MINISTRY OF FOREIGN AFFAIRS OF DENMARK

Danida Evaluation Guidelines

April 2024

Department for Evaluation, Learning and Quality (LEARNING)

PREFACE

These Evaluation Guidelines provide operational guidance for evaluations of Danish development cooperation and complements the Evaluation Policy for Danish Development Cooperation (2020), which constitutes the overall framework for evaluations of Danish development cooperation. This version of the guidelines replaces the version published in 2018. It is aligned to the new policy and organisation of the Ministry of Foreign Affairs (MFA).

The *Quality Standards for Development Evaluation*, which have been agreed between members of the OECD/DAC Evaluation Network[1], and which constitute a basis for development evaluation of all OECD/DAC members, have influenced the contents of these guidelines.

The audience for the guidelines are those who have a professional engagement in evaluation of development cooperation, as well as others interested in evaluation. These include those who are parties to an evaluation process and the users of evaluations. Moreover, the guidelines may be of interest to a broader audience, such as students, researchers, and policy makers, and the interested public.

Chapter 1 explains the role of evaluation in Danish development cooperation and provides definitions and key principles of evaluation of development cooperation in the Ministry of Foreign Affairs. Chapter 2 outlines the processes and requirements related to the formulation of the evaluation programme. Chapter 3 explains the scoping of an evaluation. The concrete design steps of an evaluation are elaborated in Chapter 4. Chapter 5 presents the MFA's approach to evaluation of multilateral cooperation. Chapter 6 lays out the organisational set-up of an evaluation and the contractual procedures around assigning an evaluation team. Chapter 7 describes the different implementation phases of an evaluation from the inception phase to the final reporting. The final chapter (Chapter 8) explains the mechanisms in place to ensure learning from and dissemination of evaluations. The guidelines include five annexes: codes of conduct (Annex 1); quality control and quality assurance (Annex 2); key issues to be covered by an inception report (Annex 3); key issues to be covered by an evaluation report (Annex 4); and an overview of analytical quality issues, pointing to validity and reliability requirements (Annex 5). The guidelines do not constitute a manual in evaluation methods and techniques, but reference is made to useful links and resources throughout the text.

The guidelines will be updated as need arises, and comments and suggestions for improvements or clarifications are welcome and may be forwarded to laering@um.dk

---

[1] The OECD/DAC network on evaluation brings together 36 bilateral donors and multilateral development agencies with the purpose of supporting robust, informed and independent evaluations of development cooperation. It is a subsidiary body to the Development Assistance Committee (DAC). See http://www.oecd.org

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS

3iE	International Initiative on Impact Evaluation
ALNAP	Active Learning Network for Accountability and Performance in Humanitarian Action
AMG	Aid Management Guidelines
CMO	Context-Mechanism-Outcome
DAC	Development Assistance Committee
Danida	Danida is the term used for Denmark's development cooperation (but not an organisation). Development cooperation falls under the mandate of the Ministry of Foreign Affairs of Denmark
DEReC	The DAC Evaluation Resource Centre – repository for evaluation reports
DIME	Development Impact Evaluation
EM	Evaluation Management
ERG	Evaluation Reference Group
EvalNet	OECD-DAC Network for Development Evaluation
GEI	Global Evaluation Initiative
LEARNING	The Department for Evaluation, Learning and Quality
MFA	Ministry of Foreign Affairs of Denmark
NGO	Non-Governmental Organisation
NONIE	Network of Networks on Impact Evaluation
OECD	Organisation for Economic Cooperation and Development
QA	Quality Assurance
RBM	Results Based Management
RTE	Real-Time Evaluation
SWE	Sector-Wide Evaluation
ToR	Terms of Reference

CHAPTER 1. DEVELOPMENT EVALUATION: DEFINITION AND OVERARCHING ISSUES

This chapter defines and explains the role of development evaluation[2] as an instrument in Danish development cooperation.

1.1 What is development evaluation?

All major development agencies involved in international development cooperation undertake formal evaluations of their activities. Denmark has integrated evaluation processes into development cooperation since the early 1980s. Both the evaluation function and international cooperation on evaluations have since been strengthened significantly.

The joint efforts of OECD/DAC donors to develop the *"Quality Standards for Development Evaluation"* (OECD/DAC 2010)[3] underscore the importance paid to evaluation by development cooperation actors. Denmark adheres to the OECD/DAC definition of evaluations:

> *Development evaluation is the systematic and objective assessment of an on-going or completed development intervention, its design, implementation and results. In the development context, evaluation refers to the process of determining the worth or significance of a development intervention.* (OECD/DAC Quality Standards section 1.1)

1.2 The role of evaluation in Danish development cooperation

Danish development cooperation includes bilateral country programmes, regional programmes, programmes and funds managed centrally (e.g. in relation to climate and to the private sector as well as strategic sector cooperation), humanitarian support, and support to civil society organisations as well as an extensive engagement through multilateral organisations and the EU.

As stated in the Evaluation Policy for Danish Development Cooperation (2020), evaluation plays a critical role in assessing and understanding the contribution of Danish development cooperation by generating learning and evidence for what worked and what did not work and why. Evaluation of Danish development cooperation serves two interrelated purposes:

- Learning with a view to improving the quality and results of development cooperation through generation of knowledge about what works, and why.

---

[2] The terms "development evaluation" and "evaluation" are used interchangeably in these guidelines.
[3] Hereafter referred to as the OECD/DAC Quality Standards.

- Accountability through reporting and communicating results from the development cooperation to stakeholders and the wider public in Denmark and abroad, including beneficiaries.

The evaluation function supplements and works in conjunction with other programme cycle processes and aid management tools. The Ministry of Foreign Affairs applies a results-based management (RBM) approach for development cooperation and makes use of a range of different aid management tools in order to plan, monitor and assess progress. These tools, which include appraisals, reviews, performance monitoring and assessments, feed into the various phases of the programme cycle and provide valuable information for evaluations. Explanations of the RBM approach and various tools used by the MFA to monitor development cooperation can be found at www.um.dk and www.amg.um.dk

The evaluation function in Danish development cooperation resides in the Ministry of Foreign Affairs. The Office of Evaluation, Learning and Quality (LEARNING) is responsible for the planning, management and quality control of evaluations of Danish development cooperation. The evaluation function has a specific budget for its work and matters concerning evaluation refers directly to the Minister for Development Cooperation and Global Climate Policy through the State Secretary for Development Policy and is independent of the operational functions of the Ministry of Foreign Affairs (HQ departments and representations abroad involved in development cooperation). In addition, the evaluation function is also responsible for management and collaboration in relation to a range of studies, including pre-studies as part of the preparation of evaluations, meta-studies and other types of studies contributing to learning of relevance for Danish development cooperation.

1.3 Ensuring quality

The Ministry of Foreign Affairs operates with three types of evaluation products:

1. **Evaluations of policies, strategies, modalities, themes, and programmes.** These are typically ex-post evaluations, and are often strategic in nature as they seek to address issues of broader interest to policy makers and the public as to the results of development cooperation as well as provide learning to inform new policies, strategies, and programmes. Only LEARNING can initiate such evaluations, with the various requirements for follow-up etc., that they entail.

2. **Evaluation studie**s are studies of a particular issue where evidence is sought for. Evaluation studies can be used for documenting results or lessons learned. They may also take the form of meta evaluations/synthesis evaluations based on evaluation results produced by the Ministry of Foreign Affairs as well as other development partners and/or research. Evaluation studies may be studies in their own right or form part of the preparation for a larger evaluation.

3. **Follow-up evaluations and studies**: These evaluations and studies are designed as investigation related to follow-up to finalised evaluations, often with the purpose of supporting efforts to enhance evaluability of the next phase of the strategy or programme.

Quality *assurance* is the responsibility of the team selected to conduct a particular evaluation or study (see Section 6.4); the standards to be followed are laid down in the technical bid and subsequently in the contract with the selected evaluation team.

LEARNING conducts quality *control* of all evaluations. The OECD/DAC Quality Standards on quality control are adhered to. This includes ensuring that the evaluation responds to the Terms of Reference (ToR), that analytical clarity and consistency is ensured, and that the evaluation is conducted in accordance with ethical standards and codes of conduct (see annexes 1 and 2). However, the exact nature of the quality control arrangements for evaluations depends on the type of evaluation and the scope and complexity (see Chapter 5). Quality control may be expanded during the evaluation process, if need be. Quality control is also carried out for studies, but without specific reference to the OECD/DAC standards.

Both quality assurance and quality control are exercised throughout the evaluation process (see also Annex 2).

CHAPTER 2. THE EVALUATION PROGRAMME

This chapter explains the process related to the preparation of the MFA's evaluation programme for development cooperation.

2.1 Preparation of the evaluation programme

Preparation of the evaluation programme falls under the mandate of LEARNING. The portfolio of evaluations is planned on a two-year rolling basis. The programme is developed through consultations and discussions with stakeholders and is updated annually. The selection of evaluations for the programme is partly based on suggestions from representations abroad, departments in the Ministry of Foreign Affairs, stakeholders and the public; and partly decided upon by LEARNING in order to meet the obligation to cover the entire development cooperation portfolio.

The consultation process includes the Council for Development Policy, internal ministerial hearings and dialogue at both headquarters and embassy level. Representations are encouraged to consult national stakeholders about possible topics for evaluation.

The dialogue with stakeholders concerning topics for the evaluation programme is important in order to secure utilisation of results and promote learning from evaluations. Key stakeholders include senior management in the Ministry of Foreign Affairs, managements, staff in departments and representations, embassies, relevant partners in development cooperation, and may further include consultants involved in programme design, researchers, and others with an interest in development cooperation. Following the hearing process, LEARNING presents the proposed evaluation programme including a section on evaluation activities conducted in the previous year to the Council for Development Policy (See Box 1.) The final evaluation programme is submitted to the Minister for Development Cooperation and Global Climate Policy, before being posted on the website Evaluation of development assistance (um.dk).

Box 1: The role of the Council for Development Policy

Programme preparation: The Council for Development Policy provides comments to the suggested evaluation programme and may suggest ideas for new evaluations.

Information and presentation: All evaluation summaries for completed evaluations are sent to the Council for Development Policy for information. The Council for Development Policy can request in-depth presentations and discussions of specific evaluations as well as the follow-up plans as expressed in the Management Response.

