

3 Using the evaluation criteria in practice

This chapter considers the practical aspects of how the criteria should be used in evaluation design and implementation. It reviews the two main principles that should guide the application of the criteria. The chapter begins by exploring different approaches that can support thoughtful application of the criteria. It then addresses how the criteria can be applied within different institutional settings with different strategic priorities, ways of working and cultures. It outlines how the criteria can prompt evaluators to consider differential experiences and impacts by applying a gender lens. Finally, it examines how the criteria can help evaluators and evaluation managers to work in ways that support achievement of the Sustainable Development Goals and the wider 2030 Agenda. The chapter includes a range of practical examples illustrating how the criteria have been applied to various evaluations.

To support evaluators and those involved in designing or managing interventions in developing evaluations that are helpful and appropriate to different contexts and stakeholders, the following two principles have been developed to guide their use. To avoid that the criteria are applied in ways that are mechanistic – discouraging critical thinking, creativity and ownership of participants – these principles should accompany the criteria whenever they are used (OECD, 2019^[1]):

- **Principle One:** The criteria should be applied thoughtfully to support high-quality, useful evaluation.
- **Principle Two:** Use of the criteria depends on the purpose of the evaluation.

The following section elaborates on these two principles and outlines additional key concepts for working with the criteria, including how to adjust the criteria to specific contexts, how to examine the criteria at different moments in time, and how the criteria relate to one another.

Applying criteria thoughtfully

Principle one stresses that the criteria should be used thoughtfully. In practice, this means thinking critically about which criteria are most useful to support high-quality, useful evaluation that will be valuable to the intended users. Box 3.1 provides an example of how the criteria can be applied thoughtfully when evaluating an intervention.

Considering the following six aspects and their related questions will assist evaluators in the thoughtful application of the criteria:

- **Context:** What is the context of the intervention itself and how can the criteria be understood in the context of the individual evaluation, the intervention and the stakeholders?
- **Purpose:** What is the evaluation trying to achieve and what questions are most useful in pursuing and fulfilling this purpose?
- **Roles and power dynamics:** Who are the stakeholders, what are their respective needs and interests? What are the power dynamics between them? Who needs to be involved in deciding which criteria to apply and how to understand them in the local context? This could include questions about ownership and who decides what is evaluated and prioritised.
- **Intervention (evaluand):** What type of intervention is being evaluated (a project, policy, strategy, sector)? What is its scope and nature? How direct or indirect are its expected results? What complex systems thinking are at play?
- **Evaluability:** Are there any constraints in terms of access, resources and data (including disaggregated data) impacting the evaluation, and how does this affect the criteria?
- **Timing:** At which stage of the intervention’s lifecycle will the evaluation be conducted? Has the context in which the intervention is operating changed over time and if so, how? Should these changes be considered during the evaluation? The timing will influence the use of the criteria as well as the source of evidence.

Box 3.1. Thoughtful use of the criteria: Evaluation of Norway’s civil society grant for developing countries

This evaluation examined the Norwegian civil society grant, which aims to strengthen civil society in developing countries and thus contribute to a stronger civil society with the ability and capacity to promote democratisation, improve human rights and reduce poverty.

This evaluation applied the criteria in a “thoughtful” and flexible manner, interpreting them in a specific way reflecting the nature of the intervention (partnerships and capacity building with civil society organisations) and taking into account Norwegian priorities and the context. The evaluation uses the following criteria, adapted in certain respects:

- impact – defined in relation to key outcomes such as democracy, income opportunities leading to reduced poverty
- relevance – no changes
- effectiveness – here defined in relation to the specific objectives of the partnership (i.e. service delivery, advocacy and capacity strengthening). Includes some mid-term outcomes such as creating space for civil society
- sustainability – no changes.

An additional criterion – “value added” – is also used. This is defined as professional competence, organisational and financial competence, and networking competence. It relates to issues that, in other contexts, might be covered under the criteria of effectiveness and efficiency.

Source: Helle et al. (2018^[2]), *From Donors to Partners? Evaluation of Norwegian Support to Strengthen Civil Society in Developing Countries through Norwegian Civil Society Organisations*, https://www.norad.no/globalassets/filer-2017/evaluating/1.18-from-donor-to-partners/1.18-from-donors-to-partners_main-report.pdf

Adapting the criteria to the evaluation’s purpose

The most important aspect of deciding how to use the criteria is relating them to the aim of the evaluation and its context and then building the evaluation criteria and questions around this purpose. Box 3.2 gives two examples of how the purpose of an evaluation can be defined.

