments both in terms of its approaches, methods and in terms of the focus and objects of evaluation. The future development evaluation agenda is a broad topic that is addressed in numerous evaluation conferences and workshops and would require a fuller treatment than what is possible in this note.

28. It should however, be noted that development evaluation has evolved. Over the past fifteen years, we have seen a wave of new approaches. Many hundreds of impact evaluations and a number of systematic reviews have been implemented. New approaches involving complexity and qualitative methods have been developed and are increasingly applied. In short, the current set of evaluation criteria has not prevented the development of new methods and experimentation.

29. In the context of a debate around the criteria, it may also be noted that Robert Picciotto in his article “The logic of development effectiveness; is it time for the broader evaluation community to take notice?” (published in Evaluation 2013) argues that the DAC criteria forged in the development field would be equally serviceable in other domains of evaluation practice. The DAC criteria are seen as useful and valid for evaluation at project, country, regional/global levels and fit for purpose not only in development but also for developed countries. The article argues for some refinements, notably related to coherence, and also notes that the criteria have proved resilient to shifts in policy doctrines and have demonstrated their worth in diverse operating environments.

30. It may also be noted that a new generation of evaluators have come on board who are now contributing to theory, methods and new applications. As they have not been party to developing some of the current sets of norms, it is likely that this group would like to contribute to inform the debate and become a part, so as to create ownership.

3. Adapting the DAC evaluation criteria

The communique of the recent DAC High Level Meeting, adopted on 31 October 2017, included an encouragement to explore adapting the criteria. As stated in para 23: “We also encourage the DAC and its respective subsidiary bodies to explore adapting the five key evaluation criteria to program evaluations in line with the 2030 agenda”.

At the 21st meeting of the Evaluation Network members agreed to start a process to explore adapting the Evaluation Criteria.

3.1. Process

As with other norms and standards, revising the DAC criteria will involve discussions within the network, but also with other providers of development co-operation, with partners and civil society organisations. A consultative process ensures a better product with stronger ownership, and the network cannot work in isolation to develop any new criteria, particularly as the current use is spread much beyond the network.

Consequently, there needs to be broad consultations. As part of such outreach, it was suggested to try to piggyback on upcoming international meetings in development evaluation and insert or add a session on the criteria. This would enable consultations at a lower cost than organising specific workshops, although this may also be an option, particularly if members would be willing to host such workshops.

As part of the consultations process, the Network could invite external comments on blog posts and online platforms.
3.2. Timeline

A possible timeline is as follows:

- **Early 2018** – a smaller meeting of interested parties of the Network and some experts to review and discuss in more detail
- **Spring 2018** – survey of members and consultations with external stakeholders
- **June 2018** – 22nd Evaluation network meeting – progress report and further discussion
- **Autumn 2018** – further consultations with partners and civil society
- **Winter 2018** – circulation of final draft for comments
- **End 2018** – final revisions
- **February 2019** – approval at the 23rd EvalNet meeting

The sequences need to be further clarified as the schedule of international evaluation events during 2018 evolves, and based on member’s willingness to support the process.

Based on earlier normative work, this would be a comparatively rapid process. The glossary took three years, the standards included a three year testing period requested by members, and the principles took approximately two years to reach agreement on. While in principle the criteria are more limited in scope, opening a discussion on them will involve a number of considerations, as illustrated in the previous section. The criteria are also widely used and swiftly changing them without due consultative processes would not be in line with the principles of partnership.

The work on adapting the criteria should keep in mind the need for clarity and for simplicity as the criteria will be used in practical work by both experienced and novice evaluation managers. The final product should go beyond the criteria and provide a user guide that clarifies how to use the criteria in an appropriate manner and not in a mechanistic way. The reality of situations in field-based evaluation and for head offices need to be kept in mind.

3.3. Expected output and outcomes

The direct output would be an adopted criteria with a user guide to explain their appropriate use and application.

The outcome would be higher quality evaluations which apply a set of criteria in evaluations in an appropriate manner, also enabling collaborative work and facilitating comparison and synthesis studies.