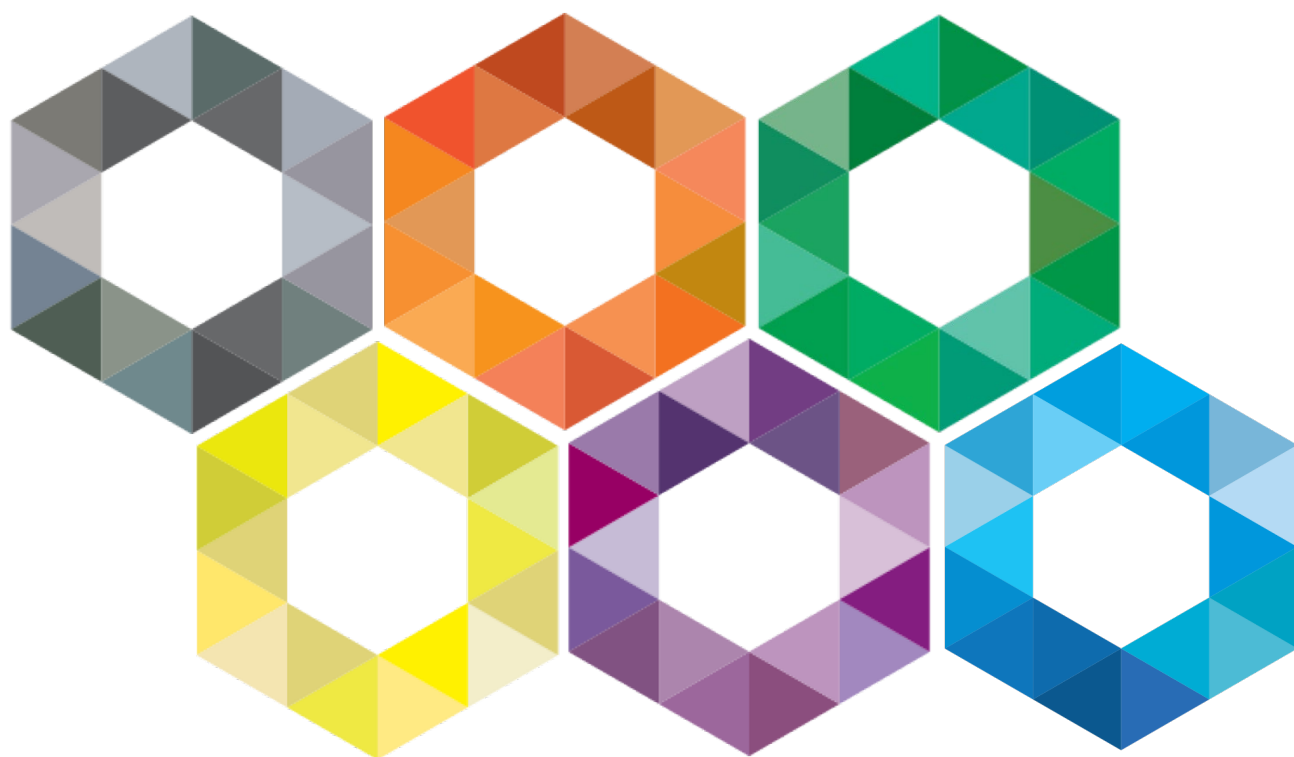


Better Criteria for Better Evaluation

Revised Evaluation Criteria
Definitions and Principles for Use

OECD/DAC Network on Development Evaluation



1. Background

1.1. Rationale for adaptation

1. The Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) first laid out the evaluation criteria (relevance, effectiveness, efficiency, impact and sustainability) in the 1991 OECD DAC *Principles for Evaluation of Development Assistance*, and later defined the terms in the 2002 *Glossary of Key Terms in Evaluation and Results Based Management*. These five criteria have come to serve as the core reference for evaluating international development and humanitarian projects, programmes and policies. Beyond development co-operation, evaluators and commissioners also use the criteria in other areas of public policy.