The criteria are not intended to be applied in a standard, fixed way for every intervention or used in a tick-box fashion. Indeed the criteria should be carefully interpreted or understood in relation to the intervention being evaluated. This encourages flexibility and adaptation of the criteria to each individual evaluation. It should be clarified which specific concepts in the criteria will be drawn upon in the evaluation and why.

The purpose of the evaluation should be carefully and clearly defined. Stakeholders involved in the evaluation should be included at this stage to ensure that they understand the goal of the evaluation and how it will be used.

Key questions to look at when determining the purpose of the evaluation include:

1. What is the demand for an evaluation, who is the target audience and how will they use the findings?
2. What is feasible given the characteristics and context of the intervention?
3. What degree of certainty is needed when answering the key questions?
4. When is the information needed?

5. What is already known about the intervention and its results? Who has this knowledge and how are they using it?

Quality and ethical standards should inform subsequent thinking about methodological approaches, design, implementation and management of the evaluation process.

When adjusting the criteria to a specific evaluation, it is also important to delve into causality and gauge the extent to which the evaluation will be able to attribute effects to the intervention being assessed. These considerations can help manage stakeholder expectations and determine which criteria will be covered in depth. This is most crucial for effectiveness and impact which is discussed in both sections below, but it is also indirectly applicable to, and may feed into, other criteria. For example, efficiency and sustainability could use the actual (or projected) benefits attributed to the intervention as assessed under effectiveness and impact. An evaluability analysis is a useful tool to check that the purpose is understood in depth.¹ Further explanation on how to interpret each criterion is covered in Chapter 4.

Box 3.2. Defining the purpose of the evaluation: Examples from Uganda, Kenya and Sweden

The Danish International Development Agency's (Danida) evaluation of water, sanitation and environment programmes in Uganda

The Evaluation Department of the Danish Ministry of Foreign Affairs commissioned an independent evaluation of Danish initiatives to improve water, sanitation and environment in Uganda over the period 1990-2017.

The evaluation's objectives were:

- to document the results and achievements in the sub-sectors
- to analyse the "value added" from Danida's support to the sub-sectors
- to extract lessons learned.

To those ends, the evaluation focused on assessing effectiveness in two stages: 1) at the sector and cross-cutting level, and 2) at sub-sector level. The evaluation also assessed the value of the initiatives and looked at the sustainability of these results.

United Nations High Commissioner for Refugees (UNHCR) and Denmark joint evaluation

The joint evaluation by the UNHCR and Denmark of the refugee response in and around Kalobeyei, Kenya, defines the purpose in the following way:

"The main purpose of the evaluation is to contribute to learning about the integrated settlement model in and around Kalobeyei. By documenting lessons learned from a concrete effort to link humanitarian and long-term development assistance, the intention is to provide evidence of the potentials and challenges of designing and implementing an integrated solutions model."

Evaluation of the Swedish International Development Cooperation Agency's (Sida) support to peacebuilding in conflict and post-conflict contexts

The evaluation report gives a useful example of how the evaluation reflects the purpose. It gives the following explanation:

The purpose or intended use of the evaluation is to systematise lessons learned from peacebuilding practice. The evaluation will serve as an input to the process of conceptualising and developing the peacebuilding approach used by the Swedish International Development Cooperation Agency (Sida). These approaches will in turn influence strategic planning and design of future Sida support in contexts

affected by conflict. Moreover, the evaluation is expected to contribute to increased understanding of peacebuilding as a concept and practice.

The specific objective of the evaluation was to evaluate how Sida has approached peacebuilding on the strategic level in different contexts. To that end, the evaluation paid particular attention to four of the criteria:

- relevance of Sida's peacebuilding work vis-à-vis contextual and beneficiary needs, and Sida's policy priorities
- effectiveness of its peacebuilding work in terms of contributing to overall peacebuilding objectives and its ability to provide a conducive framework for its partners' peacebuilding work
- impact of its peacebuilding work
- sustainability of its peacebuilding work.

Across all four criteria, special emphasis was given to results related to gender equality, female empowerment and rights, as well as inclusiveness of marginalised groups and ethnic minorities. For all four case study countries, the evaluations focused on aspects of marginalisation that were linked to peacebuilding and which have appeared in the literature and interviews. Marginalised groups were primarily considered from the perspective of ethnicity or as a consequence of being a minority. This is because, in all four cases, ethnic factors formed a major part of the root causes of conflict.

