Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Departement of Economic Affairs,
Education and Research EAER
State Secretariat for Economic Affairs SECO

GUIDELINES

Economic Cooperation and Development WE

Evaluation Guidelines

Version	Gever ID	Reference ID
2.0 (01-Feb-2025) second version	SECO-D-7FDA3401/1091	EVAL-2-GDL

Recipients / Applied to

All staff of the *SECO Economic Cooperation and Development (WE) division*, both at headquarter and abroad, and partner organisations conducting evaluations for SECO-WE.

Owner / Author	Approved/authorised by
Malacarne, Nicola Evaluation Manager	Schneider, Johannes Head Quality and Resources (WEQA)

These guidelines put the *Evaluation Policy* into practice by setting minimum standards for all SECO-WE commissioned evaluations. If a partner organisation commissions an evaluation of a SECO-funded project, these standards should be considered as well.

Table of Contents

List of abbreviations **3**

List of boxes, figures and tables **3**

0 Purpose and Scope of Guidelines **4**

1 Evaluation Fundamentals **4**

1.1 Definition 4

1.2 Purpose 5

1.3 Principles 6

1.4 Criteria 7

2 Evaluation Typology **8**

2.1 Overview 8

2.2 Mid-term, End-term and Ex-post Evaluations 8

2.3 From Process to Impact Evaluations 13

3 Evaluation Process **14**

3.0 Step 0: Choice of analysis tool or process 14

3.1 Step 1: Conceptualising the evaluation 14

3.2 Step 2: Selecting evaluators 16

3.3 Step 3: Initiating the evaluation / inception phase 16

3.4 Step 4: Conducting the evaluation 16

3.5 Step 5: Writing the evaluation report 17

3.6 Step 6: Using and communicating the evaluation results 17

4 Annexes **18**

Annex I. Chronology of Versions 18

Annex II. Example of Sustainability Definition 18

Annex III. Extended Evaluation Typology 19

Annex IV. Evaluation Report Checklist 22

Annex V. Standard Format for Evaluation Reports 23

Annex VI. Assessment tool and processes other than evaluation 24

Annex VII. General Evaluation Decision Table 25

Annex VIII. Impact Evaluation Decision Table 26

Annex IX. Evaluation Matrix 27

List of abbreviations

Abbrev.	*Signification*
DAC	Development Assistance Committee
M&E	Monitoring & Evaluation
OECD	Organisation for Economic Co-operation and Development
SDC	Swiss Agency for Development and Cooperation
SECO	State Secretariat for Economic Affairs
SEVAL	Swiss Evaluation Society
ToR	Terms of Reference
WE	Division of Economic Cooperation and Development within SECO (i.e. SECO-WE)

List of boxes, figures and tables

Box 1: Definition of Evaluation, based on OECD (2023) and SECO-WE Evaluation Policy. 4
Box 2: Considerations to conduct an evaluation 5
Box 3: Biennial Performance Report 8
Box 4: OECD DAC Criteria Rating 8
Box 5: Is a Mid-term Evaluation the right tool? 10
Box 6: Minimum standards Mid-term Evaluation 10
Box 7: Is an End-term Evaluation the right tool? 11
Box 8: Minimum standards End-term Evaluations 11
Box 9: Is an ex-post Evaluation the right tool? 12
Box 10: Minimum standards Ex-post Evaluations 12
Box 11: Support by WEQA 14

Figure 1: Purpose of Evaluation (own illustration) 5
Figure 2: Evaluation Typology (SDC 2024). 8
Figure 3: Decision tree on the importance of establishing causal relationships (own illustration). 13
Figure 4: Evaluation process flow chart 14
Figure 5: Flowchart of conceptualising an evaluation (own illustration) 15

Table 1: Linking DAC criteria with evaluation timing (own illustration) 9
Table 2: Two categories of sustainability (USAID, 2021) 12
Table 3: Top 4 tips on conceptualising the evaluation 14
Table 4: Tips on formulating evaluation questions 15

0 Purpose and Scope of Guidelines

These guidelines are designed to implement SECO-WE's Evaluation Policy. The policy defines three evaluation categories:

External Evaluation	Commissioned by the Programme Manager, overseen by the Head of the Operational Section, and carried out by independent external evaluators at the project level.
Independent Evaluation	Commissioned by WEQA, overseen by the External Evaluation Committee, and conducted by independent external evaluators at a thematic or portfolio level.
Internal Review (Completion Note)	A self-assessment conducted by the Program Manager at the project level, overseen by the Head of the Operational Section.

These guidelines explain the purpose, principles and criteria of evaluation. They further describe the evaluation typology and definitions used at SECO-WE, including concepts like process and impact evaluations. Lastly, they elaborate the evaluation process from a commissioner's perspective. The guidelines are written in a concise manner providing tips and examples.