2.2 Programme coverage and strategic considerations

The evaluation programme must, over 5-7 years, demonstrate a proper coverage of the total development cooperation portfolio. The programme coverage includes different modalities and instruments, a balance of geographical areas, large and small partner countries, and

thematic areas. Evaluations may also be undertaken as joint evaluation with other donors and/or partner country/countries, when the subject matter makes this feasible and appropriate. The following are examples of evaluation topics:

- General cooperation with a partner country or cooperation within specific sectors or themes in a partner country.

- Specific aspects of the Strategy for Denmark's Development Cooperation and Humanitarian Assistance or specific policies.

- The various instruments in development cooperation such as sector/thematic programmes and project assistance to partner countries, private sector instruments, support to civil society, mixed credits, humanitarian assistance, climate change funds, strategic sector cooperation, globalisation and stabilisation support, and support to research.

- Thematic evaluations, covering themes such as anti-corruption, decentralisation, gender equality, exit strategies, etc.

- On-going or completed projects and programmes, which may have a broader significance for development cooperation.

- Funding channelled through multilateral organisations, core as well as earmarked funding, including so-called multi-bi support (see Chapter 5).

Box 2: Main parties to an evaluation

Collaboration between key parties, with due adherence to the principles of independence and impartiality of the different parties, is key to successful evaluations. Parties include:

The commissioner. This is either LEARNING alone or LEARNING jointly with other development agencies, and/or LEARNING with partner country institutions.

The evaluators. These are the independent and impartial consultant companies or individuals that carry out the evaluation, i.e. collect and analyse data, judge the value of the cooperation intervention, and produce the evaluation report. Evaluators are often organised as an *evaluation team*.

The stakeholders. These are the agencies, organisations, groups or individuals that have an interest in the development intervention or its evaluation, but not necessarily a formal role in the subject of the evaluation.

The Users. The users of evaluations are stakeholders with a specific relationship to the intervention and its evaluation. They include policy makers, MFA management, staff, advisers, partner country decision makers and implementers, and other parties with a formal or direct role in relation to the development activities under evaluation.

Stakeholder and user participation in the evaluation process is encouraged to promote learning, ownership, and application of results and "lessons learned" to future activities. In practice, the groups may overlap.

CHAPTER 3. SCOPING AN EVALUATION

This chapter describes the process of scoping or conceptualising an evaluation. Each evaluation preparation process is unique. Conceptualising an evaluation is often exploratory and therefore not a linear process. In most cases, the scoping is followed by a design phase (see Chapter 4).

3.1 Preliminary considerations

The preliminary scoping starts during the consultations that feed into the evaluation programme (see Section 2.1). This scoping includes assessments of various topical angles and entry points, preliminary assessment of existing studies of the subject area, timing and logistics, and availability of funding and other resources. Stakeholder views and possible cooperation modalities are also solicited. Studies may be carried out as part of evaluability assessments and in order to provide background for focusing the evaluation (see 4.2).

> *The feasibility of an evaluation is assessed. Specifically, it should be determined whether or not the development intervention is adequately defined and its results verifiable, and if evaluation is the best way to answer questions posed by policy makers or stakeholders.* (OECD/DAC Quality Standards section 2.4, qualitystandards.pdf (oecd.org))

The availability of data and the evaluability of the topic are then gauged. Gauging evaluability involves recognising the barriers and complications that may challenge the evaluation process and assessing whether a topic can be evaluated fulfilling the requirements of the OECD/DAC Quality Standards. At this stage, evaluability considerations may lead to the conclusion that a topic is not feasible for evaluation, but more suited for a review, audit or research project.

The options for organisation, management and resourcing of the evaluation are also considered. This includes consideration of the potential for conducting the evaluation as a joint undertaking, for instance in collaboration with other agencies (see Section 3.2) or the possible involvement of the partner country or countries.

3.2 Considering a joint evaluation

The Danish MFA and other development agencies aim to promote mutual accountability in the management and administration of development cooperation, and various valuable networks help foster collaboration and cooperation around development evaluations. This entails being attentive to the possibility of joint evaluations, in particular when addressing coordinated and aligned programmes or other areas where the evaluation subject reflects a joint effort.

> *In order to increase ownership of development and build mutual accountability for results, a partnership approach to development evaluation is systematically considered early in the process. The concept of partnership connotes*

> *an inclusive process, involving different stakeholders such as government, parliament, civil society, intended beneficiaries and international partners. (OECD/DAC Quality Standards section 1.4)*

Joint evaluations have been conducted by various groups of donors including the Danish MFA since the late 1980s, including joint evaluations with partner countries and with bilateral donors or multilateral organisations. In later years the focus on joint evaluations in the international community has decreased somewhat, but recently an increased interest is detected, and LEARNING will continue to take initiatives for joint evaluations and assess the interest, whenever relevant.

> *A joint evaluation is an evaluation in which different donor agencies and/or partners participate. There are various degrees of "jointness" depending on the extent to which individual partners cooperate in the evaluation process, merge their evaluation resources and combine their evaluation reporting. The decision to conduct an evaluation singly or jointly should be taken on a case-by-case basis and with careful consideration of the value added and benefits and costs involved. Joint evaluations are particularly appropriate when evaluating co-financed programme support at the budget or sector level, national aid effectiveness goals, the effectiveness of a multilateral or regional development agency or issues that are too sensitive or controversial for one agency alone to tackle.* (OECD/DAC: Managing Joint Evaluations, 2010; [46868375.pdf (oecd.org)](46868375.pdf))

With regard to conducting joint evaluations, OECD/DAC has summarised the experience and found that if an evaluation has to be conducted in haste, is narrowly focused, or mainly concerned with domestic accountability needs, a joint evaluation approach may not be appropriate. It is also noted that joint evaluations are often more costly and take longer than a single donor evaluation[4]. Possible strengths of a joint approach include enhanced credibility because of broader ownership, cross-learning among institutions, and decreased transaction costs for partner countries. A possible disadvantage is that, in joint evaluations, the evaluation questions can become overly general, largely because of the incorporation of all partners' requirements.

In the case of joint programmes or common subject matters, Denmark will consider the possibility of joint evaluations. Before engaging in a joint evaluation, LEARNING considers the opportunities and challenges. It also assesses the potential benefits such as the expected learning outcome and value in relation to transaction costs. LEARNING also engages actively in joint evaluations, where other donors are contract holders. Experience has shown that a sleeping partner role is not satisfactory.

With regard to joint evaluations conducted with partner countries, the *Evaluation study on experience with conducting evaluations jointly with partner countries[5]* lists good practices for such evaluation processes with respect to both partner country institutions and international evaluation departments. The joint evaluations in which Denmark has participated are published at LEARNING's website [Evaluation reports and other publications (um.dk)](um.dk). The OECD/DAC database of evaluations (DEReC) includes additional examples of joint evaluations, oecd.org/derec/home/?hf=5&b=0&s=score.

---

[4] OECD/DAC: *Managing Joint Evaluations*, 2010.
[5] Danida: *Evaluation study on experiences with conducting evaluations jointly with partner countries*, Serial number 2009/3.

The design phase develops the considerations made in the scoping phase. The rationale and purpose of the evaluation are sharpened, and the objectives of the evaluation are formulated. A preparatory study may be undertaken in order to establish an overview of the portfolio and/or the approach and methodology of the evaluation. The study may also outline prioritisation of the evaluation criteria, the preliminary list of overall evaluation questions, and the methodology to be applied to answer the evaluation questions. The design of an evaluation is an evolving process and continues into the tendering phase (see Section 6.5) and subsequently the inception phase of the evaluation (see Section 7.1).

4.1 Rationale, purpose and objectives

When the broader topic and focus of an evaluation have been established, the rationale, purpose and intended use of the evaluation should be clarified. This requires a deeper assessment of the topic, focusing on why the evaluation is being undertaken, for whom, and if the evaluation meets accountability or learning purposes – or both.

Rationale and Purpose of the Evaluation
The rationale, purpose and intended use of the evaluation are stated clearly, addressing: why the evaluation is being undertaken at this particular point in time, why and for whom it is undertaken, and how the evaluation is to be used for learning and/or accountability functions.
For example, the evaluation's overall purpose may be to:
- *contribute to improving a development policy, procedure or technique,*
- *consider the continuation or discontinuation of a programme,*
- *account for public expenditures and development results to stakeholders and tax-payers.* (OECD/DAC Quality Standards, section 2.1)

The formulation of the objectives is a crucial step in the design process, because the objectives set the level of ambition and clarify what the evaluation tries to find out. In addition, the objectives spell out whether the evaluation has a process or a results focus. Evaluation should be objective driven rather than driven by methods or methodological considerations. Only when the objectives are clearly formulated does it become possible to determine the most suitable approach and methodology (see Section 4.4).

4.2 Preparatory studies

In the process of determining the rationale, purpose and objectives, LEARNING may use various types of preparatory studies: a pre-study, an approach paper, a status or scoping study or a combination (see Box 3). The Terms of Reference includes relevant aspects of a -study or an approach paper and constitute the main or only preparatory document (see Section 4.5).

Box 3: Preparatory studies

Preparatory studies assist the parties to specify why an evaluation is undertaken (accountability and/or learning purposes) as well as the anticipated focus, objectives and use of the evaluation.

Preparatory studies (pre-studies or approach papers) assemble the first broad overview documentation of a particular evaluation topic. They are typically used to summarise existing knowledge and factual information. They also suggest the focus and purpose of the up-coming evaluation and assess evaluability and discuss potential approaches, methods and sampling strategies.

Technically, the main purpose is to help clarify, simplify and focus an often complex and broad topic. Strategically, the preparatory studies also constitute a way of engaging stakeholders to build commitment to the evaluation, including the use of the evaluation. Therefore, preparatory studies are shared with relevant parties, both in and outside the MFA, to solicit their engagement and comments in time before an evaluation is carried out. This process helps to strengthen the foundation for the evaluation and may also be used to explore possibilities for a broader cooperation with other donors or providers of development cooperation.[6]

A status or scoping study will typically be broader in nature, for instance providing background information of relevance to deciding on whether a full-scale evaluation is the most appropriate way forward, and on helping delineating a fruitful scope, in order to ensure that an evaluation has sufficient added value.

4.3 Evaluation criteria and evaluation questions

Part of the basis for any evaluation, irrespective of type and focus, is the six OECD/DAC[7] criteria.