Sources: Danida (2019^[3]), *Evaluation of Water, Sanitation and Environment Programmes in Uganda (1990-2017)*, <http://www.oecd.org/derec/denmark/denmark-1990-2017-wash-environment-uganda.pdf>;

ADE (2019^[4]), *Joint Evaluation of the Integrated Solutions Model in and around Kalobeyei, Turkana, Kenya*, https://um.dk/en/danida-en/results/eval/eval_reports/publicationdisplaypage/?publicationid=dd54bef1-1152-468c-974b-c171fbc2452d;

Bryld (2019^[5]), *Evaluation of Sida's Support to Peacebuilding in Conflict and Post-Conflict Contexts: Somalia Country Report*, https://publikationer.sida.se/contentassets/1396a7eb4f934e6b88e491e665cf57c1/eva2019_5_62214en.pdf

Understanding the time dimension

The criteria can be applied to different points in time. Each of the criteria can be used to evaluate before, during, or after an intervention. Likewise, they can be assessed during different moments in the intervention's lifecycle. However, the interpretation of the criteria and sources of evidence may be different at different points in time. For example, before an intervention has taken place, effectiveness and sustainability would be projections; whereas after the intervention, more data will be available from which to draw more solid conclusions.

The criteria – and related evaluation questions – should therefore reflect the following two key aspects related to timing: 1) the point in the lifecycle of the intervention when the evaluation will take place; and 2) the stage of the intervention or point of the result chain on which the evaluation will focus.

The conceptual frame of each criterion does not change if an evaluation is taking place before, during or after an intervention. However, the data and evidence available to assess the criteria and the methods used do change. Evaluators should keep these differences in mind when discussing the (potential) findings with evaluation stakeholders as this may influence the perceived usefulness and credibility of the evaluation. For example, an ex-ante evaluation of sustainability (taking place before the intervention begins) could look at the likelihood of an intervention's benefits continuing by examining the intervention design and available evidence on the validity of the assumptions about the continuation of expected benefits. After the completion of the intervention, an evaluation of sustainability would look at whether or not the benefits did in fact continue, this time drawing on data and evidence from the intervention's actual achieved benefits.

When looking back in time, evaluations will have to take into account the context and data available at that time so as to make judgements based on reasonable expectations of what could or should have been done. It would be unfair to judge the actions of past programme designers based on information available today that was not known to them at the time. However, many evaluations have found that available information was not fully utilised (where it could have been), which would have made the intervention more relevant. For example, local people were not sufficiently consulted and involved in the design of the intervention. Such an oversight could reasonably have been expected to be avoided and should therefore be flagged in an evaluation.

Using the criteria for better evaluation questions

Formulating good evaluation questions is a key part of the evaluation process and the use of the six criteria interacts with and supports the process of deciding on evaluation questions.² The process starts with a reflection on the purpose of the evaluation, how it will be used and by whom. An effective engagement with stakeholders through a well-designed participatory process can help evaluators and managers understand how they will use the evaluation. A deep understanding of the intervention and its context, its objectives and theory of change should complement this discussion with stakeholders. An inception phase can be used to explore these questions and generate (a small number of) key questions that the evaluation needs to address.

Following this initial step, the evaluation criteria then provide a tool for checking from different perspectives to see if anything has been missed, enabling further development and refining of the questions, which are essential to the evaluation design. This makes the process systematic and ensures that the evaluation is comprehensive. It is a crucial part of what constitutes a “better evaluation”. It is highly recommended that this question development phase be undertaken with a clear and coherent overall approach.

The institution commissioning the evaluation may have taken a particular decision to evaluate interventions according to certain criteria. In this instance, the institutional guidance should be followed alongside thoughtful application of the criteria to ensure consistency.

An example of this is the Sida’s evaluation manual, which provides examples of standard questions under each of the criteria (Molund and Schill, 2004^[6]). Other examples of how institutions interpret and unpack the criteria are available, such as the guidance and technical notes that have been developed for evaluation managers and evaluators at the World Food Programme (WFP, 2016^[7]).

Relationships between the criteria

The criteria comprise multiple lenses through which an intervention and its results can be viewed. They are interrelated, in that the concepts underpinning each one help to analyse complementary dimensions of the process of achieving results. For instance, effectiveness and impact look at different levels of the results chain, depending on how the objectives of the intervention have been defined. The two criteria are therefore interrelated along the causal chain.

The criteria often depend on each other. For example, an intervention that was not relevant to the priorities of the beneficiaries is unlikely to have the intended impact (unless there are some rather unusual separate channels of impact). An intervention that is poorly implemented (less effective) is also likely to be less sustainable. On the other hand, it is also possible that an intervention could be highly relevant yet ineffective. Or a highly coherent intervention could be very inefficient due to increased transaction costs. Evaluators should explore and reflect on relationships and synergies between different criteria, including considering if and how they are causally related.

