



Federal Ministry
for Economic Cooperation
and Development

Evaluation criteria for German bilateral development cooperation

BMZ guidelines on using the OECD-DAC criteria in the evaluation
of German bilateral development cooperation interventions

October 2024

Contents

Preliminary remarks	3
Relevance	5
Alignment with policies and priorities	5
Alignment with the needs and capacities of the beneficiaries and stakeholders	5
Appropriateness of the design	5
Adaptability – response to change	5
Coherence	6
Internal coherence	6
External coherence	6
Effectiveness	7
Achievement of the (intended) objectives	7
Contribution to achievement of objectives	7
Quality of implementation	7
Unintended results	8
Efficiency	9
Production efficiency	9
Allocation efficiency	9
Impact (higher-level development results)	10
Higher-level (intended) development changes	10
Contribution to higher-level (intended) development changes	10
Contribution to higher-level (unintended) development changes	11
Sustainability	12
Capacities of the beneficiaries and stakeholders	12
Contribution to supporting sustainable capacities	12
Durability of results over time	12
Summary of the intervention’s contributions to the 2030 Agenda for Sustainable Development	13
Universality, shared responsibility and accountability	13
Interplay of economic, environmental and social development	13
Inclusiveness/leave no one behind	14
Reference materials	15

Preliminary remarks

These guidelines were developed in a working group of the Federal Ministry for Economic Cooperation and Development (BMZ). Under the leadership of the BMZ, the Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH (GIZ), KfW Development Bank (KfW), the Physikalisch-Technische Bundesanstalt (PTB) and the Federal Institute for Geosciences and Natural Resources (BGR), supported by the German Institute for Development Evaluation (DEval), updated the BMZ guidelines on evaluation criteria for German bilateral development cooperation that were drafted in 2006. Revision of the previous document was necessitated by increasing demands on the evaluation of complex development cooperation interventions, the principles of the United Nations 2030 Agenda for Sustainable Development and the resulting adaptation of international evaluation criteria by the Development Assistance Committee of the Organisation for Economic Cooperation and Development (OECD-DAC).

The definitions of and principles for use of the revised evaluation criteria of the OECD-DAC Network on Development Evaluation (EvalNet) are presented in the report “Better Criteria for Better Evaluation” (OECD, 2019). Practical guidelines are provided in the document accompanying the report, “Thoughtful Evaluation: A Guide to Applying Criteria” (OECD DAC, 2020).

The evaluation criteria form the conceptual basis for assessing the success of German bilateral development cooperation interventions. This paper presents the key evaluation dimensions and questions¹ under the OECD-DAC’s evaluation criteria: relevance, coherence, effectiveness, efficiency, impact and sustainability (see Figure 1 for an overview of the guiding questions for the six OECD-DAC evaluation criteria). Application of the six evaluation criteria is binding for all bilateral implementing organisations and provides orientation for non-governmental development cooperation organisations as well as for the BMZ and DEval. The evaluation criteria are embedded in the framework of the principles and standards of development cooperation evaluation set out in the BMZ’s evaluation policy.

The criteria should be applied thoughtfully to support high-quality, useful evaluation. They should be contextualised – understood in the context of the individual evaluation, the intervention being evaluated, and the beneficiaries and stakeholders involved. How the criteria are applied depends on the purpose of the evaluation and the needs of the evaluation stakeholders.

These guidelines organise the evaluation questions for the six criteria (relevance, coherence, effectiveness, efficiency, impact and sustainability) into specific evaluation dimensions. In specific instances, certain priorities may be set within the individual evaluation criteria and dimensions; transparent reasons must be provided. The individual evaluation questions can be adapted to the specific situation, depending on the subject of the evaluation and interest in the findings. The substance of the evaluation criteria, evaluation dimensions and evaluation questions should not be altered.

¹ The guidelines differentiate between 1) evaluation-related questions, i.e. questions that serve to assess success, and 2) descriptive auxiliary questions that help in answering evaluation-related questions. An upright font is used for evaluation-related questions; descriptive “learning and auxiliary questions” are displayed in italics.