Box 4: Evaluation criteria

- **Relevance** is the extent to which the intervention objectives and design respond to beneficiaries, global country and partner/institution needs, policies, and priorities, and continue to do so if circumstances change.
- **Coherence** is the compatibility of the intervention with other interventions in a country, sector or institution.
- **Effectiveness** is the extent to which the intervention achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups
- **Efficiency** is the extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way.
- **Sustainability** is the extent to which the net benefits of the intervention continue, or are likely to continue.
- **Impact** is the extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.

---

[6] The Danish MFA and some other organisations consider that preparatory analyses can be similar or equivalent to an evaluability assessment. For more information regarding evaluability assessments, see for instance Evaluability assessment | Better Evaluation

[7] , See the full overview of the OECD/DAC, Criteria, 2019 at: Evaluation Criteria - OECD

It is important to note, that not all criteria may be equally relevant in all evaluations. The OECD/DAC evaluation criteria should not become a straitjacket within which evaluations are designed, implemented and communicated. Rather, they should be understood as a tool in the evaluation process to ensure that the evaluation cover issues of importance to answer the essential questions of any evaluation: What worked, what did not work, and why, and then what now? The evaluation questions and what you intend to do with the answers, should inform how the criteria are specifically interpreted and analysed. Further, 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.[8]

Use of the standard OECD/DAC criteria does not exclude that other evaluation criteria be applied. The reason for supplementing the six standard evaluation criteria is to enhance and/or expand the focus of an evaluation.

The criteria for the evaluation of *humanitarian assistance* are a case in point. Because of the unique features of humanitarian action, the Active Learning Network for Accountability and Performance in Humanitarian Action, ALNAP, has introduced an interpretation of the OECD/DAC criteria as well as additional evaluation criteria: appropriateness, connectedness, coordination and coverage. (See Box 5.) At times, the evaluation of specific topics, such as climate change, or evaluations in specific contexts, such as conflict and high-risk, post-conflict or multiple natural disaster situations, call for the use of specific methodologies, such as remote evaluation techniques.

Considerable effort is expended in various fora to discuss how to adapt the OECD/DAC criteria to specific settings as well as how to develop context-specific evaluation guidelines. For example, in relation to the evaluation of conflict prevention and peace building interventions, there are continuous efforts to refine and field-test guidelines for the application of both the OECD/DAC and the additional ALNAP criteria. Similarly, guidelines have been elaborated for applying a human rights and gender equality lense to the OECD Evaluation Criteria.[9]

---

[8] See Applying Evaluation Criteria Thoughtfully, OECD/DAC 2021, for more. Applying Evaluation Criteria Thoughtfully | en | OECD

[9] See *Applying a Human Rights and Gender Equality Lens to the OECD Evaluation Criteria;* OECD-DAC 2023, Applying a Human Rights and Gender Equality Lens to the OECD Evaluation Criteria | Best Practices in Development Co-operation | OECD iLibrary (oecd-ilibrary.org)

Box 5: Criteria for evaluation of humanitarian action

For evaluation of humanitarian response, a set of criteria is used, that represents both a specific interpretation as well as supplementation of the OECD/DAC criteria. In addition to effectiveness and impact, the following criteria are considered:

- **Appropriateness:** The need to tailor humanitarian activities to local needs, increasing ownership, accountability, and cost-effectiveness accordingly. Replaces the relevance criterion used in development evaluations.
- **Coverage:** The extent to which major population groups facing life-threatening suffering were reached by humanitarian action.
- **Connectedness:** The need to assure that activities of a short-term emergency nature are carried out in a context which takes longer-term and inter-connected problems into account. Replaces the sustainability criterion used in development evaluations.
- **Coordination:** The extent to which the interventions of different actors are harmonised with each other, promote synergy, avoid gaps, duplication, and resource conflicts.[10]

The evaluation objectives are translated into relevant and specific evaluation questions. Evaluation questions are decided early on in the process and inform the development of the methodology. The evaluation questions also address cross-cutting issues, such as gender equality, environment, climate, and human rights. (OECD/DAC Quality Standards section 2.7)

Evaluation questions are formulated on the basis of the purpose and specific criteria chosen for a particular evaluation. These questions are intended to operationalise the criteria and guide the focus of the evaluation. Evaluation questions can be normative (did the intervention meet its objectives to a satisfactory degree?), descriptive (what happened during the development intervention?), or focus on the cause-and-effect of change that has been observed.

4.4 Approach and methodology

Over the years, evaluation approaches have evolved away from classical categorisations, such as summative and formative approaches[11]. Today donors often use the term "approach" to cover both the evaluation type and organisational aspects of an evaluation. In each evaluation, the approach is developed on the basis of the purpose and objectives and an agreement between partners regarding the overall organisation. The approach then

---

[10] OECD/DAC (1999) *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies*. Paris, pp 30-32, reviewed and updated in: ALNAP, Evaluating Humanitarian Action (EHA), 2016. ALNAP has reviewed the use of criteria in humanitarian evaluation, see ALNAP 2023, alnap-oecd-guidance-literature-review-report.pdf (cdn.ngo)

[11] *Formative evaluations* (often called process evaluations) are generally conducted during implementation to provide information on what is working and how efficiently, in order to determine how improvements can be made. '*When a cook tastes the soup, that's formative evaluation*' (Professor Robert E. Stake). *Summative evaluations* are undertaken (i) at or close to the end of an intervention or at a particular stage to assess effectiveness and results and also (ii) sometime after the conclusion of an intervention to assess impact. '*When a guest tastes the soup, that's summative evaluation*' (Professor Robert E. Stake).

summarises the key elements of an evaluation: rationale, purpose, objectives, methodology, setting, and organisation.

Therefore, rather than labelling an evaluation approach according to a specific evaluation typology, it is of essence that each specific evaluation should have clear objectives, and that the purpose and emphasis of the evaluation should be tailored to meet the objectives most appropriately. It should be clear if the emphasis is on policy, process and management issues; or on results, including outcomes and impact of the interventions under study; or on a mix of both process issues and results at various levels.

> *The purpose, scope and evaluation questions determine the most appropriate approach and methodology for each evaluation. An inception report can be used to inform the selection of an evaluation approach.*
>
> *The methodology is developed in line with the evaluation approach chosen. The methodology includes specification and justification of the design of the evaluation and the techniques for data collection and analysis. The selected methodology answers the evaluation questions using credible evidence. A clear distinction is made between the various result levels (intervention logic containing an objective-means hierarchy stating input, output, outcome, impact).* (OECD/DAC Quality Standards section 2.9)

The evaluation methodology is the term covering the methods to be applied to meet the overall purpose and objectives of the evaluation. The particular methodology to be used for data collection and analysis is determined by the subject and purpose of the evaluation; it is developed to ensure that professionally adequate methods are used to meet the objectives and answer the questions posed by the evaluation.

The methodology encompasses the main scope (duration of evaluation period and activities to be covered); sampling considerations at various levels (countries, sectors, themes, cases); and the mix of quantitative and qualitative methods of data collection and analysis to answer the evaluation questions (See Box 6). Other methodological elements may be included in a specific evaluation. The available budget and the time frame also influence the boundaries for what is methodologically possible, and vice-versa; the chosen methodology has implications for the budget. Practical opportunities, constraints and new technologies can lead to the application of interesting and innovative combinations of methods; thereby contributing to making evaluation a dynamic field of methodological evolution.

> **Box 6: Case studies**
>
> Case study is a methodology that is often applied in development evaluation, not least when dealing with complex thematic evaluations. Overall, a case study is a method for learning about a complex instance (a case), based on a comprehensive understanding of the case obtained through extensive description and analysis of the case as a whole and in its context. Case studies can be used for a range of purposes, including the following:
>
> - **Illustrative case studies,** adding depth and nuance through exemplification;
> - **Exploratory case studies,** capturing new experiences and forming hypothesis;
> - **Explanatory case studies,** assessing causal links for implementation or effects;
> - **Critical instance case studies,** examining best or worst case scenarios; a critical test.
> - **Cumulative case studies**, combining findings from many case studies (and potentially other data sources), to answer an evaluation question.[12]
>
> In current evaluation, case studies often aim to contribute to conclusions with application beyond the specific case, and with case studies as one part of a theory-based mixed methods design. This is highly challenging, and the international evaluation community has explored and clarified the potential – and the analytical requirements - for using case studies for causal inference and for reaching conclusions with a degree of generalisability or transferability.[13]
>
> What is true for all case studies is that the sampling of cases is paramount to the conclusions and to their potential for providing conclusions of relevance and with the needed solidity. This again requires a thorough overview and understanding of the "universe" for sampling, as well as considering the case studies in relation to an understanding of the Theory of Change for the intervention(s).

The selected evaluation methodology aims to ensure that the most appropriate methods for data collection and analysis are applied in relation to particular evaluation objectives and questions. Evaluation methodologies are derived from research standards and methods. The close link between research and evaluation methods is beneficial to evaluation and vice versa. Research methods that are both tested and innovative inspire and strengthen the methodological rigour of evaluations.

There are many combinations of approaches and methodologies for evaluations and new combinations are constantly being generated, making each evaluation unique. LEARNING encourages triangulation of methods, data collection, and data analysis based on a thorough understanding of the evaluation topic. All evaluations conducted for LEARNING must be based on evidence and must explicitly consider limitations related to the analysis conducted (e.g. due to security constraints or lack of data). In real-life settings, availability or quality of existing data may be limited, and it is often important to collect and consider a broad range of types of data, information and evidence, to establish as credible and nuanced picture as possible. Overly reliance on one data type will often either limit the analysis and/or endanger the credibility, and multiple types and sources of data will also support triangulation (see also section 7.2).

---

[12] See Morra, L, Friedlander, A: Case Study Evaluations. World Bank Operations Evaluations Department [no date].

[13] See for instance Raimondo, E. The Rigor of Case-Based Causal Analysis, World Bank Independent Evaluation Group, 2023 for an overview of elements of this work, including contributions from MFA of Denmark.

Internationally, some evaluation types, most notably various forms of impact evaluations, are receiving considerable interest and resources from some donors. At the centre of the discussion on the methodology of impact evaluations are the issues of attribution, contribution, and establishment of a counterfactual (see Box 7). LEARNING has participated in impact evaluations and follows the discourse around new methodologies closely.[14]

Box 7: Attribution, contribution and counterfactual analysis

Different usage and definitions exist for the concepts below. Thus, the definitions in this box may not be shared be everyone; the intent here is to explain the broad meaning of the concepts.