Most evaluations draw conclusions based on the findings for each criterion, as well as an overall conclusion based on all criteria, sometimes using a numerical score to rate performance. In drawing conclusions about the intervention, and depending on the purpose of the assessment, evaluators should look at the full picture and consider how to appropriately weight all of the applied criteria. Criteria may be weighted with some institutions defining a dominant (“knock out”) criterion. If performance is not satisfactory on that criterion, no matter how well other criteria scored, the intervention will be considered unsuccessful (or if *ex ante*, would not be funded).

More specific linkages between each criterion are discussed in Chapter 4.

Choosing which criteria to use

As described above, the purpose, priorities, scope and context of the intervention and the evaluation will shape the relative focus on different criteria. Evaluators should consider the relative value each criterion will add. This would include two basic decisions: Is this criterion an important consideration for this evaluation? Is it feasible to answer questions about this criterion?

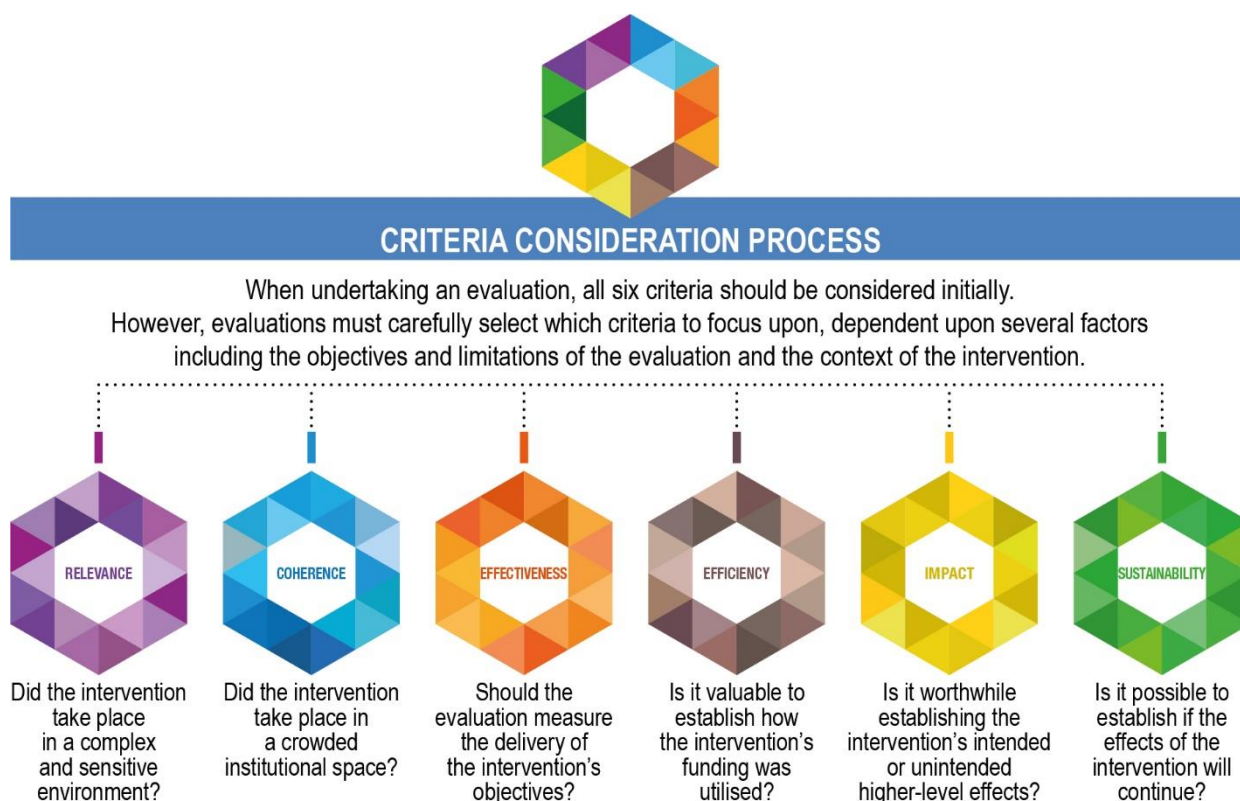
While users may be tempted to simply apply all six criteria regardless of context, the better approach (i.e. the one that is consistent with the original intent of the criteria and that leads to the highest quality evaluation) is that of deliberately selecting and using the criteria in ways that are appropriate to the evaluation and to the questions that the evaluation is seeking to answer.