Most elements of these guidelines are harmonised with SDC's (2024) Evaluation Guidelines and the corresponding SDC (2025) How to-Note. Common Evaluation Minimum Standards of the Swiss International Cooperation (IC)[1] as well as close coordination in a joint Evaluation Steering Committee, enabled this harmonisation.

1 Evaluation Fundamentals

1.1 Definition

The SECO-WE Evaluation Policy defines "evaluation" in line with OECD standards. In this guideline, we clarify the definition further:

Box 1: Definition of Evaluation, based on OECD (2023)[2] and SECO-WE Evaluation Policy.

(i) Evaluation is the systematic and objective assessment of a (or several) planned, on-going or completed project or programme (or any of its components), covering its design, implementation, and/or results.

(ii) Through evaluation, SECO-WE aims to determine relevance, coherence, effectiveness, efficiency, impact, and/or sustainability. To ensure accuracy and feasibility, evaluations should focus on the most relevant criteria, maintaining clarity of purpose and objectives.

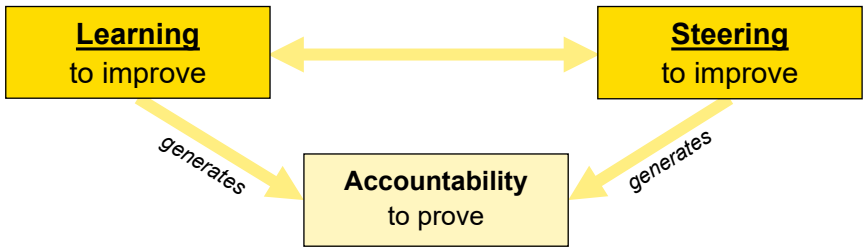
(iii) Evaluations should be useful, feasible, correct, and accurate, helping to incorporate learnings into decision-making by SECO, its partners, and recipients.

[1] SDC, SECO-WE, PHRD (2025). Swiss International Cooperation: Evaluation Minimum Standards.
[2] OECD (2023). Glossary of Key Terms in Evaluation and Results-Based Management.

1.2 Purpose

The Evaluation Policy outlines *Learning*, *Steering* and *Accountability* as three purposes of evaluation.

Figure 1: Purpose of Evaluation (own illustration)



Evaluations should prioritise which of the three purposes is most important. Since SECO-WE's Monitoring & Evaluation (M&E) system focuses on improving projects, evaluations that emphasise learning and steering are encouraged. By demonstrating continuous improvement through M&E, SECO-WE will also contribute to accountability to Parliament, the public, and other stakeholders.

SECO-WE has no prescriptions, but indicative guidance, on what, when, or where to evaluate for Independent and External Evaluations. For Independent Evaluations, the Evaluation Committee, supported by WEQA, sets priorities. For External Evaluations, the Operational Sections decide what and when to evaluate, depending on operational needs.

Box 2: Considerations to conduct an evaluation

Being clear on **why an evaluation will be conducted** is crucial as it establishes its foundation. The following questions can help guide this process:

- Where are the gaps in evidence or knowledge? What do we already know, what do we not know, and what do we want to learn more about?
- Which of these gaps bear importance beyond a single project for the wider programmatic orientation of the organisation (Operational Sections or SECO-WE)?
- Which areas of our projects are not meeting expectations (based on monitoring and reporting)? Where do we want to improve and explore further?
- What information is needed for upcoming steering decisions (e.g. follow-up phase)?
- Are there any accountability requirements that monitoring and reporting cannot fulfil? What do we need to demonstrate to Parliament, the public, and other stakeholders?

1.3 Principles

SECO-WE evaluations are guided by the four principles of the Swiss Evaluation Society.[3] For each principle operational minimum requirements are elaborated. These requirements are in line with the Evaluation Minimum Standards of Swiss International Cooperation (IC)[4] jointly formulated by SDC, SECO-WE and PHRD.