Attribution refers to the extent to which observed development effects can be attributed to the evaluated intervention itself (rather than being caused by other factors). In practice, the question of attribution is complicated and involves a process of identifying multiple determining factors. Testing of attribution requires availability of good quality data and information. Attribution analysis is typically based on experimental or quasi-experimental quantitative methods.

Contribution refers to whether or not the evaluated intervention is part of the cause of an observed change, possibly including "ranking" the evaluated intervention among the various causal factors in play. Contribution analysis is often based on mixed methods and theory-based approaches.

When assessing whether observed change can be *attributed* to an intervention, a *counterfactual* analysis is a core element It is a comparison between what has actually happened after an intervention, and what would have happened in the absence of an intervention. This means that data must be collected from two different groups: the "treatment group" and a "comparison group". It is important that the comparison group possess similar observable characteristics to the treatment group. The comparison may be done by combining a well-designed baseline study, carried out on both a treatment and a comparison group, before the interventions are initiated, with an ex-post study for the same two groups. In the absence of a baseline study, it may be possible to construct a reliable counterfactual with existing good quality datasets (such as population censuses, national surveys etc.).[15]

In many cases, development interventions and results cannot be linked in a meaningful way to specific treatment and comparison groups. Here, other approaches for assessing whether and how observed changes are caused by an intervention, are relevant. This include realist evaluation, contribution analysis, qualitative-comparative analysis etc.[16] These approaches investigate the causal links or pathways behind the observed change, considering not just the intervention at hand, but the role of context and other influencing factors.[17] This way, it is possible to assess whether changes at outcome or impact level are results of the intervention at hand, as well as the role and importance of other factors; thereby identifying the so-called "credible contribution".

---

[14] For more on impact evaluations, see: The International Initiative for Impact Evaluations (3iE) www.3ieimpact.org and the World Bank Development Impact Evaluation initiative (DIME) www.worldbank.org/dime

[15] See for instance Hansen, H; Klejnstrup, N; Andersen O.W. (2013) "A Comparison of Model-Based and Design-Based Impact Evaluations of Interventions in Developing Countries." American Journal of Evaluation 34 (3).

[16] see Stern, E et al, 2012 "Broadening the range of designs and methods for impact evaluations", DFID.

[17] For more on approaches to assessing cause-effect relationships without quantitative counterfactual analysis see White, H and Phillips, D, 2012 "Addressing attribution of cause and effect in small n impact evaluations". 3ie.

4.5 Terms of Reference

The design phase concludes by the finalisation of the Terms of Reference (ToR) by summarising the decisions of the scoping and design phases. The ToR also describe the competences of the evaluation team, which are used as selection criteria in the tender process, and the processes of implementation and management. The ToR usually specify the requirements for the various phases of the evaluation, i.e. inception phase, literature or desk study, fieldwork, analysis and synthesis/reporting.

The ToR for an evaluation are structured along the sequential logic of design discussed in the sections above. Below, the common structure for ToR of Danida evaluations is shown. The structure may be different in joint evaluations, but the point is that the ToR should follow a logical sequence.

- Background,

- Objectives and scope,

- Evaluation criteria and evaluation questions,

- Approach and proposed methodology,

- Competencies required of the evaluation team,

- Tentative timing of the evaluation process,

- Key documents.

The ToRs for ongoing evaluations are posted on LEARNING's website ([Evaluation of development assistance (um.dk)](um.dk)) under "ongoing evaluations".

CHAPTER 5. EVALUATION OF MULTILATERAL DEVELOPMENT COOPERATION

This chapter lays out key aspects of evaluations of multilateral organisations and cooperation with multilateral organisations on particular evaluations; it elaborates on the specifics of designing, managing and implementing such evaluations. It should also be noted that the planning of the evaluation programme (discussed in Chapter 2) may include planning of evaluations of and with multilateral organisations. Similarly, the Danida evaluation policy principles (Chapter 1), the analytical standards of evaluation (Chapter 4), the scoping phase (Chapter 3), and the discussion on dissemination on evaluation findings (Chapter 8) also all apply to work with multilateral development cooperation.

5.1 Multilateral cooperation

Funds to multilateral organisations are channelled as core funding (not earmarked) and non-core funds (earmarked). In the first case, funds are pooled and the results and impact of individual contributions cannot be assessed in isolation. Assessments must focus on the effectiveness of the receiving institution as such.

According to the OECD/DAC, The United Nations, the World Bank and other 200 multilateral agencies and global funds received about one third of total ODA in 2022.[18] When including earmarked funding provided to multilaterals, this goes up to 40%. For Denmark, the similar figures have gone up from approximately 40% in 2013 to approximately 60% in 2021 (covering both core support and earmarked, so-called multi-bi support. This indicates donors' views of it as an important channel of development co-operation.

Evaluating multilateral cooperation poses particular accountability and evaluation challenges. The effectiveness of multilateral aid and the possible benefits to individual donors and recipients of funding channelled through multilateral organisations have been recurrent issues in many countries, including Denmark, not least because donor visibility is generally less in multilateral aid than in bilateral cooperation. There is a perceived information gap with regard to effectiveness and efficiency of core multilateral support, and a range of initiatives are in place to remedy this, as yet without fully closing the gap.

5.2 Collaborating for mutual accountability

The changed architecture of aid and the increasing interest from the general public and policy makers have contributed to a renewed interest among donor agencies in better evidence of multilateral impacts and the effectiveness of multilateral contributions. At the same time, there is an overall agreement for bilateral donors to avoid carrying out their own individual evaluations of core support to multilateral organisations. Thus, Denmark works

---

[18] OECD/DAC 2022 Multilateral Development finance, Multilateral Development Finance - OECD , and OECD/DAC 2022 Comparing Multilateral and Bilateral Aid. https://web-archive.oecd.org/2022-11-28/626382-2022-mdf-comparing-multilateral-bilateral-aid.pdf

through governing boards of multilateral organisations to enhance the organisations' own evaluation functions, and has supported the creation of a new Sector-Wide Evaluation office under the UN Secretary General (SWE). Denmark is also ready to support peer reviews of multilateral evaluation functions with a view to supporting efforts to improve evaluations. Further, Denmark is an active partner in the Multilateral Organisation Performance Assessment Network (MOPAN). This is an expanding network of 22 donor countries (of which two are observers). Together, the MOPAN members, observers and collaborators provide USD 100 billion in annual contributions to and through the multilateral system – the majority of the system's ODA funding – and the members have a common interest in assessing the organisational effectiveness of the major multilateral organisations they fund. The MOPAN assessments provide a multidimensional snapshot of organisational performance, assessing strategic, operational, relationship, and knowledge management, of selected organisations, as well as their results. As such, the assessments are important tools for member states to push for greater effectiveness in the multilateral system.[19]

Some multilateral agencies (e.g. the World Bank) act as "pace-setters" in development of evaluation approaches and methodologies. Being large organisations with strong pressures for accountability from their donors and boards and the beneficiaries of their programmes, they have an obligation to continuously develop and refine their approaches and methodologies and to publish and share their evaluation results. Therefore, they are often at the forefront of the learning curve on evaluation, and collaboration with these institutions on evaluation may at times have a capacity-building effect on bilateral donors and in partner countries.

Contributions through multilateral organisations may be evaluated as part of a broader evaluation of a development theme or modality. Denmark will also seek to work in collaboration with multilateral organisations and other donors on joint evaluation of issues of common interest, which will as a by-product also enhance Danish insight into the evaluation function of a given organisation.

Efforts are also being made through the OECD/DAC evaluation network to explore the possible usefulness of meta evaluations, i.e. reports that synthesise results from evaluations by the multilaterals' own evaluation functions. The aim is to contribute to the determination of the development effectiveness of particular multilaterals.

Despite these initiatives, there is still a perceived information gap, but it should be noted that new initiatives such as the SWE are too recent to have come to fruition at this point.

5.3 Approach to evaluating multilateral cooperation

Most evaluations of the interventions conducted by multilateral organisations are designed and conducted by the evaluation units established in and by the multilateral organisations

---

[19]The methodology used in the assessments has been adjusted several times since the launch of MOPAN in 2002, currently with the "MOPAN 3.1", that has taken the UN development system reform into consideration. Due to the diversity of organisations' mandates and structures, MOPAN does not compare or rank the organisations being assessed. For more information regarding MOPAN, the methodology and the various assessments, see MOPAN | Multilateral Organisation Performance Assessment Network (mopanonline.org)

themselves.[20] The evaluation units differ in size, structural set-up and approach to designing and managing evaluations. The quality of the evaluation functions and evaluations produced likewise vary from one agency to another, but considerable efforts are being made to improve the standards of the evaluation function and the quality of evaluations. Peer reviews of UN agencies are instrumental in this regard (See Box 8).

Box 8: Peer reviews

As a way of supporting improvements in the quality of evaluation functions of multilaterals, the OECD/DAC Evaluation Network and the United Nations Evaluation Group (UNEG) have joined forces to conduct peer reviews of the evaluation functions of UN agencies. The peer reviews use UNEG norms and standards and provide pointers, which gauge the extent to which donors (including the Danish MFA) can rely on the organisations' evaluations to feed into their accountability systems. Recognising the importance and usefulness of peer reviews[21], Denmark has been actively involved in their initial conceptualisation and participates regularly in them(see www.uneval.org or www.oecd.org/dac for further information).

The development banks (WB, IMF and the regional banks) have developed their own network of evaluation functions; and some peer reviewing of evaluation functions within this group of institutions is also taking place. Furthermore, MOPAN assessments include reviews of the organisation's evaluation function.

LEARNING also actively engages in and supports Danish representations who follow the evaluation practices and programmes of the multilateral organisations. This may include advising representations with responsibilities for particular multilateral organisations on their dialogue with these organisations concerning development of evaluation programmes and specific evaluations (TOR, draft reports, use of evaluation findings and follow-up).

In some cases, joint evaluations of a particular theme or intervention with the evaluation office/department in a multilateral organisation are also conducted. In such cases, the evaluation departments involved need to develop and agree on a joint understanding of the principles, approach and methodology to be used for the evaluation based on their respective policies and guidelines.

---

[20] World Bank's Independent Evaluation Group www.worldbank.org/ieg; The African Development Bank's Operations Evaluations Department www.afdb.org/opev; United Nations Evaluations Group www.uneval.org; The Asian Development Bank's Independent Evaluation Department www.adb.org/evaluation

[21] Participation in peer reviews of other evaluation functions also serves as a "mirror function" in so far as good practices in the multilateral organisations can act as inspiration for the bilateral evaluation units.