To achieve this, ask questions such as:

- If we could only ask one question about this intervention, what would it be?
- Which questions are best addressed through an evaluation and which might be addressed through other means (such as a research project, evidence synthesis, monitoring exercise or facilitated learning process)?
- Are the available data sufficient to provide a satisfying answer to this question? If not, will better or more data be available later?
- Who has provided input to the list of questions? Are there any important perspectives missing?
- Do we have sufficient time and resources to adequately address all of the criteria of interest, or will focusing the analysis on just some of the criteria provide more valuable information?

It is important to strike a balance between flexibility (avoiding a mechanistic application of all the criteria) and cherry-picking (selecting only the easiest criteria, or those that are likely to generate positive results) when using the criteria. Notably, one should not shy away from answering critical questions on impact and coherence, even though these questions can be more challenging at times. Some points for consideration are shown in Figure 3.1 below; it should be noted that these are examples and not a comprehensive checklist.

Figure 3.1. Considering how to cover the criteria for a specific evaluation



A good knowledge of the stakeholders involved – in both the intervention and the evaluation – can help identify potential tensions between their different interests and priorities when it comes to the design and implementation of the evaluation. Beneficiaries may be most interested in understanding effectiveness (e.g. whether their children's health is improving through participation in a malnutrition treatment programme) while implementers may be more interested in understanding efficiency, with an eye to scaling up treatment to more families. In most cases, not all potentially interesting questions can be answered in a single evaluation and choices will have to be made. To increase the likelihood that questions that are left out of the evaluation will be covered elsewhere, it is good practice to document the process and outcomes of discussions about prioritising different stakeholder needs and deciding on evaluation questions.

Operationalising and adapting the criteria at an institutional level

The criteria definitions and this guidance provide a common platform and set of agreed definitions on which to build. However, tailoring them to the institutional context is crucial. Evaluators and evaluation managers often use the criteria on behalf of a development organisation, ministry, or other institution that has its own specific mandate, policy priorities, evaluation policy, standards and guidance – all as decided by their governing bodies.

When considering how to operationalise and apply the criteria, it is important that evaluators and commissioners carefully consider the organisation's strategic priorities, culture and opportunities as a background to decision making. The way certain terms – such as impact – are used varies, and it is important to pay close attention to the potential for confusion or misinterpretation of the criteria and their intended focus. This will assist evaluators and commissioners as they apply the criteria, maximise the use of the evaluation's findings and enhance relevance to the intended user's needs. Managers should encourage discussion between evaluators, commissioners and the target audience of the evaluation to

consider how the criteria should be applied and interpreted. Such a process can support the design of credible and timely evaluations that better meet the users' needs.

Methodological requirements set by an institution may also have a bearing on how the criteria are applied. Evaluators should refer to the specific requirements and guidance of their own organisation or commissioner. Other relevant sources such as the United Nations Evaluation Group (UNEG) guidance or the Evaluation Co-operation Group's (ECG) good practice standards and the Active Learning Network for Accountability and Performance in Humanitarian Action's (ALNAP) guidance on evaluation in humanitarian settings are also very useful where applicable.

Responding to the 2030 Agenda and the Sustainable Development Goals

Along with adapting the criteria to the institutions in which they are being used, the way the criteria are understood and applied will reflect the broader policy context, influencing how evaluation managers, evaluators and stakeholders use the criteria. For the next decade, particularly for evaluators working in international development co-operation, the Sustainable Development Goals (SDGs) and the 2030 Agenda are the single most important overarching policy framework and set of global goals.

Key elements of the 2030 Agenda, include:

- universal access to the benefits of development
- inclusiveness, particularly for those at greatest risk of being left behind
- human rights, gender equality and other equity considerations
- environmental sustainability, climate change and natural resource management
- complexity of context and of development interventions
- synergies among actors engaged in the development process.

This framework influences both interventions and their evaluation, including how the criteria are interpreted as well as the process of evaluation itself (including who is involved in applying the criteria, identifying priority questions). Box 3.3 provides guidance for using the 2030 Agenda to inform national evaluation agendas. Box 3.4 gives an example of how this was done in the German development evaluation system (BMZ, 2020^[8]). Similar efforts have been made by several national governments – including Costa Rica, Nigeria, Finland – and non-governmental organisations (NGOs) and provide useful lessons for evaluators and evaluation managers (D'Errico, Geoghe and Piergallini, 2020^[9]).

The revised wording of the criteria definitions also reflect these elements in several ways. For example, by giving particular consideration to context and beneficiary perspectives and priorities when looking at relevance, effectiveness and impact; by taking account of equity of results under effectiveness and impact; and by adopting an integrated approach when looking at coherence. The guiding principles also reflect the 2030 Agenda by encouraging an integrated way of thinking.