2. Building from learning gathered over 25 years of applying the criteria, the global evaluation community began discussing revisiting the criteria following the 2015 agreement of the 2030 Agenda for Sustainable Development, including the Sustainable Development Goals (the 2030 Agenda), and the Paris Agreement within the United Nations Framework Convention on Climate Change (Paris Agreement). In its High Level Communiqué of 31 October 2017, the OECD DAC decided to “explore adapting the five key evaluation criteria to program evaluations in line with the 2030 Agenda”. The DAC Network on Development Evaluation (EvalNet) – the DAC subsidiary body responsible for supporting evaluation and building the evidence base for policymaking and for learning – pursued this adaptation in 2018-2019.

3. The goal of the adaptation process was to take stock of lessons and experiences, in order to improve the criteria. Better criteria will support better evaluation. Better evaluation will contribute to better policies to advance the 2030 Agenda, achieve national contributions to the Paris Agreement, and other goals.

1.2. Consultation on the criteria: strengths and areas for improvement

4. The adaptation process included a far-reaching consultation, conducted May to November 2018, with a public survey, interviews, discussion at international meetings, and a review of literature. Based on initial drafts, consultations continued into late 2019. A summary of findings from the consultation is available.¹

5. The consultations found widespread support for and use of the criteria. Respondents clearly preferred to maintain the main body of the current set of evaluation criteria, recognising their universal acceptance and usefulness. Many respondents highlighted the value of the criteria in bringing standardisation and consistency to the evaluation profession and evaluative practice. It was also clear that there was a need for continued simplicity, by retaining a limited set of evaluation criteria and keeping the definitions coherent.

¹ See DAC Network on Development Evaluation (2018), OECD DAC Evaluation Criteria: Summary of Consultation Responses (November 2018). Available at: oe.cd/criteria

6. At the same time, requests were received for clarifications of certain concepts. Many pointed to challenges with the way the criteria are applied in practice. Particularly problematic is the tendency to cover too many criteria and questions. While the Quality Standards for Development Evaluation are clear that the use of all the criteria is not mandatory, and that other criteria may be used, in practice they can end up being applied mechanistically without sufficient consideration of the evaluation context and intended purpose. There was also concern that the original set of criteria did not adequately encompass the 2030 Agenda narrative and current policy priorities. Some felt the criteria were too project-focused and did not sufficiently address issues such as complexity and trade-offs, equity, and integration of human rights and gender equality. Many requested enhanced guidance on implementation of the evaluation criteria, to improve their use and to contribute to enhanced evaluation quality.

7. In addition, the consultations brought to light a number of weaknesses in evaluation practice and systems that go beyond the criteria and their definitions. In line with its mandate to support better evaluation, EvalNet is committed to working with partners in the global evaluation community to address these concerns, and is currently exploring options for additional work.

1.3. Key features of the adapted criteria

8. To respond to the request of the OECD DAC's High Level Communiqué, as well as to address findings from the consultation, EvalNet developed an adapted set of definitions and principles for the use of the criteria. EvalNet members and partners commented on two drafts, and then a series of webinars were held permitting an in-depth exchange and interaction on the definitions. International evaluation experts were also invited to provide feedback on drafts.

9. Drawing on all of this input, the adapted criteria have the following features:

- New and improved definitions of the original five criteria: Preserving and enhancing the key strength of conceptual clarity, definitions were refined and notes used to explain the concepts, while keeping the text as short and simple as possible.
- Adding one major new criterion – coherence – to better capture linkages, systems thinking, partnership dynamics, and complexity.
- Supporting use and addressing confusion by adding: an introduction on the intended purpose of the criteria; guiding principles for use; and an accompanying guidance to explain further the dimensions of each criteria, and how they apply to different evaluations (forthcoming).
- Ensuring applicability across diverse interventions, and beyond projects: Recognising the diverse range of development and humanitarian activities and instruments subject to evaluation – and the use of the criteria beyond development co-operation – the term “intervention” is used (rather than e.g. external funding, programme or project, as previously). References to “donors” have likewise been removed.
- Better responding to current policy priorities, including equity, gender equality, and the “leave no-one behind” agenda: The definitions of relevance and effectiveness in particular, encourage more in-depth analysis of equity issues. The criteria are useful for evaluating efforts (whether national, sub-national or international) aimed at achieving the Sustainable Development Goals as set out

in the 2030 Agenda and the Paris Agreement. At the same time, the criteria are high-level enough to ensure they will be widely applicable, and will remain relevant as policy priorities and goals change.