CHAPTER 6. MANAGING AND ASSIGNING AN EVALUATION

This chapter explains the roles of the various bodies managing, advising and implementing an evaluation commissioned by or in cooperation with the Danish MFA. The chapter also describes the contractual procedures and the role of the evaluation team.

6.1 Functions in implementation of an evaluation

The overriding principle is that an evaluation process must be free of any bias and open and transparent in order to be credible. The institutional framework, both of the planning and management of an evaluation, must adhere to that principle throughout the process and in the functions established. Those managing and advising the evaluation must be free of decision-making and implementation interests in the Danish MFA and other agencies. Those implementing the evaluation (the evaluation team) must be free of external pressure, and there should be no conflict of interest or any previous involvement in the intervention being evaluated.

A management structure is set up for each evaluation. Generally, the main management and oversight bodies are the Evaluation Management (EM), which is in charge of day-to-day management and quality control, and the Evaluation Reference Group (ERG), which is a technical and advisory body. In joint evaluations with a large number of partners, the governance structure may differ from the normal practice. To strengthen the quality control aspects, the Evaluation Management may engage external evaluation peer reviewers, who are called upon to comment on draft reports because of their particular subject matter expertise or their in-depth knowledge of evaluation methodology/report writing (See also Annex 2 on quality assurance and quality control principles).

The tasks assigned to various bodies may vary depending on the types of evaluation conducted and the agencies and other partnerships involved.

The point to be underlined is that, in each and every evaluation, the organisational set-up must be unambiguous and the roles of each party must be clear. A Memorandum of Understanding is at times used to formalise the agreements, e.g. in the case of joint evaluations with partner countries.

> *The governance and management structures are designed to fit the evaluation's context, purpose, scope and objectives.*
>
> *The evaluation governance structure safeguards credibility, inclusiveness, and transparency. Management organises the evaluation process and is responsible for day-to-day administration. Depending on the size and complexity of the evaluation, these functions may be combined or separate.* (OECD/DAC Quality Standards section 2.11)

The role of the evaluation team is usually described in the TOR for the evaluation and further elaborated on in the contract between the commissioning party (LEARNING or lead partner on the evaluation, if different from LEARNING, for instance for joined evaluations) and the company or organisation that has been selected for the assignment.

6.2 Evaluation management

When the MFA is the contract holder for an evaluation, it is LEARNING who, in collaboration with OKO (the Ministry of Foreign Affairs' Finance Department), is responsible for managing the contractual relationship with the selected evaluators. LEARNING must ensure that the evaluation is carried out in accordance with the Terms of Reference, Danida's evaluation policy and guidelines, and good evaluation practice. This includes safe-guarding the independence and impartiality of the evaluation, and ensuring that there is no undue influence or pressure in play in the evaluation process. LEARNING also holds the responsibility to ensure that evaluations are carried out in a cost-effective and timely fashion.

For joint evaluations, the Evaluation Management (EM) typically comprises of one or more representatives from participating evaluation departments, including LEARNING and when relevant, institutions from partner countries or other donor agencies with a mandate to conduct evaluations, depending on the specific evaluation.

During the preparation, the main tasks of the EM include drafting of the TOR and participation in tender processes, often supported by an independent tender consultant. The EM also provides relevant background documentation to the evaluation team and organises meetings with key informants at headquarters and in representations.

An important function of the EM is to carry out quality control throughout the implementation of the evaluation process. In so doing, management seeks to ensure that the evaluation addresses all the evaluation questions listed in the ToR and that the evaluation report assembles findings based on solid evidence and high quality and consistent analysis; the report must also be consistent and set out a clear link between findings, conclusions and recommendations. Moreover, the role of the EM includes sharing of draft reports among relevant stakeholders and calling for comments; conveying consolidated comments to the evaluation team and discussing perspectives to facilitate the work of the team, engaging possible peer reviewer(s); and signing off on/approving final versions of the inception report, work plan, progress reports, and the evaluation report. An important part of the EMs role is thus to engage with sufficient depth and knowledge of details to be able to ensure the quality of the evaluation, but at the same time not overstep or become involved in micro-management, as this may lead to a situation where the evaluation is not seen as sufficiently independent. A key dividing line is thus between assessing or posing questions regards the strength and consistency of the evidence and the analysis (as the foundation for the conclusions and verdicts), which is typically a core element in the quality control process; and questioning the conclusions and verdicts in their own right (Annex 1 explains more regarding codes of conduct, and annex 2 describes the quality assurance and quality control processes).

The EM may also organise, facilitate and participate in workshops to discuss and disseminate (preliminary) findings and recommendations.

Once the evaluation is completed, LEARNING presents the evaluation results and the follow-up as suggested by the responsible embassy or department to relevant Danish authorities, i.e. the Danida Programme Committee, senior management, the Council for

Development Policy, the Minister for Development Cooperation and Global Climate Policy. The evaluation summary and information regarding the publication of the evaluation is also forwarded to the Finance Committee and the Committee on Foreign affairs of the Danish Parliament (Section 8.1 has further details on follow-up and dissemination activities).

6.3 Evaluation Reference Group

The Evaluation Reference Group (ERG) is usually advisory and comprises in-depth expertise of relevance for the evaluation in question. It may also include representatives from relevant representations and departments in Copenhagen as well as Danish or international resource persons. In addition, the ERG may include partner country or partner organisation representatives; alternatively, a special reference group may be established in the partner country/countries involved, depending on the specific evaluation.

The tasks of an ERG include advising on factual information and methodological issues, including commenting the draft evaluation report. Depending on the specific process, ERG members may also comment on draft ToR. The ERG engages at different stages of the evaluation process, participates in evaluation workshops (as relevant); and supports the dissemination and follow-up of the evaluation recommendations.

The work of the group is in most cases based on a combination of in-person meetings and virtual communication through email and video conferences. The virtual communication enables the participation of members based in different countries, thereby making possible a broader feedback on draft reports and other aspects of the evaluation. This is often complemented by at least one stakeholder workshop, either at headquarters in MFA Copenhagen or in the country/countries covered by the evaluation. Such workshops help validate evaluation findings and may also be seen to contribute to the control of the quality of the evaluation.

6.4 Evaluation team

The independence and impartiality of the evaluation team is a core requirement in establishing the credibility of an evaluation. Screening takes place as part of the tendering process to ensure that this requirement is met. Members of the evaluation team represent relevant professional areas; a mix of international and country specific expertise is often requested. National/regional team members are included in most evaluation teams. The team leader is responsible for the team's performance, according to professional evaluation principles and standards.

The evaluation team prepares and carries out the evaluation according to the ToR and the contract, and in line with the present evaluation guidelines and the OECD/DAC quality standards. The team is accountable to the Evaluation Management. It is expected to conduct the evaluation process with a high degree of integrity and to apply the approved methodology to produce evidence-based and reliable findings, conclusions, lessons learnt and recommendations.

The evaluation team reports to the Evaluation Management regularly, organises stakeholder workshops in case study countries (when this is part of the ToR), participates in or organises validation stakeholder workshops towards the end of the evaluation process, and ensures systematic and documented processes for quality assurance.

6.5 Procuring the evaluation team

LEARNING procures the evaluation team through a tendering process. The contracting is typically with a consulting company, a research organisation, or a consortium. In cases of joint evaluations, the process may involve the participation of the partner country and/or representatives of other agencies involved in the evaluation. In such cases the lead partner's procurement policies and processes usually apply. Key selection criteria are the quality of the technical proposal of the tenderer, experience in evaluation, field-level competences, and experience relevant to the task.

Due to the magnitude of almost all evaluations, the budgets are above the EU threshold and thus follow the EU procurement directive. The Contract Notice for such assignments is advertised on EU's Tender Electronic Daily (TED) for all interested parties to respond to. Smaller assignments follow the relevant MFA tender procedure, dependent on the specific size and type of the project.

Tenders are appraised on the basis of topical content, professional composition, competence of the evaluation team, and price. Tender procedures, according to the EU procurement directive, require a preparation period of about four months; sometimes longer. A two-step process is often used. First, there is a short-listing of organisations among those who have forwarded an expression of interest. The short listing is based on an assessment of the references submitted of similar assignments previously carried out. On the basis of this assessment, a certain predetermined number (usually three to five) of prospective tenderers are invited to submit a proposal. The second step is the tender evaluation. As part of the evaluation of compliant tenders, interviews with the proposed Team Leaders may be carried out to clarify and confirm the details of the information provided in the submitted Team Leader CV. If interviews form part of the process, this will be communicated in the tender material, and the interviews will be carried out based on an interview guide that is identical for all tenderers. The MFA procurement website has further details and guidelines.[22]

> *A transparent and open procurement procedure is used for selecting the evaluation team.*
>
> *The members of the evaluation team possess a mix of evaluative skills and thematic knowledge. Gender balance is considered and the team includes professionals from partner countries or regions concerned.* (OECD/DAC Quality Standards section 3.1)

In situations where conflict of interest could occur, potential evaluators are excluded from participation, if their participation puts into doubt the independence and impartiality of the evaluation. This is typically linked to the risk of evaluators having to assess their own earlier

---

[22] Procurement (um.dk)

work or their earlier decisions, implying that having been involved in preparation or implementation of an engagement will typically lead to exclusion. In addition, any firm or expert having otherwise been involved in the engagement(s) under evaluation may be excluded from participation in the tender, if the involvement constitutes unfair competition.

> *Evaluators are independent from the development intervention, including its policy, operations and management functions, as well as intended beneficiaries. Possible conflicts of interest are addressed openly and honestly. The evaluation team is able to work freely and without interference. It is assured of co-operation and access to all relevant information.* (OECD/DAC Quality Standards section 3.2)

The specific criteria are set out in the tender material for each evaluation. Decisions on whether a conflict of interest exists rest with the Tender Evaluation Committee and are made on a case-by-case basis. Companies and individuals submitting proposals/bids shall provide all necessary information of relevance to the decision-making process as part of their tender. They shall self-assess the likelihood of a conflict of interest occurring. If a conflict of interest or issues implying unfair competition has not been disclosed as it should have been, the tender process will be cancelled.

CHAPTER 7. IMPLEMENTATION AND REPORTING

This chapter presents issues related to the steps of implementation and reporting in evaluations. Different phases of implementation are discussed, namely inception, field work, data collection and analysis. (See Box 9 for the sequence of the evaluation process.)