Box 3.3. Using the SDGs to inform national evaluation agendas

“Evaluation to connect national priorities with the SDGs. A guide for evaluation commissioners and managers”, outlines five considerations to support countries in developing evaluation agendas and enhancing the value of evaluation. Evaluators are encouraged to apply “complex systems thinking” and:

- think beyond single policies, programmes and projects
- examine macro forces influencing success or failure
- have a nuanced understanding of ‘success’
- recognise the importance of culture
- adopt evaluative thinking and adaptive management

Source: Ofir et al. (2016^[10]), *Briefing: Five considerations for national evaluation agendas informed by the SDGs*, <https://pubs.iied.org/sites/default/files/pdfs/migrate/17374IIED.pdf>

Box 3.4. Mapping the criteria to the 2030 Agenda and the SDGs

The following questions have been developed by the German Federal Ministry for Economic Cooperation and Development (BMZ) with the support of the German Institute for Development Evaluation (DEval) to help evaluators assess overall contributions to the 2030 Agenda and the SDGs. Each question below is a response to the SDG principles and also relates to the criteria.

Universality, shared responsibility and accountability

- To what extent does the intervention contribute to achieving the SDGs? (see impact criterion)
- To what extent is the intervention designed to use existing systems and structures (of partners/other donors/international organisations) for the implementation of their activities and to what extent are these used? (see coherence criterion)
- Is division of labour with other donors and development partners used when implementing the intervention? If so, to what extent? (see coherence criterion)
- To what extent are common systems used for monitoring, learning and accountability? (see coherence criterion)

Interaction of economic, environmental and social development

- To what extent does the intervention follow a holistic approach to sustainable development (social, environmental and economic)? (see relevance criterion)
- To what extent were there intended or unintended positive or negative interactions between the social, economic and environmental outcomes and what was the overall impact of the intervention? (see effectiveness and impact criterion)
- What contribution did the intervention make to promoting intended or unintended positive or negative interactions between the social, economic and environmental outcomes and what was the overall impact of the intervention? (see effectiveness and impact criterion)

Inclusiveness

- To what extent is the intervention consistent with international norms and standards on the participation and promotion of particularly disadvantaged and vulnerable groups? (see coherence criterion)
- To what extent were there intended or unintended positive or negative overarching developmental changes at the level of particularly disadvantaged and vulnerable groups (possible differentiation according to age, income, gender, ethnicity, etc.)? (see impact criterion)
- What contribution did the intervention make to the intended or unintended positive or negative overarching developmental impacts at the level of particularly disadvantaged and vulnerable groups (possible differentiation according to age, income, gender, ethnicity, etc.)? (see impact criterion)
- To what extent did the intervention contribute to strengthening the resilience of particularly disadvantaged or vulnerable groups (possible differentiation according to age, income, gender, ethnicity, etc.)? (see sustainability criterion)

Source: BMZ (2020^[8]), *Evaluation Criteria for German Bilateral Development Co-operation*

Applying a gender lens to the criteria

Evaluators should work in ways that thoughtfully consider differential experiences and impacts by gender, and the way they interact with other forms of discrimination in a specific context (e.g. age, race and ethnicity, social status). Regardless of the intervention, evaluators should consider how power dynamics based on gender intersect and interact with other forms of discrimination to affect the intervention's implementation and results. This may involve exploring how the political economy and socio-cultural context of interventions influence delivery and the achievement of objectives.

Applying a gender lens can provide evidence for learning and accountability while supporting the achievement of gender equality goals. Practical steps to apply a gender lens to the evaluation criteria include:

- evaluators, managers, and commissioners working in ways that are inclusive and lead to appropriate participation in decision making, data collection, analysis and sharing of findings
- considering the extent to which gender interacts with other social barriers to jeopardise equal opportunity in the intervention
- considering how an intervention interacts with the legislative, economic, political, religious and socio-cultural environment to better interpret different stakeholder experiences and impacts
- considering socially constructed definitions of masculinity, femininity and any changes to gender dynamics and roles
- analysis of evaluator skills in gender-sensitive evaluation approaches and experiences of working in different contexts when selecting evaluators.