Utility	Evaluations should meet the needs of intended users, providing relevant and timely information. Evaluators must understand the audience, address their information needs, and communicate findings clearly. *Operational Minimum Standards:* • An evaluation takes place before important decisions/changes. Decision-makers are open to evidence. • A reasonable number of clear evaluation questions are formulated. • Recommendations are actionable, addressed and prioritised. • Relevant stakeholders participate in the evaluation process.
Feasibility	Evaluations should be practical and adapted to existing conditions, using resources efficiently. They should maximise stakeholder acceptance while only using resources - time, material, staff, and budget - to the extent necessary. *Operational Minimum Standards:* • Professional as well as methodologically, thematically and contextually qualified evaluators are recruited. • Report and management response are published. • Agreed measures are acted upon and tracked.
Correctness[5] (and Independence)	Evaluations must be conducted ethically and legally, respecting the rights and interests of those involved, requiring sensitivity by evaluators. Evaluators must maintain impartiality and independence.[6] *Operational Minimum Standards:* • Evaluators are independent, impartial and unbiased. • Evaluators are granted access to documents and stakeholders. • There are no conflicts of interest, undue influence and pressure. • Evaluations should embrace and harness diversity (in the team, approach, data collection). • The evaluation processes consistently adhere to ethical and legal standards at every stage.
Accuracy	Evaluations should provide valid, reliable and relevant information using robust methods. They should ensure findings are logically connected to the data collected. *Operational Minimum Standards:* • A sound evaluation approach, design and methods are employed. • Data collection and analyses use triangulation, are verifiable, transparent and fair. • Intended and unintended results are considered.

[3] SEVAL (2016). Evaluation Standards of the Swiss Evaluation Society. https://www.seval.ch/en/standards-competences/standards/ (retrieved: 07.10.2024).
[4] SDC, SECO-WE, PHRD (2025). Swiss International Cooperation: Evaluation Minimum Standards.
[5] The SEVAL Evaluation Standards use the term "propriety".
[6] Independence of SECO-WE's evaluation system is anchored in chapter 4 of the Evaluation Policy.

Based on these principles, SEVAL elaborates 27 standards, which can be found in the SEVAL Evaluation Standards.

1.4 Criteria

SECO-WE uses the OECD DAC[7] evaluation criteria to structure evaluations. The last column notes possible considerations if an evaluation criterion should be applied.

Evaluation criteria	*Definition*	*Consideration if criteria is relevant*
Relevance: is the project doing the right things?	The extent to which the project objectives and design respond to beneficiaries', global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change.	Did the project take place in a complex and sensitive environment?
Coherence: how well does the project fit?	The compatibility of the project with other projects in a country, sector or institution.	Did the project take place in a crowded institutional space? Is there a potential that other projects support or undermine the project under evaluation? Is the project testing a relatively novel approach?
Effectiveness: is the project achieving its objectives?	The extent to which the project achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups.	Should the evaluation measure the delivery of the project's objectives? is it feasible to evaluate the causal relationship between project and results?
Efficiency: how well are resources being used?	The extent to which the project delivers, or is likely to deliver, results in an economic and timely way.	Is it valuable to establish how the project's funding was utilised?
Impact: what difference does the project make?	The extent to which the project has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.	Is it worthwhile establishing the project's intended and unintended higher-level effects? Is it possible to evaluate the causal relationship between project and the effects?
Sustainability: will the benefits last?	The extent to which the net benefits of the project continue or are likely to continue.	Is it possible to establish if the effects of the project will continue?

Not all criteria need to be applied to every evaluation. They should be chosen based on the purpose and objectives of the evaluation to ensure it is useful and of high quality.[8]

[7] OECD (2019). Better Criteria for Better Evaluation. https://www.oecd.org/en/publications/better-criteria-for-better-evaluation_15a9c26b-en.html (retrieved 08.10.2024).
[8] OECD (2021). Applying Evaluation Criteria Thoughtfully. https://www.oecd.org/en/publications/applying-evaluation-criteria-thoughtfully_543e84ed-en.html (retriever 08.10.2024).

Box 3: Biennial Performance Report

WEQA carries out a synthesis and meta-evaluation every two years. As part of this, WEQA's backstopping team reviews all external project evaluations. The backstopping team assesses evaluation quality and rates the evaluated projects with respect to the applicable DAC criteria. It is not seen as unfavourably if only some of the DAC criteria were used. In fact, focusing on a few criteria can reflect a deliberate choice to evaluate those in greater depth.