- Reflecting the integrated nature of sustainable development, the definitions and guidance for use promote an interconnected approach to the criteria, including examination of synergies and trade-offs.



2. Adapted Evaluation Criteria

2.1. Purpose of the evaluation criteria

10. The purpose of the evaluation criteria is linked to the purpose of evaluation. Namely, to enable the determination of the merit, worth or significance of an intervention.² The term “intervention” is used throughout this document to mean the subject of the evaluation (see Box 1). Each criterion is a different lens or perspective through which the intervention can be viewed. Together, they provide a more comprehensive picture of the intervention, the process of implementation, and the results.

11. The criteria play a normative role. Together they describe the desired attributes of interventions: all interventions should be relevant to the context, coherent with other interventions, achieve their objectives, deliver results in an efficient way, and have positive impacts that last.

12. The criteria are used in evaluation to:

- Support accountability, including the provision of information to the public;
- Support learning, through generating and feeding back findings and lessons.

13. The criteria are also used beyond evaluation – for monitoring and results management, and for strategic planning and intervention design.

14. They can be used to look at processes (how change happens) as well as results (what changed). All criteria can be used to evaluate before, during or after an intervention.³

Box 1. Evaluating interventions

We use the term intervention throughout the document to refer to the subject of the evaluation. Intervention encompasses all the different types of development and humanitarian efforts that may be evaluated using these criteria, such as a project, programme, policy, strategy, thematic area, technical assistance, policy advice, an institution, financing mechanism, instrument, or other activity. It includes development interventions, humanitarian aid, peacebuilding, climate mitigation and adaptation, normative work, and non-sovereign operations.

The criteria can be used to evaluate international co-operation activities, as well as the interventions of private sector actors, non-government actors, and national or local governments in domestic policy contexts.

² See reference to worth and significance in OECD DAC (1991) Principles of Evaluation of Development Assistance.

³ However, the way the criteria are defined here reflects the dominant practice of interim, final and ex-post evaluation.

2.2. Principles for use

15. The following principles guide the use of the criteria. Further advice and examples are provided in the accompanying guidance (forthcoming). In addition, the OECD DAC's *Quality Standards for Development Evaluation* describe standards for evaluation planning and implementation. It is important that the definitions of the criteria be understood within a broader context, and read in conjunction with other principles and guidance on how to conduct evaluations in ways that will be useful and of high quality.

Principle One

The criteria should be applied thoughtfully to support high quality, useful evaluation. They should be contextualized – understood in the context of the individual evaluation, the intervention being evaluated, and the stakeholders involved. The evaluation questions (what you are trying to find out) and what you intend to do with the answers, should inform how the criteria are specifically interpreted and analysed.

Principle Two

Use of the criteria depends on the purpose of the evaluation. The criteria should not be applied mechanistically. Instead, they should be covered according to the needs of the relevant stakeholders and the context of the evaluation. More or less time and resources may be devoted to the evaluative analysis for each criterion depending on the evaluation purpose. Data availability, resource constraints, timing, and methodological considerations may also influence how (and whether) a particular criterion is covered.⁴

2.3. Definitions and explanations of changes

16. The following section defines each criterion. Within the definitions, notes are used to clarify concepts. Explanatory boxes describe changes made to the original definitions in the *Glossary of key terms in evaluation and results based management* (OECD, 2002). A second edition of the *Glossary* is being produced and will serve as a useful reference point for readers as it defines many terms used in the document – including intervention, results, output, outcome, and objective.

