

## **DAC Network on Development Evaluation**

### **Summary of the workshop on OECD DAC Evaluation Criteria: Progressing the dialogue**

**15 March 2018, Paris, OECD Boulogne**

Following the 21<sup>st</sup> OECD DAC EvalNet meeting of 14<sup>th</sup> November 2017, the subsequent workshop in March 2018, sought to ‘Progress the Dialogue’ around the possible adaptation of the five main evaluation criteria. The workshop’s specific objectives were:

- To identify key opportunities and challenges posed by the new development agenda to each of the five existing evaluation criteria and their definitions
- To identify areas of the new development agenda currently uncaptured in the five existing evaluation criteria
- To develop a collective sense of the task, and awareness of the roadmap going forward.

Hans LUNDGREN, Head of Unit, Evaluation, [hans.lundgren@oecd.org](mailto:hans.lundgren@oecd.org)

Julia BETTS, Consultant, [julia@jbetts.net](mailto:julia@jbetts.net)

## Draft Summary Record of the Workshop: OECD DAC Evaluation Criteria: Progressing the dialogue

Thursday 15 March 2018

The meeting was chaired by Per Øyvind Bastoe.

### Item 1. Implementation of the criteria

1. Hans Lundgren presented broad issues identified to date around the implementation of the current criteria. These included stressing the *merits* of the current framework, including a unifying language and ‘brand’ for the evaluation field, simplicity and their limited number; as well as clarifying for external audiences the evidence-based nature of evaluation.
2. *Challenges* identified to the current set of criteria included: the interpretation of the criteria as implying a linear logic; the limited ability to reflect complexity and synergetic effects; and gaps in some key themes raised under Agenda 2030.
3. *Implementation weaknesses* included: formulaic and un-contextualised implementation; definitions which do not reflect current developing thinking; and the fact that the criteria are currently most suited for project and programme level, being less useful for strategy/policy/corporate evaluations. Specific limitations were also identified for individual criteria.
4. Participants commented on the *strengths* of the current criteria; including the fact that they are already applied within national evaluation systems; the importance of their neutrality; and the ability of relevance to encompass design dimensions and investment prioritisation. However, *additional weaknesses* were also identified in the implementation of the criteria. These included: jargon around the use of criteria, which can alienate policy/operational actors; the formulaic use of all criteria without adequately considering the specific context and needs of each evaluation; challenges around methods for assessing efficiency; and the lack of definition of the many different dimensions of relevance.
5. *Gaps* noted in the existing criteria included coherence and flexibility/ability to innovate, with possible lessons to be drawn from humanitarian evaluation; and for an integrated approach across the criteria. The need for stronger guidance on how to apply the criteria was also raised.

### Item 2. Broader considerations

6. Facilitated by Susanne Frueh, presenters Caroline Heider, Zenda Ofir and Marcus Palenberg presented their considerations of the current issues surrounding the criteria. Key issues raised included:
  - Caroline Heider recounted the experience of the IEG blog series on the evaluation criteria, which occurred in response to IEG’s responsibility to stimulate discussion and debate. She described the extensive reaction to the blog, with 20,000 views collectively, and which culminated in a dialogue piece, published at UNDP’s 5th International Conference on National Evaluation Capacities in Istanbul, October 2017. She observed that the criteria created a platform for the broad and diverse evaluation community to come together, with the criteria serving as a rallying point; whilst also recognising the importance of staying dynamic and keeping the criteria up to date. She emphasised the importance of

uniformity/comparability across criteria – although skilled evaluators can and should adapt the criteria to context. Seven key points raised included (i) Changes in the wider context, such as a recognition of complexity and the need for a systems analysis, which sit alongside the declining importance of aid; (ii) The need to engage with stakeholders across world; (iii) The role of the criteria in incentivising both positive and negative behaviours in evaluators; (iv) The need to consider whether and where new criteria are needed because of contextual change, for example the role of synergetic effects, which is important for the SDGs; (v) The need to consider revisions to the existing criteria e.g. whether relevance should be adapted to consider selectivity/strategic choice; (vi) The need to consider the application of the criteria – for example including guidance, training and other forms of support for the new generation of evaluators; and finally (vii) The importance of being open to connecting the evaluation community to the data science community, and overcoming dichotomies to make maximum use of the possibilities offered by ‘big data’, technology for data generation, analysis and presentation.