The following table has been developed to help evaluators to reflect on how they can apply a gender lens to the criteria:

Table 3.1. Applying a gender lens to the criteria

Criteria	Guiding questions for applying a gender lens
Relevance	<p>Was the intervention designed in ways that respond to the needs and priorities of all genders? If so, how?</p> <p>To what extent does the intervention's design reflect the rights of persons of all genders and include feedback from a diverse range of local stakeholders including marginalised groups?</p> <p>Does the intervention meet the practical and strategic needs of all genders?</p>
Coherence	<p>To what extent are the intervention's design, delivery and results coherent with international laws and commitments to gender equality and rights, including the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW), the Beijing Declaration and Platform for Action, the Programme of Action of the International Conference on Population and Development, and the 2030 Agenda?</p> <p>To what extent does the intervention support national legislation and initiatives that aim to improve gender equality and human rights? What lessons can be learned?</p>
Effectiveness	<p>Did the intervention achieve its objectives and expected results in ways that contribute to gender equality? If so, how?</p> <p>Were there differential results for different people? If so, how and why? Were different approaches necessary to reach people of different genders? Was there sufficient monitoring and analysis of differential effects? Was the intervention adjusted to address any concerns and maximise effectiveness?</p> <p>Was the theory of change and results framework informed by analysis of gender equality, political economy analysis and human rights? If so, to what extent?</p> <p>To what extent and why is effectiveness different for people of different genders?</p>
Efficiency	<p>Were different resources allocated in ways that considered gender equality? If so, how were they allocated? Was differential resource allocation appropriate?</p> <p>Do the investment costs per person targeted meet the differentiated needs of people of different genders?</p>
Impact	<p>Were there equal impacts for different genders or were there any gender-related differences in engagement, experience and impacts? If so, why did these differential impacts occur?</p> <p>To what extent did gender-related impacts intersect with other social barriers including race/ethnicity, disability, age and sexual orientation to contribute to differential experiences and outcomes?</p> <p>How did gendered norms and barriers within the wider political, economic, religious, legislative and socio-cultural environment impact outcomes?</p> <p>To what extent have impacts contributed to equal power relations between people of different genders and to changing of social norms and systems?</p>
Sustainability	<p>Did the intervention contribute to greater gender equality within wider legal, political, economic and social systems? If so, how and to what extent? Did it result in enduring changes to social norms that are harmful to people of all or some genders? If it did not achieve this, why not?</p> <p>Will the achievements in gender equality persist after the conclusion of the intervention? Have processes contributed to sustaining these benefits? Have mechanisms been set up to support the achievement of gender equality in the longer term?</p>

Using other criteria

The six criteria are intended to be a complete set that fully reflects all important concepts to be covered in evaluations. If applied thoughtfully and in contextually relevant ways they will be adequate for evaluations across the sustainable development and humanitarian fields.

Nonetheless, in certain contexts, other criteria are used. For instance, in their evaluation policies many institutions will mandate analysis of a particular focus area. In 2020, an evaluation of Italy's health programmes in Bolivia used nine criteria: relevance, effectiveness, efficiency, impact, sustainability,

coherence, added value of Italian co-operation, visibility of Italian co-operation and ownership (Eurecna Spa, 2020^[11]). Another example involves applying the criteria to humanitarian situations, where criteria such as appropriateness, coverage and connectedness are highly relevant.³ Additionality is a criterion that is sometimes applied, often in the fields of blended finance, non-sovereign finance and climate finance. Various definitions for additionality, including different types of financial and non-financial additionality, are used.⁴ Depending on the definition used, additionality may be examined under the criterion of relevance, effectiveness or impact. Others treat it as a distinct cross-cutting criterion.

Users should be cautious when considering whether to add criteria, as this can lead to confusion and make an evaluation too broad (providing a less useful analysis). Having a limited number of criteria is useful for ensuring sufficient depth of analysis and conceptual clarity – a point that was repeatedly made during the consultation process when updating the definitions in 2017-2019.

When using other criteria, it is important to define them. An explanation of why they are being added can help ensure that other people understand how the additional criteria fit with the six described here. To support learning across evaluations, it is critical that the same concepts or elements are assessed under the same criteria.

Regardless of which criteria are used, the core principles and guidance provided here should be applied.