Figure 1. Overview of the guiding questions for the six OECD-DAC evaluation criteria



Source: OECD DAC (2019).

Using the relevant evaluation criteria as a basis, one major measure of the success of a development cooperation intervention is also its contributions to the 2030 Agenda for Sustainable Development (UN, 2015). The guiding principles of the 2030 Agenda are universality, shared responsibility and accountability, the interplay of economic, environmental and social development, and inclusiveness. In view of the Agenda's great development policy significance, the evaluation criteria listed above have been fleshed out – both by the OECD (2019; 2020) and within the German development cooperation system – by the addition of further evaluation questions based on the principles of Agenda 2030. Evaluations of German official bilateral development cooperation also include a summary of the evaluation's findings on these specific questions, taking into account the individual principles.

Relevance

IS THE INTERVENTION DOING THE RIGHT THINGS?

*The “relevance” criterion focuses on the intervention’s design. It refers to the extent to which the objectives and design of a development intervention are consistent with the (global, country and institution-specific) requirements, needs, priorities and policies of beneficiaries and stakeholders (individuals, groups, organisations and development partners). It also identifies the ability of the intervention’s design to adapt to a change in circumstances. “Relevance” is assessed in relation to 1) the **time of the intervention design**² and 2) from **today’s perspective**³.*

Alignment with policies and priorities

- To what extent are the intervention’s objectives aligned with the (global, regional and countryspecific) policies and priorities of the BMZ and of the beneficiaries and stakeholders and other (development) partners? To what extent do they take account of the relevant political and institutional environment?

Alignment with the needs and capacities of the beneficiaries and stakeholders

- To what extent are the intervention’s objectives aligned with the development needs and capacities of the beneficiaries and stakeholders involved (individuals, groups and organisations)?
- To what extent are the intervention’s objectives geared to the needs and capacities of particularly disadvantaged and vulnerable beneficiaries and stakeholders (individuals, groups and organisations)? With respect to groups, a differentiation can be made by age, income, gender, ethnicity, etc.

Appropriateness of the design⁴

- To what extent is the intervention’s design appropriate and realistic (in terms of technical, organisational and financial aspects)?
- To what extent is the intervention’s design sufficiently precise and plausible (in terms of the verifiability und traceability of the system of objectives and the underlying assumptions)?
- To what extent is the intervention’s design based on a holistic approach to sustainable development (interaction of the social, environmental and economic dimensions of sustainability)?

Adaptability – response to change

- To what extent has the intervention responded to changes in the environment over time (risks and potentials)?

2 The “time of the intervention design” is the point in time when the offer/most recent modification offer was approved or to the corresponding point in time laid down in the guidelines for financial contributions.

3 In relation to the current standards, knowledge and environment.

4 The design of an intervention is usually assessed by evaluating its intervention logic. The intervention logic depicts the system of objectives used by an intervention. It maps out the systematic relationships between the individual results levels. At the time an intervention is designed, the intervention logic, in the form of a logical model, is described in the offer for the intervention both as a narrative and generally also on the basis of a results framework. The model is reviewed at the start of an evaluation and adjusted to reflect current knowledge. Comprehensive (re)constructed intervention logics are also known as “theories of change”.

Coherence

HOW WELL DOES THE INTERVENTION FIT?

*This criterion refers to the intervention's compatibility with other interventions in a country, sector or institution as well as with international norms and standards. **Internal coherence** addresses the synergies and division of tasks between the intervention and other interventions of German development cooperation and also the intervention's consistency with the relevant international norms and standards to which German development cooperation adheres. **External coherence** considers the intervention's complementarity, harmonisation and coordination with the interventions of other partners, donors and international organisations. The "coherence" criterion relates both to the intervention's design as well as to the results it achieves.*

Internal coherence

- Within German development cooperation, to what extent is the intervention designed and implemented (in a sector, country, region or globally) in a complementary manner, based on the division of tasks?
- To what extent are the instruments of German development cooperation (Technical and Financial Cooperation) meaningfully interlinked within the intervention (in terms of both design and implementation)? Are synergies leveraged?
- To what extent is the intervention consistent with international and national norms and standards to which German development cooperation is committed (e.g. human rights)?

