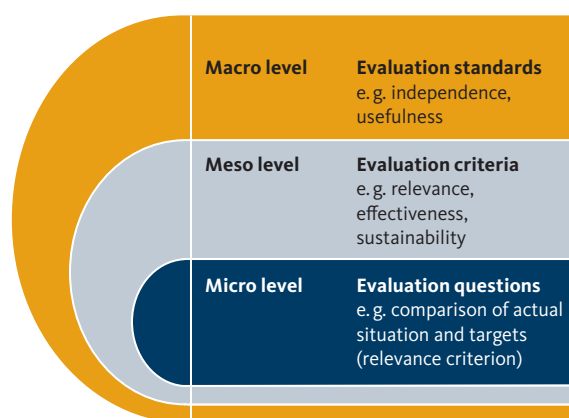


Figure on Box 1



Source: own figure.

BOX 1: Evaluation Frameworks: standards, criteria and guiding questions

In the field of DC, particular evaluation questions and methodological designs are typically embedded in a normative framework. Evaluation standards are generally located at the macro level of such a framework or evaluation regime. Standards specify key aspects of 'good' evaluation, e.g. independence, usefulness, accuracy, transparency or ethical responsibility. The list of criteria under review here can be placed at the meso level and provides content direction for the overall design of an evaluation. Thus, the criteria list is made to help structuring the focus of the particular evaluation questions in terms of relevant success factors of development cooperation to be addressed by the evaluation. At the micro level, a wider number of guiding questions below each evaluation criterion allow for a more specific perspective related to a given criterion at the meso level (see Box 2).

1) Neutrality vs. content-related focus

As yet, no particular content-related aspects have been assigned to the criteria list drawn up in 1991 and the standards of the OECD on which the list is based; instead, they are of a fairly timeless nature. No explicit reference is made to the trend towards political and economic liberalisation of the 1990s or to the United Nations Millennium Development Goals of the 2000s.

Moreover, in contrast to more activist and developmental evaluation approaches, the OECD/DAC evaluation criteria comes as a more technical instrument for the learning and accountability function of evaluation, which aims at improving DC in an indirect manner.

Against this background, some critics suggest that the criteria set should serve the SDGs more directly and that evaluation should be used as a more active instrument with direct effects on development goals. However, integrating content-related developmental aspects poses serious risks. It could detract from efforts to keep evaluation as neutral as possible and hence contains the risk of indirectly damaging evaluation's credibility.

At the same time, the principles of the 2030 Agenda themselves open up new prospects that can be assessed without making too strong value judgements about development contents. In a revised list of criteria, key and innovative aspects of the 2030 Agenda could be included, e.g. through *mainstreaming* in all the criteria, new or revised examination questions, or by summarizing key findings related to the 2030 Agenda in a particular chapter of an evaluation. As a result, evaluations could be geared more heavily towards assessing the contribution of an intervention to the principles of the 2030 Agenda.

2) Comprehensive appraisal vs. feasibility

The current set of evaluation criteria attempts to provide an encompassing perspective on a development intervention. After considering the initial question as to appropriate objectives (*relevance*), the evaluation criteria analyse whether these objectives have been achieved (*effectiveness*), which changes have occurred (*impact*), whether resources were invested efficiently (*efficiency*) and in how far the observed changes have been persistent over time and consistent across different dimensions of development (*sustainability*).

Beyond these core criteria – sometimes called the big five – some donor organizations have already enhanced the list. For instance, in the version of the list adapted by the German Ministry of Economic cooperation and development (2006), a separate criterion on *coordination, coherence and complementarity* has been introduced. In the wake of the current debate, further criteria have been proposed to be added to the list. *Innovation* and *resilience* have been mentioned in addition to *coherence* as potential criteria, given that these aspects have become increasingly important for successful development cooperation. However,

having an increasingly large number of criteria makes it more difficult to carry out an appropriate in-depth assessment – not only because of limited resources but also because of increasing difficulties to draw clear lines between different criteria and their respective guiding questions. As is illustrated in Box 2, even the current set of “only” five criteria has created various technical challenges for implementing the evaluation criteria in a conceptually consistent and empirically manageable way. Adding further criteria may easily exacerbate the existing challenges. A reform of the list of criteria should therefore primarily concentrate on working on the existing lack of clarity and the existing overlaps between the criteria. If possible, it should attempt to integrate new important aspects at the micro level of the guiding questions. For instance, the question of promoting *resilience* through an intervention could be addressed under the umbrella of the sustainability criterion, while aspects related to *coherence* and *synergies* could be dealt with as examination questions under the relevance and impact criteria. Such a process of integrating aspects, however, would require sufficient resources and hence additional analytical investment in precise and coherent definitions of the existing criteria.