References

- ADE (2019), *Joint Evaluation of the Integrated Solutions Model in and around Kalobeyei, Turkana, Kenya*, UNHCR and Danida, https://um.dk/en/danida-en/results/eval/eval_reports/publicationdisplaypage/?publicationid=dd54bef1-1152-468c-974b-c171fbc2452d (accessed on 11 January 2021). [4]
- Bamberger, M., J. Vaessen and E. Raimondo (2015), *Dealing With Complexity in Development Evaluation - A Practical Approach*, https://www.betterevaluation.org/en/resources/dealing_with_complexity_in_development_evaluation (accessed on 11 January 2021). [12]
- BMZ (2020), *Evaluation Criteria for German Bilateral Development Co-operation*. [8]
- Bryld, E. (2019), *Evaluation of Sida's Support to Peacebuilding in Conflict and Post-Conflict Contexts: Somalia Country Report*, Sida, https://publikationer.sida.se/contentassets/1396a7eb4f934e6b88e491e665cf57c1/eva2019_5_62214en.pdf (accessed on 11 January 2021). [5]
- D'Errico, S., T. Geoghe and I. Piergallini (2020), *Evaluation to connect national priorities with the SDGs | Publications Library*, IIED, <https://pubs.iied.org/17739IIED> (accessed on 22 February 2021). [9]
- Danida (2019), *Evaluation of Water, Sanitation and Environment Programmes in Uganda (1990-2017)*, Evaluation Department, Ministry of Foreign Affairs of Denmark, <http://www.oecd.org/derec/denmark/denmark-1990-2017-wash-environment-uganda.pdf> (accessed on 11 January 2021). [3]
- Davis, R. (2013), "Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations", No. 40, DIFD, <https://www.gov.uk/government/publications/planning-evaluability-assessments> (accessed on 12 January 2021). [13]

- Eurecna Spa (2020), *Bolivia - Evaluation of Health Initiatives (2009-2020)*, Italian Ministry of Foreign Affairs and International Cooperation, http://www.oecd.org/derec/italy/evaluation-report-of-health-initiatives-in-Bolivia-2009_2020.pdf (accessed on 11 January 2021). [11]
- Helle, E. et al. (2018), *From Donors to Partners? Evaluation of Norwegian Support to Strengthen Civil Society in Developing Countries through Norwegian Civil Society Organisations*, Norad Norwegian Agency for Development Cooperation, https://www.norad.no/globalassets/filer-2017/evaluering/1.18-from-donor-to-partners/1.18-from-donors-to-partners_main-report.pdf (accessed on 11 January 2021). [2]
- Molund, S. and G. Schill (2004), *Looking Back, Moving Forward Sida Evaluation Manual*, Sida, <https://www.oecd.org/derec/sweden/35141712.pdf> (accessed on 11 January 2021). [6]
- OECD (2019), *Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use*, DAC Network on Development Evaluation, OECD Publishing, Paris, <https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf> (accessed on 11 January 2021). [1]
- OECD (2002), *Evaluation and Aid Effectiveness No. 6 - Glossary of Key Terms in Evaluation and Results Based Management (in English, French and Spanish)*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264034921-en-fr>. [14]
- Ofir, Z. et al. (2016), *Briefing: Five considerations for national evaluation agendas informed by the SDGs*, IIED, London, <http://dx.doi.org/10.3138/cjpe.30.3.02./11>. [10]
- WFP (2016), *Technical Note: Evaluation Methodology*, DEQAS, World Food Programme, <https://docs.wfp.org/api/documents/704ec01f137d43378a445c7e52dcf324/download/> (accessed on 11 January 2021). [7]

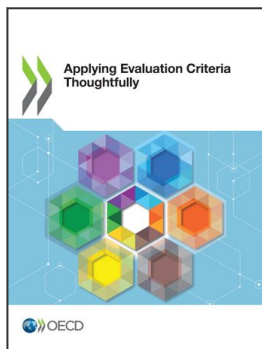
Notes

¹ Evaluability is the extent to which an activity or a programme can be evaluated in a reliable and credible fashion. Evaluability assessment calls for the early review of a proposed activity in order to ascertain whether its objectives are adequately defined and its results verifiable (OECD, 2002^[14]). See also Davis (2013^[13]).

² The process of formulating evaluation questions and involving stakeholders, while also taking into account the complexity of the intervention is discussed in more detail by, for example, Bamberger, Vaessen and Raimondo (2015^[12]).

³ The Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP) is currently updating its 2006 guidance on using the criteria in humanitarian settings, as a compliment to ALNAP's comprehensive *Evaluation of Humanitarian Action Guide*.

⁴ EvalNet's Working Group on Evaluating Blended Finance supported research on the definitions of additionality and related concepts. Findings will be published in early 2021.



From:
Applying Evaluation Criteria Thoughtfully

Access the complete publication at:
<https://doi.org/10.1787/543e84ed-en>

Please cite this chapter as:

OECD (2021), “Using the evaluation criteria in practice”, in *Applying Evaluation Criteria Thoughtfully*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/d1aca6d0-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. Extracts from publications may be subject to additional disclaimers, which are set out in the complete version of the publication, available at the link provided.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.