External coherence

- To what extent does the intervention complement and support the partner's own efforts (principle of subsidiarity)?
- To what extent has the intervention's design and implementation been coordinated with other donors' activities?
- To what extent has the intervention's design been designed to use existing systems and structures (of partners/other donors/international organisations) for implementing its activities? To what extent are these systems and structures used?
- To what extent are common systems (together with partners/other donors/international organisations) used for M&E, learning and accountability?

Effectiveness

IS THE INTERVENTION ACHIEVING ITS OBJECTIVES?

“Effectiveness” refers to the extent to which the intervention has achieved, or is expected to achieve, its objectives (at outcome level), including any differential results across beneficiary and stakeholder groups. It examines the achievement of objectives in terms of the direct, short-term and medium-term results.

Achievement of the (intended) objectives

- To what extent has the intervention achieved, or is the intervention expected to achieve, the (intended) objectives as originally planned (or as modified to cater for changes in the environment)?

Contribution to achievement of objectives

- *To what extent have the intervention’s outputs been delivered as originally planned (or as modified to cater for changes in the environment)?*
- To what extent have the delivered outputs and increased capacities been used and equal access (e.g. in terms of physical, non-discriminatory and affordable access) guaranteed?
- To what extent has the intervention contributed to the achievement of objectives?
- To what extent has the intervention contributed to the achievement of objectives at the level of the intended beneficiaries?
- To what extent has the intervention contributed to the achievement of objectives at the level of particularly disadvantaged or vulnerable groups of beneficiaries and stakeholders? (These may be broken down by age, income, gender, ethnicity, etc.)
- *Which internal factors (technical, organisational or financial) were decisive for achievement/non-achievement of the intervention’s intended objectives?*
- *Which external factors were decisive for achievement/non-achievement of the intervention’s intended objectives (taking into account the anticipated risks)?*

Quality of implementation

- What assessment can be made of the quality of steering and implementation of the intervention in terms of the achievement of objectives?
- What assessment can be made of the quality of steering and implementation of, and participation in, the intervention by the partner/executing agency?

Unintended results

- To what extent can unintended positive/negative direct results (social, economic, environmental and among vulnerable beneficiary groups) be observed/anticipated?
- What potential benefits/risks arise from the positive/negative unintended results? What assessment can be made of them?
- How has the intervention responded to the potential benefits/risks of the positive/negative unintended results?

Efficiency

HOW WELL ARE RESOURCES BEING USED?

*This criterion describes the extent to which the intervention delivers results in an economic and timely way (relationship between input and output, outcome and impact level). The evaluation dimension “**production efficiency**” refers to the appropriateness of the relationship between inputs and outputs. The evaluation dimension “**allocation efficiency**” refers to the appropriateness of the relationship between the inputs and the results achieved (project/development objective; outcome/impact level) by the intervention. The “efficiency” criterion relates both to the intervention’s design and implementation and to the results it achieves.*

Production efficiency

- *How are the intervention’s inputs (financial, human and material resources) distributed (e.g. by instruments, sectors, sub-interventions, taking into account the cost contributions of partners/executing agencies/other beneficiaries and stakeholders etc.)?*
- *To what extent have the intervention’s inputs (financial, human and material resources) been used economically in relation to the outputs delivered (products, investment goods and services)? If possible, refer to data from other evaluations in a region or sector, for instance.*
- *Possible comparison with alternatives: To what extent could the intervention’s outputs (products, investment goods and services) have been increased through the alternative use of inputs (financial, human and material resources)? If possible, refer to data from other evaluations of a region or sector, for instance.*
- *Were the outputs (products, investment goods and services) produced on time and within the planned time frame?*

Allocation efficiency

- *By what other means and at what cost could the results achieved (higher-level project objective) have been attained?*
- *To what extent – compared with alternative designs for the intervention – could the results have been attained more cost-effectively?*
- *Possible comparison with alternatives: To what extent – compared with alternative designs for the intervention – could the positive results have been increased using the existing resources?*

Impact (higher-level development results)

WHAT DIFFERENCE DOES THE INTERVENTION MAKE?