7.1 Inception phase

The inception phase provides an opportunity for the evaluation team, in agreement with the Evaluation Management, to further operationalise the ToR and finalise the evaluation approach and methodology. Key stakeholders may also be consulted. In most cases, the inception phase includes a documentation review, which helps strengthen the final development of the methodology. This includes clarification of potential sampling strategies and sampling criteria, the methods to maximize solidity and relevance of case studies, and the inclusion of relevant assumptions or requirements.

The inception phase may also be used to discuss issues related to the intervention logic and Theory of Change including e.g. the non-linear nature of intervention logics in complex interventions and how to conduct fieldwork in fragile situations. In such cases, methodological complexities, often combined with logistical challenges, must be taken into account. In fragile situations, poor or limited security may reduce the possibility of ensuring robust sampling. Alternative methods of data collection must be proposed at this stage. While the purpose and objectives for the evaluation has shaped both the selection of the evaluation criteria and the specific evaluation questions, the inception phase is also when additional clarification of interpretation and operationalisation will take place, typically involving the development of a detailed evaluation matrix.

At the end of the inception phase, the evaluation team produces an inception report, including a detailed operational plan for the conduct of the evaluation fieldwork. The draft report is usually shared with the ERG for comments and the draft report is discussed at a meeting or workshop. The Evaluation Management approves the final version of the report (see Annex 3 for a list of key issues to be covered by the inception report).

The evaluation team must apply and document its quality assurance process starting from this phase. Quality assurance must address key questions of methodology, e.g. reliability and validity of findings as well as security and confidentiality of data and information.

7.2 Fieldwork and data collection

The purpose of fieldwork (if part of the evaluation) is to systematically collect data and information from a range of relevant sources, using a solid sampling frame and methods for data collection. The data collection builds on the documentation analysis, it validates existing information, fills knowledge gaps, and ensures triangulation of information (see

Section 7.3). A mix of methods is applied to compensate for the respective weaknesses and biases in each of the data collection methods.

Data collection methods may include interviews, focus group discussions, surveys, SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis, participatory appraisal methods, structured observation, review of log books, analysis of data generated through monitoring systems, and new data sets generated from existing surveys.

Based on initial data analysis, preliminary findings of facts are formulated and validated with concerned parties.

Data collection and choice of methods should take into account the specific challenges such as language and adaptation to the specific contexts. Similarly, cultural sensitivities should be considered and respected both in the choice of data collection methods and in the manner in which the evaluation team gathers information.

Because evaluations often produce findings that are subject to disagreements, the choice of methods of data collection and analysis is critical; it needs to be of high standard and well substantiated.

LEARNING, as part of its quality control function, monitors that robust methodologies are applied, i.e. that evaluations use the methods that best answer the evaluation questions in order to ensure validity and reliability of findings and conclusions (see Annex 5 for further information on validity, reliability, sufficiency and evidence in data collection and analysis).

7.3 Analysis

The purpose of the analysis is to transform data into credible evidence. The collected data, using a robust sampling strategy as well as thorough contextual knowledge, is the basis for the analytical process.

The first step in the analytical process involves the cleaning, verification and organisation of data, so that they meet standards of quality, i.e. validity and reliability. Data also need to be organised according to the evaluation criteria and questions.

The second step is to aggregate the data to constitute findings that are relevant to the evaluation questions at hand. Although judgment is involved in the formulation and selection, findings describe facts and do not constitute or contain evaluative judgments in themselves. At this step, the evaluation team needs to ensure that the collected data have been triangulated, i.e. cross verified with data emanating from diverse sources and/or various methods of data collection, and that such data are sufficiently robust for analysis. Testing the findings through application of various analytical tools or perspectives may also strengthen the validity of the analysis.

The third step in the analytical process is to categorise, contrast, compare, and interpret the findings, i.e. formulate conclusions. This entails judging findings in relation to a reasonable expectation, e.g. a standard, criterion, benchmark, target, indicator, or good practice.

The "lessons learned" from an evaluation present selected key findings and conclusions and assesses these in the broader perspective of generality and wider applicability.

The evaluation team translates the conclusions and lessons learned into main recommendations, which can be considered by management and subsequently translated into more specific and detailed decisions and follow-up. Recommendations must be well founded and clearly supported by the data analysis and the conclusions drawn in the evaluation. In order to be useful, the recommendations also have to be as specific as possible, considering the evidence and the subject at hand.

A dialogue with key stakeholders (e.g. during validation workshops) aims to enhance the understanding of lessons learned and recommendations and thereby the usefulness of the evaluation. To ensure independent views, the final formulation of lessons learned and recommendations, however, rests with the evaluation team.

Box 9: Building sequential consistency of the evaluation process
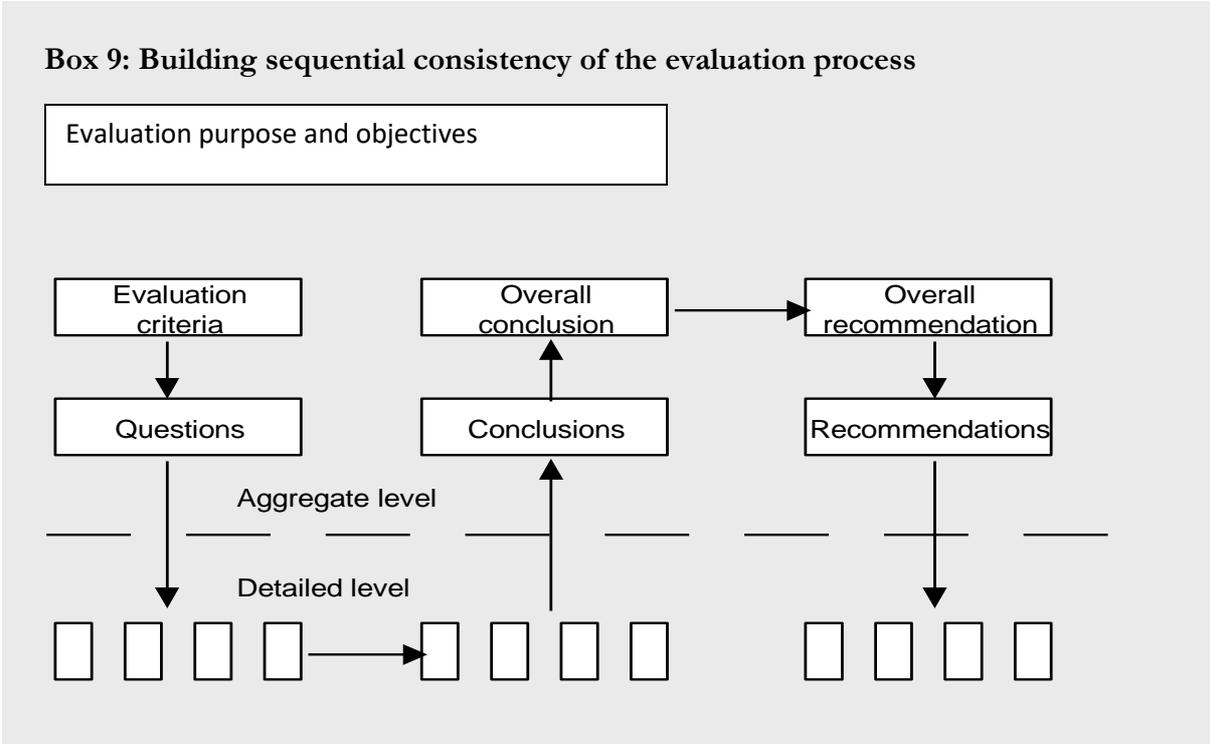
7.4 Reporting

The reporting on an evaluation starts as early as the inception phase when a report outline is presented based on the ToR and the introductory discussions with LEARNING. The reporting ends with the submission of the final evaluation report. The evaluation team is bound by contract to ensure that the contents of the ToR are adequately addressed in the evaluation report. Moreover, as stated in the OECD/DAC Quality Standards, reporting must also take into consideration the interests and background of the intended audience for the evaluation.

> *The evaluation report can readily be understood by the intended audience(s) and the form of the report is appropriate given the purpose(s) of the evaluation.*
> (OECD/DAC Quality Standards section 3.5)

Throughout the process, it is important to keep in mind that the report should be based on findings from *all* stages of the evaluation process. It is a common pitfall that findings from fieldwork dominate at the expense of findings from document reviews and/or initial interviews with key resource persons for instance.

LEARNING, other MFA staff, and the commissioning partners have no say on the evaluative judgements of the report. They assess the quality of the report and submit comments regarding factual information, methodological issues, consistency, conclusions, and clarity of recommendations, after which the team prepares the final draft version.

In accordance with the OECD/DAC Quality Standard, LEARNING emphasises and seeks short and precise reports, written in understandable and reader-friendly language.

The team leader is responsible for submitting the final report and summary in a structure and a layout that comply with the formal requirements of LEARNING and the MFA. In joint evaluations that are not managed by the MFA, other formal requirements may apply. Lay-out and writing guidelines are available and can be accessed online at Evaluation of development assistance (um.dk).

CHAPTER 8. FOLLOW-UP, LEARNING AND DISSEMINATION

This chapter provides an overview of MFA's internal follow-up procedures and the external dissemination of evaluations, as well as the learning aspects related to both.

8.1 Management response to evaluations and internal follow-up

When an evaluation has been finalised, a Management Response and Follow-up Note is prepared and discussed in the MFA's internal Programme Committee. This serves a dual purpose: First, the note provides the Programme Committee with information on the key findings and recommendations of the evaluation. Second, the note proposes a management response to the recommendations of the evaluation.

LEARNING coordinates the preparation of Follow-up Notes, while the responsibility for the draft management response rests with the representation/department responsible for the programme (or multilateral support) being evaluated. The management response is finalised by the responsible representation/department, based on the discussions in the Programme Committee, and submitted to LEARNING. LEARNING is responsible for a short summary of the evaluation, including a key elements of the management response. This is prepared in Danish.

Since evaluation plays a critical role in the learning cycle, the management of the MFA continuously monitors how findings and recommendations of completed evaluations are followed-up. The Secretary of State for Development Policy chairs an evaluation meeting once a year where follow-up to evaluations is reported and discussed with LEARNING and the units responsible for the implementation of the follow-up. This happens approximately two years after finalisation of the management response.