- Zenda Ofir referred to the importance and power of the current criteria, which have often come to determine evaluation questions, and raised the central question of whether and how evaluation criteria can support/enable development, particularly under the framework of the SDGs. As a point of departure, she argued that the evaluation criteria should be more powerful in supporting development that transforms, and positive development trajectories that sustain. She raised six key challenges that require addressing in any potential adaptation of the criteria, including: (i) Omissions/weaknesses in the theory of assessment; (ii) Belief in ‘stakeholder questions / interests’ (only) – without considering their meaningfulness for development; (iii) ‘Refinement’ which may transpire into incorporating many items under each criterion (particularly under the broad framework of the SDGs); (iv) Conceptualisation and implementation weaknesses of the current criteria; (v) The need to address the boundaries and limitations of evaluation practice; and (vi) The need to evaluate development through a Complex Adaptive Systems lens. She suggested a framework for development-focused evaluation criteria, including three different considerations: (i) A focus on the nature of development, especially seen as a complex adaptive system, with a set of core criteria which are not flexible; (ii) A focus on global, societal & organisational norms and mandates, which suggests a moderately flexible set of criteria and (iii) A focus on specific stakeholder interests, which suggests a totally flexible set of criteria. Potential models under each consideration were presented.
- Markus Palenberg discussed changing paradigms of cause and merit; from ‘owning results’ (an ‘intervention’ perspective) to ‘assisting change’ (a ‘systems’ view). His presentation noted that causal attribution is not always a good basis for attributing responsibility; whilst in turn, the attribution of responsibility requires more than causal attribution. He also discussed the challenges of assessing efficiency, with no agreed terminology and a discrepancy between the criterion and the OECD DAC’s Glossary. He described efficiency types by measure and purpose, and presented examples of three different levels of efficiency analysis, with Level 0 providing basic information, and fulfilling basic accountability requirements, to describe efficiency; Level 1 seeking to improve the efficiency of individual interventions and identify efficiency improvement potential in one intervention; and Level 2 aiming to select the most efficient interventions (rational decision-making) and compare the efficiency of entire aid interventions with alternatives or benchmarks.

7. Participants commented that the SDGS are extremely broad (a ‘Christmas tree’) and that evaluation criteria cannot necessarily cover everything (*‘If we want to eat an elephant, then we should do*

*it in small pieces*’). To attempt to be comprehensive would be to run the risk of generating superficial evaluation reports which lack new insights. Some participants proposed retaining only a few core criteria and then adding other key issues such as human rights, equity and other cross-cutting issues and well as coherence/co-ordination and resilience.

### Item 3. Fit for the future: Revisiting the existing evaluation criteria

8. This session was led by Jörg Faust and Ida Verspohl, who presented some, identified conceptual challenges to the existing criteria and their usage. These included changing normative agendas, including Agenda 2030 and the SDGs, international agreements and agenda such as the 2015 Paris agreement; the fragility agenda and the human rights agenda, as well as national interests. Challenges to evaluation included the need to encompass synergies and coherence; the multitude of involved actors; complexity; and methodological development. Global developments, including those in international relations, as well as thematic trends such as governance and state fragility and inequality and social pressure, human rights; and finally different scenarios, such as the crystallisation of a single priority vs a multitude of priorities, and prevailing national interests, also influence the development of the criteria.

9. Key conceptual challenges were identified to each of the five evaluation criteria; including:

- **Relevance** - Whose understanding of relevance? - as well as definitional issues, such as strategic relevance, relevance to needs; relevance to actions of others, relevance for what; and the need to clarify whether the criterion emphasises intention or accomplishment. Also the need to strengthen the perspective on added value /critical constraints as benchmarks;
- **Efficiency** - Weak conceptualisation/definition of the criterion, and the fact that social and environmental costs are not currently taken into account; the need to balance standard setting and feasibility;
- **Effectiveness** – The inability to consider synergetic effects and varied definitions of ‘outcomes’; assumptions around causality (outcomes-impact); challenges with ‘attribution/contribution’ and insufficient consideration of external effects;
- **Impact** - Unclear distinctions from “effectiveness” and assumptions around causality; insufficient linkage to sustainability; and a lack of consideration of negative social/environmental/ technological/economic impacts;
- **Sustainability** – Weak conceptualisation in relation to the SDGs and insufficient linkage to impact; an unclear distinction between sustainability during implementation, and the likelihood of sustainability in future; as well as a lack of clarity on the distinction between the durability of effects and sustainable development from a multidimensional perspective; and the need to analyse influencing factors in a more systematic manner, differentiating between capacity/ ownership challenges, and influenceable/non-influenceable factors.