⁴ Evaluability assessments (carried out before an evaluation begins) can be useful in setting realistic expectations for what information the evaluation can provide, what evidence can be gathered, and how the evaluation will answer questions.

RELEVANCE: IS THE INTERVENTION DOING THE RIGHT THINGS?

The extent to which the intervention objectives and design respond to beneficiaries',⁵ global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change.

Note: "Respond to" means that the objectives and design of the intervention are sensitive to the economic, environmental, equity, social, political economy, and capacity conditions in which it takes place. "Partner/institution" includes government (national, regional, local), civil society organisations, private entities and international bodies involved in funding, implementing and/or overseeing the intervention. Relevance assessment involves looking at differences and trade-offs between different priorities or needs. It requires analysing any changes in the context to assess the extent to which the intervention can be (or has been) adapted to remain relevant.

Box 2. Explanation of changes made to the definition of relevance

We add "design" to capture other quality elements related to the relevance of the intervention – i.e. is the intervention well-designed to address relevant priorities / needs (i.e. its objectives, underlying theory of change, its theory of action, its *modus-operandi*, risk analysis, context analysis, etc.). This is intended to deepen the analysis of relevance, while the achievement of objectives is assessed under effectiveness, and how well the intervention is implemented is evaluated under efficiency.

The note describes key elements of the context to which an intervention should be sensitive, in order to encourage deeper analysis of the contextual conditions.

We add beneficiaries' "priorities", to clarify the importance of affected people (not just donors or governments) deciding what is most important/urgent. The focus on beneficiaries also reflects the current policy priority to leave no-one behind. We use the term "needs", which is more easily understood than "requirements" (requirements being sometimes confused with legal requirements).

During the consultation we heard some discomfort with the word "beneficiaries", because it may imply a passive recipient role, or because it assumes people are benefiting, when that may not be the case. We retain the term here because it has specific meaning for evaluating relevance in regards to the intentions of reaching certain people and *for whom* the intervention should be relevant. To avoid confusion, we define the term in a footnote and offer alternatives.

We retain "global priorities", which are today reflected in the 2030 Agenda and the Paris Agreement. We do not explicitly mention these in the definition, because policy priorities will change, and the relevance criterion will be applied to different evaluands.

We make the time dimension explicit in the definition (previously this was included in a note), because relevance can be evaluated both at the time of the intervention design and later. This is useful to include because adapting and responding to complex or changing circumstances is important for effectiveness. The previous version of relevance was often used as a one-off assessment against policy priorities.

To better reflect the use of the criteria beyond development co-operation, we remove the term "donors" and add a note to explain that partner/institution includes funding, implementing and oversight partners (which could be local, multilateral, private, etc.).

⁵ Beneficiaries is defined as, "the individuals, groups, or organisations, whether targeted or not, that benefit directly or indirectly, from the development intervention." Other terms, such as rights holders or affected people, may also be used.

We do not include the concepts of participation and ownership in the definition because these are factors that affect relevance (as well as effectiveness and sustainability), rather than dimensions of the criterion itself. These concepts will be discussed in the guidance.

Glossary definition of relevance: “The extent to which the objectives of a development intervention are consistent with beneficiaries’ requirements, country needs, global priorities and partners’ and donors’ policies. Note: Retrospectively, the question of relevance often becomes a question as to whether the objectives of an intervention or its design are still appropriate given changed circumstances”

COHERENCE: HOW WELL DOES THE INTERVENTION FIT?

The compatibility of the intervention with other interventions in a country, sector or institution.

Note: The extent to which other interventions (particularly policies) support or undermine the intervention, and *vice versa*. Includes internal coherence and external coherence: **Internal coherence** addresses the synergies and interlinkages between the intervention and other interventions carried out by the same institution/government, as well as the consistency of the intervention with the relevant international norms and standards to which that institution/government adheres. **External coherence** considers the consistency of the intervention with other actors’ interventions in the same context. This includes complementarity, harmonisation and co-ordination with others, and the extent to which the intervention is adding value while avoiding duplication of effort.