Based on recognisable higher-level development changes (at impact level), the criterion of “higher-level development results (at impact level)” relates to the extent to which the intervention has already produced significant positive or negative, intended or unintended results at the overarching level (contributions to the observed changes), or is expected to do so in the future. This includes any differential results across different stakeholders and beneficiaries. This criterion refers to the results of the development intervention.

Higher-level (intended) development changes

- To what extent can the higher-level development changes (social, economic and environmental dimensions and the interactions between them) to which the intervention will/is designed to contribute be identified/foreseen? (Specify time frame where possible.)
- To what extent can the higher-level development changes (social, economic, environmental dimensions and the interactions between them) be identified/foreseen at the level of the intended beneficiaries? (Specify time frame where possible.)
- To what extent can higher-level development changes to which the intervention will/is designed to contribute be identified/foreseen at the level of particularly disadvantaged/vulnerable groups of beneficiaries and stakeholders? (These may be broken down by age, income, gender, ethnicity, etc.) (Specify time frame where possible.)

Contribution to higher-level (intended) development changes

- To what extent has the intervention actually contributed to the identified and/or foreseeable higherlevel development changes (social, economic, environmental dimensions and their interactions, taking into account political stability) that it was designed to bring about?
- To what extent has the intervention achieved its intended (original and, where applicable, revised) development objectives?
- To what extent has the intervention achieved its (original and, where applicable, revised) development objectives at the level of the intended beneficiaries?
- To what extent has the intervention contributed to higher-level development changes/changes in the lives of particularly disadvantaged or vulnerable groups of beneficiaries and stakeholders that it was designed to bring about? (These may be broken down by age, income, gender, ethnicity, etc.).
- *Which internal factors (technical, organisational or financial) were decisive for achievement/ non-achievement of the intervention’s intended development objectives?*

- *Which external factors were decisive for the achievement/non-achievement of the intervention's intended development objectives?*
- To what extent has the intervention achieved structural or institutional changes (e.g. for organisations, systems and regulations)?
- To what extent did the intervention serve as a model and/or achieve broad-based impact?
- *How would the situation have developed without the intervention?*

Contribution to higher-level (unintended) development changes

- To what extent can higher-level, unintended development changes (social, economic and environmental dimensions and their interactions, taking into account political stability) be identified/foreseen? (Specify time frame where possible.)
- To what extent has the intervention brought about foreseeable/identifiable unintended (positive and/or negative) higher-level development results?
- To what extent has the intervention contributed to foreseeable/identifiable unintended (positive and/or negative) higher-level development results at the level of particularly disadvantaged or vulnerable groups of beneficiaries and stakeholders? (These may be broken down by age, income, gender, ethnicity, etc.)

Sustainability

WILL THE BENEFITS LAST?

The “sustainability” criterion relates to continued long-term benefits (at the outcome and impact level) or the probability of continued long-term benefits – taking into account observed or foreseeable risks – over time, particularly after assistance has ended.

Capacities of the beneficiaries and stakeholders

- To what extent do the beneficiaries and stakeholders (individuals, groups and organisations, partners and executing agencies) have the institutional, human and financial resources as well as the willingness (ownership) required to sustain the positive results of the intervention over time (once assistance has drawn to a close)?
- To what extent do the beneficiaries and stakeholders (individuals, groups and organisations, partners and executing agencies) have the resilience to overcome future risks that could jeopardise the intervention’s results?