All evaluations and Danish summaries of evaluation reports are distributed to relevant representations and departments in the MFA as well as to the Minister for Development Cooperation and Global Climate Policy. It is the responsibility of operational departments and embassies to ensure that relevant past experience is built into the design and preparation of future activities. In addition, the evaluation summaries, including information regarding the publication of the evaluations, management responses etc. is also forwarded to the Finance Committee and the Committee on Foreign Affairs of the Danish Parliament, with an invitation to receive a verbal presentation, should this be wished for.

> *Recommendations are systematically responded to and action taken by the person(s)/ body targeted in each recommendation. This includes a formal management response and follow-up. All agreed follow-up actions are tracked to ensure accountability for their implementation.* (OECD/DAC Quality Standards section 4.2)

LEARNING also attends the presentation and discussions of early drafts of new programmes or new phases of existing programmes in the Programme Committee. The link established through this participation aims to ensure inclusion of evaluation-based knowledge in the preparation of new programmes or phases of programmes. More details on the functioning of the Programme Committee can be found at www.amg.um.dk

In addition, "brown-bag lunches", online meeting and seminars are organised in the MFA to disseminate lessons learned and follow-up actions related to evaluations as deemed relevant. Uptake of new knowledge based on evaluations may thus take place at individual level, but may also take place through a community of practice, such as a group of technical advisers or specialists with expertise in the topic, focal points etc. Similarly, internal meetings may be organised to discuss evaluation studies or other reports commissioned by LEARNING.

8.2 External dissemination

External dissemination of evaluations takes place through publishing of evaluations and evaluation studies at LEARNINGs website Evaluation of development assistance (um.dk)

Electronic links to evaluation reports are distributed through information on the MFA's news-site. Links to evaluation reports are also submitted to the database of the OECD/DAC Evaluation Resource Centre (DeREC).

LEARNING contributes to international efforts towards developing methods for evaluation of development cooperation through e.g. the organisation of international seminars or workshops and active participation in international networks such as OECD/DAC EvalNet and the Nordic+ evaluation network.

Other means of communicating evaluation-based knowledge include the organisation of public meetings in Denmark to present evaluations, lectures at universities in Denmark, and explaining the role of evaluation through video and film. This happens on an ad hoc basis, depending on the relevance and important of a particular evaluation and will often be in collaboration with a relevant actor, such as the "Global Focus", IDA Global, or the Danish Institute for International Studies.

> *The evaluation is designed, conducted and reported to meet the needs of the intended users. Conclusions, recommendations and lessons are clear, relevant, targeted and actionable so that the evaluation can be used to achieve its intended learning and accountability objectives. The evaluation is delivered in time to ensure optimal use of the results.*
>
> *Systematic dissemination, storage and management of the evaluation report are ensured to provide easy access to all development partners, to reach target audiences, and to maximize the learning benefits of the evaluation.* (OECD/DAC Quality Standards section 4.1)

The external users of evaluations are a diverse group of persons and organisations located in Denmark, partner countries and the international community. They comprise authorities, the media, politicians, civil society, private sector associations, researchers, consultants and professional agencies, among others.

> *The evaluation results are presented in an accessible format and are systematically distributed internally and externally for learning and follow-up actions and to ensure transparency. In light of lessons emerging from the evaluation, additional interested parties in the wider development community are identified and targeted to maximise the use of relevant findings.* (OECD/DAC Quality Standards section 4.3)

LEARNING continuously assesses how evaluations are used. The main feedback on existing outreach includes use of evaluation in planning of new development interventions, in management of existing activities, in development of policies and strategies, and in training of staff members and external stakeholders as well as the assessment of follow-up. In addition, specific follow-up studies may be carried out on an ad hoc basis.

ANNEX 1: CODES OF CONDUCT

The evaluation management

- At the inception stage, the Evaluation Management (EM) is responsible for briefing the evaluation team on the operations, the expected role of all parties involved in the process, relevant documents and data sources. After the draft report has been presented, all communication between the evaluation team and MFA/partner staff and other stakeholders should go through the EM.

- The EM is required to react to all requests for assistance relating to situations in which the evaluation team feels their independence questioned or threatened. If the evaluation team encounters insufficient assistance or outright resistance during the evaluation process, including when carrying out the fieldwork, it is the responsibility of the EM to contact the persons involved and ensure that proper cooperation be established. In the extreme situation where comments can be interpreted as subtle or overt pressure against the evaluation team to achieve specific conclusions, the EM has a particular responsibility to take immediate action.

- If the team feels that the EM does not address concerns regarding threats to the independence, or that the EM applies undue pressure, the team must take their concerns to the Head of Department for LEARNING. Ensuring critical assessments and scrutinising the consistency of statements or the foundations for conclusions in a draft report are an important part of the role of EM, consistent with the need to ensure independent evaluations of high quality. However, the option to take the dialogue to the level of MFA management, in case there is need to assess whether EM oversteps its role, is seen as a further process safeguard for independence. If so, EM must be ready to explain and clarify how comments, suggestions etc. are appropriate.

- If the evaluation team suspects mismanagement, corruption or other illicit practices, the evaluation team must inform the EM, which must ensure that the information is passed on to relevant departments or representations for appropriate action.

MFA Staff and other stakeholders responsible for the activities under evaluation

Actors in partner countries such as line ministries, private sector actors, and civil society organisations may in some cases be, along with MFA staff, key parties to the evaluation process.

Other parties responsible for the activities under evaluation may include organisations in Denmark (e.g. Danish civil society organisations), multilateral organisations, and other international institutions supported by Denmark. These stakeholders have the complex role of being both the object of an evaluation, key resource persons during the evaluation, and/or users of the results. They have a key role in assuring the usefulness of evaluation findings for their own learning processes. They facilitate evaluations, and they comment on the reports by pointing out factual errors and inaccuracies.

- When facilitating evaluations by providing contacts, references, information about activities and logistical support to the evaluation team, MFA staff and

other stakeholders are expected to respect the integrity of the evaluation team in making its own decisions about where to go and whom to see. They may certainly provide comments or background information on suggested sites for field visits and persons to be interviewed, but the final decision rests with the evaluation team in consultation with the EM.

- MFA staff and other stakeholders should assist with the identification of relevant documents, even if the material has not been specifically requested.

- If contributing to quality control of the report, MFA staff and other stakeholders should observe the right of the evaluation team to make conclusions and recommendations, which may not be shared by the MFA.

- Reservations regarding the competence of evaluation team members, the quality of the fieldwork, the quality of analyses, etc. should be reported immediately to the EM.

- With a view to protecting the evaluation team against undue pressure, all contact between the evaluation team and stakeholders should go through the EM, after the draft report has been produced. If the evaluation team meets with stakeholders in this phase, the EM should be present, and all correspondence between the evaluation team and stakeholders should be copied to the EM.

Evaluation teams

Independent evaluation teams carry out all evaluations and it is their responsibility to collect the information necessary to fulfil the tasks set out in the Terms of Reference for the assignment and to present findings, conclusions and recommendations in a well-structured and well-written report. The evaluation team prepares and undertakes the fieldwork and in this process consults with relevant MFA and partner staff as necessary.

Conclusions and recommendations of the evaluation report are discussed with the EM, MFA staff and, in case of joint evaluations, also with staff from other agencies, partners and possibly a reference group, but it is the evaluation team who has the ultimate responsibility for conclusions and recommendations.

- The evaluation team should be prepared to engage in discussions with MFA staff and other stakeholders without perceiving difference of opinions as a form of pressure. There is an inherent tension in any evaluation assignment, and the evaluation team should expect their analysis to be scrutinised and their interpretations to be challenged.

- Evaluation teams who believe that they have been subject to undue pressure or inappropriate behaviour from MFA staff or other stakeholders (including those responsible for the activities under evaluation) during fieldwork or while preparing the report, should report this immediately to the EM or, if needed, to the Head of Department for Evaluation, Learning, and Quality.

- It is the responsibility of the team leader to decide, whether it is appropriate for the EM and possible stakeholders to participate in meetings, interviews, and field visits. Stakeholders may include Embassy staff, Danida advisers and company advisers, representatives of other donors, partner representatives, other parties responsible for the activities under evaluation, and others. Issues relevant to the team leader's decision include ensuring the impartiality of the evaluation, ensuring openness of discussions, and avoiding imbalance in numbers (too many 'evaluators' present can make stakeholders uncomfortable).

- If the evaluation team encounters or suspects corruption or other illicit practice, while carrying out their evaluation, it is their responsibility to report this to the EM, which will ensure that the MFA can take appropriate action. Only cases sustained by legally valid evidence, or reported as c-cases and published at the MFA website, can be reported in an evaluation report.

- During the preparation and the undertaking of the fieldwork, the evaluation team is responsible for informing the EM about the time schedule of the evaluation and the persons to be contacted and interviewed. This information ensures coverage of important sources of information and helps to avoid duplication and the evaluation becoming a burden upon a small group of key resource persons.

- After the draft report has been produced, the evaluation team should avoid any direct contact to MFA staff and other stakeholders without prior agreement with the EM. If the evaluation team meets with MFA staff, the EM should be present, and all correspondence between the evaluation team and MFA staff should be copied to the EM.

- The evaluation team should correct all factual errors and inaccuracies and make changes related to report structure, consistency, the analytical rigour, the validity of evidence, and requirements in Terms of Reference, as pointed out by the EM. However, the evaluation team should only agree to make changes to conclusions and recommendations of the evaluation that they regard as qualitative improvements. Where consultants disagree with changes proposed by the EM, other MFA staff, and other stakeholders, they must present counter arguments. In cases of serious disagreement, the alternative assessments and counterarguments by the evaluation team should be presented in the report, if appropriate as footnotes. In cases of minor issues, the consultant must explain in writing to the EM why suggested changes are not considered relevant or appropriate.

- Statements should not be made on behalf of the evaluation team if all team members have not had the opportunity to express their agreement. Unless disagreement has been clearly expressed in the report, all team members involved in an assignment are expected to be loyal to the conclusions of the report. If internal team disagreement has not been reported, it is not appropriate that team members subsequently criticise the quality or the conclusions of the evaluation after it has been published.

- A particular case of internal team disagreement regards the situation in certain partner countries, where criticism can have serious consequences for national

members of the evaluation team. Therefore, the team leader should ensure that national consultants either endorse a report in its entity or that their limited roles are clearly described in the report.