Box 3. Explanation of the addition of coherence and of its definition

This criterion captures a perspective that was not covered previously. A lack of coherence can lead to duplication of efforts and undermine overall progress. Adding the criteria will help raise the bar on analysis of these important issues. Including coherence also incentivises evaluators to understand the role of an intervention within a particular system (organisation, sector, thematic area, country), as opposed to taking an exclusively intervention- or institution-centric perspective.

In today’s context, there is a need for greater attention to coherence, with an increased focus on the synergies (or trade-offs) between policy areas and growing attention to cross-government co-ordination, particularly in settings of conflict and humanitarian response, and to address the climate emergency. In addition, the sources (both international and domestic) of financing for sustainable development are diversifying.

The reference to “international norms and standards” encourages analysis of the consistency of the intervention with the actor’s own commitments under international law or agreements, such as anti-corruption statutes or human rights conventions. This applies to those agreements to which the entity has already committed, and is therefore covered under internal coherence. Previously, this type of coherence has not often been sufficiently analysed. International norms and standards may also be assessed under relevance from the viewpoint of responsiveness to global priorities, which is a complementary angle.

EFFECTIVENESS: IS THE INTERVENTION ACHIEVING ITS OBJECTIVES?

The extent to which the intervention achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups.

Note: Analysis of effectiveness involves taking account of the relative importance of the objectives or results.

Box 4. Explanation of changes to the definition of effectiveness

We clarify that effectiveness should analyse progress towards objectives along the results chain / causal pathway. In contrast to impact, which looks at higher-order effects and broader changes to which an intervention may be contributing, effectiveness is concerned with more closely attributable results.

We add “results” and “differential results” to open the door to asking important questions about the distribution of results across different groups, and to look beyond intended objectives. This is in line with the policy priority to leave no-one behind. It encourages evaluators to examine equity issues and results for groups that have been marginalized, while not assuming that equity is an objective of the intervention.

The phrasing allows flexibility for evaluations to focus on the objectives and/or results that are of most interest.

We retain the reference to “relative importance” in the note because one should weigh the importance of the achieved/not achieved/expected objectives and results when drawing conclusions about overall effectiveness. On the other hand, under relevance, we assess the extent to which the intervention is a priority for key stakeholders. The overall significance of the intervention effects is captured under impact.

Glossary definition of effectiveness: “The extent to which the development intervention’s objectives were achieved, or are expected to be achieved, taking into account their relative importance. Note: Also used as an aggregate measure of (or judgment about) the merit or worth of an activity, i.e. the extent to which an intervention has attained, or is expected to attain, its major relevant objectives efficiently in a sustainable fashion and with positive institutional development impact.”

EFFICIENCY: HOW WELL ARE RESOURCES BEING USED?

The extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way.

Note: “Economic” is the conversion of inputs (funds, expertise, natural resources, time, etc.) into outputs, outcomes and impacts, in the most cost-effective way possible, as compared to feasible alternatives in the context. “Timely” delivery is within the intended timeframe, or a timeframe reasonably adjusted to the demands of the evolving context. This may include assessing operational efficiency (how well the intervention was managed).

Box 5. Explanation of changes to the definition of efficiency

We clarify that efficiency may look at inputs relative to the entire results chain (outputs, outcomes and impacts), in line with good evaluative practice. It is recognised that analysing the entire results chain, and in particular looking at the efficiency of inputs to impacts, is methodologically challenging. Benchmarking data are often missing. Nonetheless, this is an area of inquiry of interest, so we retain the reference to impacts. This will allow users the flexibility to focus the analysis of efficiency on the part of the results chain that is most relevant for their evaluation.

Adding “as compared to feasible alternatives in the context” reinforces the idea that valid cost-benefit analysis requires comparing the value of the intervention with pertinent counterfactuals: the cheapest intervention may not necessarily be the best option if it does not generate adequate benefits.

We add the timeliness dimension because this important aspect of intervention value is sometimes missed.