Contribution to supporting sustainable capacities

- To what extent has the intervention contributed to the beneficiaries and stakeholders (individuals, groups and organisations, partners and executing agencies) having the institutional, human and financial resources as well as the willingness (ownership) required to sustain the intervention’s positive results over time and to limit the impact of any negative results?
- To what extent has the intervention contributed to strengthening the resilience of the beneficiaries and stakeholders (individuals, groups and organisations, partners and executing agencies)?
- To what extent has the intervention contributed to strengthening the resilience of particularly disadvantaged groups? (These may be broken down by age, income, gender, ethnicity, etc.)

Durability of results over time

- *How stable is the context in which the intervention operates?*
- *To what extent is the durability of the intervention’s positive results influenced by the context?*
- To what extent can the positive (and any negative) results of the intervention be deemed durable?

Summary of the intervention's contributions to the 2030 Agenda for Sustainable Development

WHAT CONTRIBUTION DOES THE INTERVENTION MAKE TO IMPLEMENTING THE PRINCIPLES OF THE 2030 AGENDA FOR SUSTAINABLE DEVELOPMENT?

The summary of contributions to the 2030 Agenda focuses on the principles of universality, shared responsibility and accountability, the interplay of economic, environmental and social development, and inclusiveness. Drawing on the evaluation questions for each of the evaluation criteria, the summary makes the connection between the various findings; it provides a synopsis of the intervention's overarching contributions to the 2030 Agenda.

Universality, shared responsibility and accountability

- To what extent does the intervention contribute to achieving the Sustainable Development Goals (SDGs) [see “impact” criterion]?
- To what extent is the intervention's design geared to the use of existing systems and structures (of partners/other donors/international organisations) for implementation of its activities? To what extent are these systems and structures used [see “coherence” criterion]?
- When implementing the intervention, to what extent are tasks divided up with other donors and development partners [see “coherence” criterion]?
- To what extent are shared systems used for monitoring, learning and accountability [see “coherence” criterion]?

Interplay of economic, environmental and social development

- To what extent does the intervention pursue a holistic approach to sustainable development (social, environmental and economic dimensions) [see “relevance” criterion]?
- To what extent have there been intended or unintended positive or negative interactions between the social, economic and environmental results achieved by the intervention [see “impact” criterion]?
- What contribution does the intervention make to fostering intended or unintended positive or negative interplay between the social, economic and environmental results it achieves [see “impact” criterion]?

Inclusiveness/leave no one behind

- To what extent does the intervention conform with international norms and standards for participation and the support of particularly vulnerable groups [see “coherence” criterion]?
- To what extent have there been intended or unintended positive or negative changes at the level of particularly disadvantaged or vulnerable groups of beneficiaries and stakeholders? (These may be broken down by age, income, gender, ethnicity, etc.) [See the criteria “effectiveness” and “impact”.]
- What contribution does the intervention make to intended or unintended, positive or negative results at the level of particularly disadvantaged or vulnerable groups of beneficiaries and stakeholders? (These may be broken down by age, income, gender, ethnicity, etc.) [See the criteria “effectiveness” and “impact”.]
- To what extent has the intervention contributed to building resilience of particularly disadvantaged or vulnerable groups of beneficiaries and stakeholders? (These may be broken down by age, income, gender, ethnicity, etc.) [See “sustainability” criterion.]

Reference materials

BMZ (2023), *Evaluating German Development Cooperation. BMZ Evaluation Policy*, Federal Ministry for Economic Cooperation and Development (BMZ), Bonn.

OECD (2019), *“Better Criteria for Better Evaluation. Revised Evaluation Criteria Definitions and Principles for Use”*, OECD/DAC Network on Development Evaluation, Paris.

OECD DAC (2020), *“Thoughtful Evaluation: A Guide to Applying Criteria”*, OECD/DAC Network on Development Evaluation, Paris.

UN (2015), *“Transforming our World. The 2030 Agenda for Sustainable Development”*, New York.