- Participation by the evaluation team in the dissemination of evaluation findings after the publication of the evaluation report is appreciated. If the evaluation team contributes to a public debate about the interpretation of the evaluation, however, this has to be agreed with the EM. Particular attention should be paid to the issue of confidentiality of information obtained during the evaluation process, and the evaluation team should not orally or in writing present information, which is not included in the report.

ANNEX 2: QUALITY ASSURANCE AND QUALITY CONTROL

LEARNING makes a distinction between quality assurance and quality control, and this separation is also used in the text below:

Quality assurance -principle

Quality *assurance* is the responsibility of the evaluation team, and the standards to be followed are laid down in the technical bid and subsequently in the contract with the selected evaluation team. Quality assurance by the evaluation team must ensure that the evaluation follows the OECD/DAC quality standards for evaluations and the Danida Guidelines for Evaluation (2023) including the lay-out guidelines.

Minimum requirements in implementation

The evaluation team must:

- Understand and conduct all processes of the evaluation according to OECD/DAC Quality Standards. This includes upholding the independence, impartiality of the team and i.e. addressing key questions of methodology, e.g. reliability and validity issues, as well as security and confidentiality of data and information systematically throughout the contract period.

- Document its quality assurance processes in a transparent manner and be ready to share documentation of the quality assurance with LEARNING.

- Report to the Evaluation Management on a regular basis throughout the evaluation process e.g. by forwarding brief progress reports (or summary of progress in e-mails).

Quality assurance must be exercised throughout the evaluation process. It is the responsibility of the team to ensure that the procedures for quality assurance set out in the technical proposal and the contract with MFA, are adhered to before reports are submitted to LEARNING.

Quality control principle

LEARNING conducts quality *control* of all evaluations. Quality control is exercised throughout the evaluation process.

The OECD/DAC standards of quality control are adhered to, but the exact nature of the quality control arrangements for evaluations depends on the scope and complexity, and is decided upon when organisation and management for a particular evaluation is established. Further, the specific focus of the quality control will of course reflect the quality issues encountered in a given evaluation.

Quality control implementers

The Evaluation Management is in charge of day-to-day management and quality control.

The Evaluation Reference Group (if established) is in most cases a technical and advisory body, which contributes to quality control through its oversight at particular stages of the evaluation.

External peer reviewers (engaged on many evaluations) are used to strengthen the quality control function in particular subject matters, but may also contribute to control of evaluation methodology, consistency of the analytical work conducted etc. Internal sparring partners are typically in place to support the EM in their work related to quality throughout the process.

Participants in validation workshops may also contribute to the quality control of the evaluation.

Quality control processes

Quality control is carried out throughout the evaluation process and should help ensure that:

- The preparatory studies and Terms of Reference are coherent to ensure a clear logic between rationale, purpose and objectives and resources available for a planned evaluation.

- Tender procedures stipulate standards for quality assurance and clearly stipulate that these are part of the requirements of the tenderer. The tenderer's quality assurance set-up and approach is also rated as part of the technical proposal.

- The principles of independence and impartiality of the evaluation team is adhered to right from selection to completion.

- The inception report is coherent, and the approach and methodology meets the OECD/DAC Quality Standards.

- The fieldwork applies robust methodologies i.e. uses methods that best answer the evaluation questions in order to ensure validity and reliability of findings and conclusions.

- The evaluation report addresses all evaluation questions listed in the Terms of Reference, and the evaluation report is drawn up on the basis of evaluation findings based on solid evidence, high quality and consistent analysis and with a clear link between findings, conclusions and recommendations.

- Quality control also encompasses calling for comments to draft reports among relevant stakeholders and signing off/approval of final versions of the inception report, work plan, progress reports, and the evaluation report.

The list above is illustrative rather than exhaustive.

ANNEX 3: KEY ISSUES TO BE COVERED IN THE INCEPTION REPORT

The following outlines the requirements of the inception report. The report should include, but not necessarily be limited to:

- An overall logic model or Theory of Change for the intervention, depicting the linkages between resources (inputs), intervention activities (processes), intervention results (outputs or deliverables), intended outcomes (intervention objectives), overall impacts, as well as relevant assumptions and prerequisites. Further, the evaluation criteria in focus for the evaluation should be explained in terms of interpretation and operationalisation.

- Relevant aspects of design and approach, the final evaluation methodology including discussion of sufficiency and appropriateness of methods and alternatives if need be, explanation of sampling and data collection strategy with careful attention to the selection and use of cases, analytical framework and reporting outline. The description of the methodology must have sufficient focus on analytical issue and sufficient level of depth to present a clear argument for how valid conclusions and assessments will be reached.

- The hierarchy of evaluation questions starting from the general ones that are presented in the ToRs through to the specific ones that will produce data and information. In case of suggestions for adjusting selected questions, this should be explained and justified.

- A matrix indicating the nature and sources of evidence for each specific question with links to the criteria in play

- An explanation of the yardstick for evaluation assessment, if this is not self-evident from the methodology, matrix, and the criteria.

- A preliminary explanation of challenges and limitations encountered or envisaged.

- A schedule of activities.

- A communication and consultation plan (with stakeholders).

- In the case of evaluations with complex evaluation team organisation and logistics, e.g. joint evaluations, a systematic management plan that addresses key issues of management, coordination, authorities, responsibilities, etc.

- Draft outline of evaluation report.

ANNEX 4: KEY ISSUES TO BE COVERED BY THE EVALUATION REPORT

Because the focus and types of evaluations vary, there is no common format for reporting. However, LEARNING expects the report to be consistent with the DAC Evaluation Quality Standards and the following principles:

- The main report is relatively short, maximum 50 pages, unless otherwise agreed. It contains the aggregated information and highlights. All detailed information is referred to annexes.

- Information is organised in a simple format, which is easily accessible to the reader. This means that discussion of the evaluation criteria is a focus in the report.

The evaluation report consists of three levels of information. First, the executive summary is written as a self-contained paper that provides the bare essentials for decision-makers on background, major conclusions, recommendations and lessons learned (maximum 10 pages). The second level is the main report including conclusions and recommendations (40-50 pages). The third level in the report is the annexes. These provide all information necessary to substantiate major conclusions and recommendations in the main report. The Terms of Reference, the team's itinerary, list of persons met, and list of documents used should also be annexed.

All evaluation reports are published only on the web. Annexes are also available on the website (Evaluation of development assistance (um.dk)).

An evaluation report *could* be structured as follows:

- Executive Summary (which covers most aspects of the main report but in abbreviated form).

- Introduction, with the background for and implementation of the evaluation as well as a presentation and discussion of the methodology used.

- Setting in which cooperation interventions are implemented, in terms of geography, policy, sectors and organisations, as appropriate.

- Description of the planning, and implementation of development activities. This includes the presentation of objectives, target groups, components, financing, management, etc.

- Findings.

- Conclusions based on findings and analysis, including efficiency, effectiveness, impact, relevance, and sustainability.

- Lessons learned.

- Recommendations.

ANNEX 5: EVALUATION VALIDITY, RELIABILITY, SUFFICIENCY AND EVIDENCE[23]

This annex explains the concept validity, reliability and evidence, which may be used as a checklist in evaluation processes and support the quality control and quality assurance processes.

Validity

Validity is a measure of the extent to which, taken together, the evaluation's design, data collection methods and analyses provide a reasonable basis for conclusions about the evaluation's questions.

Some of the most common pitfalls of validity may be addressed and mitigated in the design phase. In a (mainly) qualitative evaluation design, this includes assessing the comprehensiveness of data sources, consider the cultural competence of data collectors and consider the adequacy of data analysis techniques and team capacity. In a (mainly) quantitative design, this includes considering whether random sampling is appropriate, the sampling size is sufficient and if there is a potential sampling bias. It is relevant to consider whether key indicators have been properly identified and whether measures of them are likely to be accurate, as well as to consider whether statistical procedures have been appropriately selected and whether there is sufficient expertise for their use.[24]

Reliability

Reliability is a measure of the quality of measurement; information is reliable if the measurement procedure yields the same results if applied repeatedly. Reliability is a key factor for the quality of the evaluation and, as such, the evaluation team is expected to incorporate into its methodology ways of estimating the reliability of the data it gathers.

One of the most commonly used techniques for approximating the reliability of data is triangulation, applying the same measurement procedure to different sources to obtain data that can be compared for similarity or using different data collection methods with the same type of respondents.

There are different forms of reliability. For instance, where triangulation is applied to three different categories of data (verbal, documentary, and observation); or the consistency of a measure is tested from one time to another, e.g. where the same household survey is administered at periodic intervals. It can also be when it is tested to which extent different raters/observers give consistent estimates of the same phenomenon, e.g. an assessment of the fairness and freedom of elections by different observers.

The evaluation team is expected to consider the relationship, i.e. the trade-offs, between issues of validity and reliability, as part of the development of the methodology.

---

[23] This annex is in parts an updated version of a section in Danida: Evaluation Guidelines 2006 and 2018.
[24] Bamberger,M., Rugh,J.,and Mabry,L.,2006: Strengthening the Evaluation Design and the Validity of the Conclusions. Chapter 7 in Realworld Evaluation: Working under Budget, Time, Data and Political Constraints, Sage Publications, California.

Sufficient and appropriate evidence

When making choices about the amount and nature of data to gather, it is expected that the evaluation team will collect only the information required to answer the evaluation questions. Sufficiency has to do with the amount of information required to provide persuasive support for the contents of the evaluation report, i.e. will the collective weight of the evidence be sufficient to persuade a reasonable person that the observations and conclusions are valid, and the recommendations appropriate.

Some of the factors to consider when judging sufficiency are:

- The quality of the data, i.e. its relevance, reliability and validity;

- The significance of the finding and conclusion the data are intended to support, e.g. how important is it?

- How much assurance is intended, e.g. is the evaluation important for accountability purposes?

- What is the risk of making an incorrect observation or reaching an invalid conclusion? Does absence of documentation of results imply an absence of evidence or an evidence of absence?

- What is the cost of obtaining additional information in relation to its additional benefits, i.e. in terms of support for observations and conclusions?

Appropriateness of data includes questions of reliability and validity, and of relevance, i.e. the extent to which information bears a clear and logical relationship to the evaluation criteria and questions.

Data are commonly categorised as verbal, documentary, or observational. Ss a rule of thumb, observations are considered the most robust type of data followed by documentary and verbal. It should be stressed that in real-world situations, data collection and use will often require building on a mix of information sources and data types, to allow for reaching evaluative verdicts based on as comprehensive picture as possible, and with the required triangulation.