We make explicit the reference to operational efficiency, because implementation processes and management issues are of great interest to evaluation stakeholders. This is useful, for example, in examining whether co-ordination mechanisms have improved coherence while increasing transaction costs.

Glossary definition of efficiency: “A measure of how economically resources/inputs (funds, expertise, time, etc.) are converted to results.”

IMPACT: WHAT DIFFERENCE DOES THE INTERVENTION MAKE?

The extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.

Note: Impact addresses the ultimate significance and potentially transformative effects of the intervention. It seeks to identify social, environmental and economic effects of the intervention that are longer term or broader in scope than those already captured under the effectiveness criterion. Beyond the immediate results, this criterion seeks to capture the indirect, secondary and potential consequences of the intervention. It does so by examining the holistic and enduring changes in systems or norms, and potential effects on people's well-being, human rights, gender equality, and the environment.

Box 6. Explanation of changes to the definition of impact

We have introduced “higher-level” to capture the significance, the scope, and the transformative nature of the effects. The note explains this meaning. This meaning has policy relevance in a context where the 2030 Agenda and Paris Agreement call for transformative change.

The definition is now more in line with common usage of the word impact for meaningful or important changes.

We recognise that the term impact has sometimes been confused with the notion of valid attribution adopted by some users (and as commonly used in “impact evaluation”), but we consider that valid attribution is embedded in all the criteria.

We remove “direct and indirect” and “primary and secondary” because direct and primary effects are covered by effectiveness. The note further helps to clarify the difference between effectiveness and impact.

Glossary definition of impact: “Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.”

SUSTAINABILITY: WILL THE BENEFITS LAST?

The extent to which the net benefits of the intervention continue, or are likely to continue.

Note: Includes an examination of the financial, economic, social, environmental, and institutional capacities of the systems needed to sustain net benefits over time. Involves analyses of resilience, risks and potential trade-offs. Depending on the timing of the evaluation, this may involve analysing the actual flow of net benefits or estimating the likelihood of net benefits continuing over the medium and long-term.

Box 7. Explanation of changes to the definition of sustainability

The definition is concise and the explanatory note makes clear that sustainability has various dimensions (financial, economic, social and environmental).

The original definition was too donor centric, and focused only on external funding (major development assistance), which did not sufficiently reflect the context for development evaluation today. In addition, it may be useful to evaluate sustainability even while funding or activities are ongoing. Removing the reference to external funding will support evaluation of sustainability in a variety of contexts.

The note also encourages analysis of potential trade-offs, and of the resilience of capacities/systems underlying the continuation of benefits. By contrast, the previous definition included repetitive phrasings (“continued long-term benefits” and “resilience to risk”).

We retain the term “net” benefits to focus on the continuation of the overall value of the intervention, taking into account any ongoing costs related to the intervention.

Glossary definition of sustainability: “The continuation of benefits from a development intervention after major development assistance has been completed. The probability of continued long-term benefits. The resilience to risk of the net benefit flows over time.”



Abstract

This document describes how the OECD DAC Network on Development Evaluation (EvalNet) revisited the definitions and use of the OECD DAC evaluation criteria in 2018-2019. The document lays out adapted definitions for relevance, effectiveness, efficiency, impact and sustainability, and for one new criterion, coherence. The document describes how the criteria should be used thoughtfully, and adjusted to the context of the intervention and the intended users' needs.

These revised definitions and principles for use are the result of a global consultation on the criteria and a review of how they are used in evaluation and beyond. Following the consultation, members of EvalNet and outside evaluation experts discussed the concepts in depth and reviewed several drafts. The adapted definitions are clearer and will support more rigorous, nuanced analysis, including of equity issues and synergies, in line with current policy priorities. This adaptation also addresses confusion, by adding an introduction on the intended purpose of the criteria and guiding principles for use. Detailed guidance on the application of the criteria is to be provided separately, after adoption.

This document was APPROVED by the DAC Network on Development Evaluation on 20 November 2019; it was ADOPTED and declassified by the DAC at its meeting on 10 December 2019.