10. Participants then discussed the individual evaluation criteria in groups, with two groups covering effectiveness and none efficiency. Key issues raised included:

- **Relevance** – The group noted that different dimensions of relevance require specification and definition – this includes strategic, policy, operational relevance etc. Diplomatic and humanitarian/development relevance can also be different. A proposal of ‘*significance*’ was made to replace relevance, to reflect the role/scope of the evaluation within the wider development landscape.
- **Effectiveness** - Participants noted that most challenges were related to implementation; with the main conceptual challenge being the limited requirement to assess the quality of targets

set. The criterion does not currently allow for changes in intended results and outcome in response to a changing context. Some participants felt that capacity building should also be part of effectiveness.

- **Impact** - Participants considered that the definition of impact is currently confused, with the difference between short-term and longer-term impacts not clearly defined. Participants reflected that the normative agendas shape desired impacts, which also require further definition. Implementation needs to consider both positive, negative and unintended impacts, and a clear link to sustainability should be made.
- **Sustainability** – Participants in this group reflected that sustainability is often only considered from a limited perspective, whilst economic, social and environmental aspects of sustainability remain important elements. The possibility of integrating sustainability into the other four criteria was raised, for example through a matrix with four objective criteria and then other normative criteria that could be applied more selectively.

#### Item 4. Potential additions to the criteria

11. This plenary discussion was facilitated by Wendy Asbeek Brusse. Participants voiced the need for a clear underlying logical framework for any adapted set of criteria, starting from what is intended to be achieved rather than by identifying new criteria as the first step. Some advised that the logic of universal applicability should be unpicked, including the characteristics of ‘good evaluation criteria’; these might include being neutral, implementable in every type of evaluation, and with the ability to contribute to accountability and/or learning.

12. Several participants felt that the current criteria should be retained, and cautioned against a full revision, remarking that complexity has always been part of development. However, participants recognised that definitions probably require amendment. Some participants felt that gender, environment and human rights may be important but they should not be criteria in themselves. Others felt that coherence and complementarity could be a means to an end to achieve other evaluation criteria, rather than being included as ‘new’ criteria in themselves. Overall, the need for criteria to stand the test of time was stated.

#### Item 5. Outreach and consultation

13. Chaired by Per Øyvind Bastoe, in this session participants made a range of suggestions to broaden the proposed consultation. Stakeholder additions included:

- Réseau francophonie d’évaluation.
- The African Evaluation journal.
- Evaluation contractors and suppliers
- National evaluation functions e.g. in Africa
- UNEG meetings
- International group for supreme audit institutions- performance audit.

14. Caroline Heider shared experience from the World Bank on running a consultation platform, and offered to share expertise and ideas to broaden the consultation process.

---

## Item 6. Next steps

15. The process from here was described by Hans Lundgren, with key milestones including:
- April-May 2018: Pilot test and launch of survey with members and set up of website for online consultations with external stakeholders;
  - June 2018: 22nd meeting DAC EvalNet - update and further discussion;
  - July-October 2018: Wider consultations with external stakeholders (potential opportunities : AfDB Evaluation week, Asia Evaluation week, Evalpartners, EES) and continued online consultations
  - September -November 2018: Analysis of results of various consultative processes (interviews, workshops and conferences, survey, online consultations); drafting of adapted version of criteria and work on user guidance
  - November-December 2018: Draft version of adapted/new DAC Evaluation criteria and of user guidance circulated for comments
  - January 2019: Final draft version of criteria
  - February 2019: Approval at the 23rd meeting of DAC EvalNet
  - March 2019: Publication and communication on new criteria and user guidance.