Box 2: Current debate on the challenges in designing and implementing the DAC criteria

The criterion of **relevance** examines the extent to which the topics and objectives of a development intervention are reflected by stakeholder and context characteristics and whether activities carried out are consistent with the objectives. With regards to practical implementation, critics raise the challenge of the large number of national and international agendas and objectives, impeding priority-setting along key deficits or positive dynamics. Due to this lack of orientation, the relevance of a measure is often rated as being above average.

The criterion of **effectiveness** entails a comparison between planned targets and objectives actually achieved. In practice, however, the levels examined often differ, such as the level of the direct outputs of an intervention against the level of the effects on the target group. Another realm of the debate criticizes that confounding factors are only rarely taken into account systematically.

The assessment of a measure's **impact** focuses on intended and unintended positive and negative changes as a result of a development intervention. From the evaluators' point

of view, assigning changes to an intervention at impact level still poses a serious challenge. Meta evaluations also show that unintended changes only quite rarely are recorded systematically.

An assessment of **efficiency** relates the costs entailed by a development intervention with its results achieved. The efficiency criterion usually aims to identify room for improvement in terms of resources and achievements. However, only rarely different approaches and instruments of DC are directly compared. Instead, due to methodological challenges and high resource requirements, efficiency analysis mostly takes the forms of analysing production efficiency.

The criterion of **sustainability** assesses to which extent results that have been achieved are maintained after the end of a development intervention. In addition, sustainability incorporates the idea of equity between present and future generations (or the interaction between social, economic and environmental impacts). Combining these elements in a single criterion, however, often leads to a situation where only one of the two aspects is taken into account in an evaluation, or to the two aspects not being dealt with in sufficient depth.

3) Standardisation vs. greater flexibility

Developing a set of coherent examination questions within each evaluation criterion provokes the issue of an appropriate degree of flexibility in the practical application criteria list. On the one hand, a certain amount of freedom for development organizations to define the most relevant guiding questions for their interventions creates flexibility in the application. Such greater flexibility allows an evaluation to contribute to learning and monitoring requirements that are of particular importance for an individual development intervention and take into account particularities of specialized development agencies.

At the same time, greater flexibility limits the comparability of evaluation results and hinders cross-organisational learning from evaluations. Overall, a reform of the criteria list should not neglect the advantages of standardisation and the resulting collective learning potential. This can be done on the one hand by applying more precise definitions to criteria and guiding questions, so that the evaluative work follows the same direction

in different organizations. At the same time, it would make sense to set different levels of priorities among a probably increasing number of guiding questions. In this regard, a recommendation as to which questions should be addressed as a matter of priority among a criteria's guiding questions could be made. Such a prioritization would increase inter-organizational comparability for the most important questions, while at the same time leaving sufficiently operational flexibility.

Conclusions

Despite its great achievements, the criteria list for the evaluation of development measures needs to be revised. Yet, in order to continue to harness and reinforce the existing advantages and functions of the criteria catalogue, any such revision should be more of a reform than a fundamental transformation. In this context, we advocate

- gearing the list towards the principles of the 2030 Agenda for Sustainable Development to a moderate degree in terms of content,
- rendering the existing criteria and guiding questions more precise while moderately adding aspects that have become relevant such as *resilience*, *synergies* or *coherence* and
- providing recommendations on giving priority to individual guiding questions, in order to promote comparability between evaluations across organisations, while still leaving scope for flexibility.

This kind of reform could strengthen the existing advantages of the list and promote overarching lessons learnt from evaluations without detracting from the recognisability and the reference function of the criteria catalogue. Such a reform might also increase the list's use outside the field of DC and hereby would do justice to the principle of universality enshrined in the 2030 Agenda.

References

OECD DAC (1991), *DAC Criteria for Evaluating Development Assistance*, Paris.

BMZ (2006), *Evaluierungskriterien für die deutsche bilaterale Entwicklungszusammenarbeit*, Bonn.

UN (2015), *Resolution 70/1 – Transforming our world: The 2030 Agenda for Sustainable Development*, New York.

Authors



Prof. Jörg Faust
Director of DEval



Ida Verspohl
Evaluator

The German Institute for Development Evaluation (DEval) is mandated by the German Federal Ministry for Economic Cooperation and Development (BMZ) to independently analyse and assess German development interventions. Evaluation reports contribute to the transparency of development results and provide policy-makers with evidence and lessons learned, based on which they can shape and improve their development policies.