Box 4: OECD DAC Criteria Rating

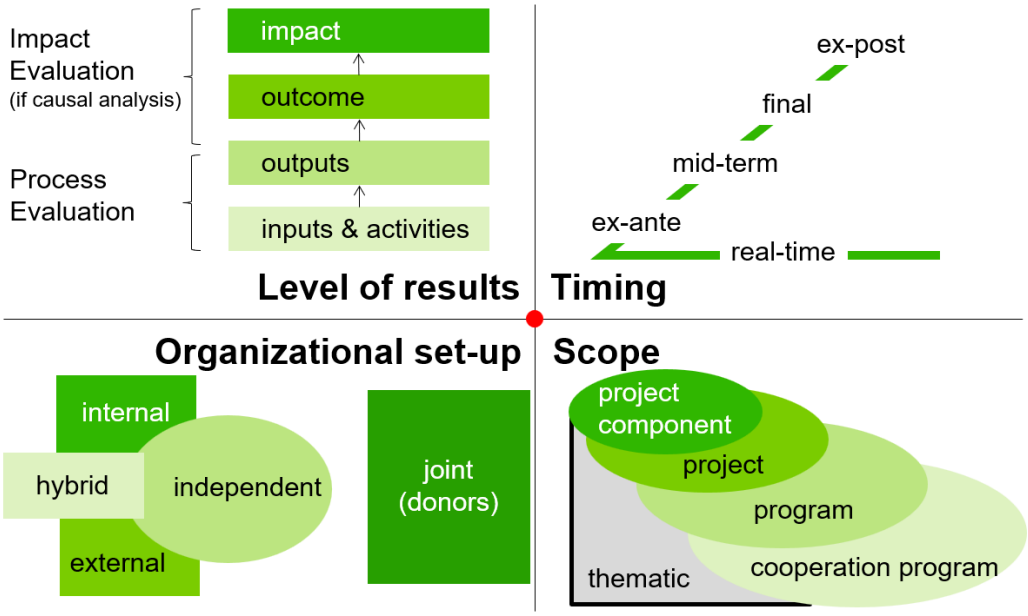
External evaluators no longer need to complete and submit the SDC-SECO Assessment Grid for DAC criteria. WEQA will arrange the rating based on the evaluation reports (refer to Box 3).

2 Evaluation Typology

2.1 Overview

Evaluations can take various forms depending on their *Purpose*, *Timing*, *Organisational set-up* and *Level of results*.

Figure 2: Evaluation Typology (SDC 2024[9]).



Further explanations and definitions can be found in Annex III, which complements this chapter.

The typology provides a simple way to understand and organise different evaluation purposes, objectives, approaches, and designs. However, in reality, evaluation is more dynamic, complex and nuanced. It's a constantly evolving field, with new approaches emerging all the time.

2.2 Mid-term, End-term and Ex-post Evaluations

The timing of an evaluation affects the results that can be assessed, the DAC criteria to be considered, and the questions that can realistically be answered. Different evaluation timings offer unique advantages and challenges. Below is an overview of the three key types of evaluations based on timing.

[9] Adapted from: SDC (2024). PCM Training. Module: Evaluation of Projects/Programmes.

Table 1: Linking DAC criteria with evaluation timing (own illustration)

	Evaluation timing		
DAC criteria	Mid-term	End-term	Ex-post
Relevance	✓	✓	-
Coherence	✓	✓	-
Effectiveness	+	✓	✓
Efficiency	✓	✓	-
Sustainability	+	+	✓
Impact	-	+	✓

Legend:
- : examination difficult or impossible → not recommended
+ : examination possible based on forecast, likelihood → not optimal
✓ : examination possible based on actual situation → optimal

2.2.1 Mid-term Evaluations

When: Evaluation performed towards the middle of the period of implementation of the project.

Mid-term evaluations have the key advantage of allowing adjustments during project implementation (learning and steering).

Mid-term evaluations are more suited for assessing efficiency and processes. They help determine whether the implementation has been efficient and whether inputs have led to the expected activities and outputs. In dynamic contexts, they can also be useful for reassessing relevance or checking coherence, especially if many actors are involved or if internal policies have changed.

However, it's difficult to fully assess effectiveness and sustainability at this stage. Evaluators can estimate the likelihood of achieving results (effectiveness) and whether systems are in place to ensure benefits last beyond the project (sustainability), but these predictions always involve uncertainty. Assessing long-term impacts is not feasible at mid-term, as these effects take time to develop.

Box 5: Is a Mid-term Evaluation the right tool?

- ✓ The main purpose is learning or steering.
- ✓ Make course corrections during implementation to (further) improve.
- ✓ Assess DAC criteria of Efficiency, Relevance and/or Coherence.
- ✓ Focus on assessing how inputs translated into activities and outputs (process evaluation).

Box 6: Minimum standards Mid-term Evaluation

- ☐ Conduct a mid-term evaluation halfway through the project/programme, with at least one year completed and one year remaining.
- ☐ Take stock of what is already known. Then ask: What information is needed to guide future decisions? → evaluation questions
- ☐ Focus on the DAC efficiency criterion: What specific aspects of efficiency need to be addressed?
- ☐ Decide if there is enough information to assess the likelihood of effectiveness and sustainability. If not, leave these criteria out.
- ☐ Do not evaluate the DAC impact criterion.
- ☐ Consult WEQA for feedback on the Terms of Reference (ToR), and ideally on the inception and final reports as well.

2.2.2 End-term Evaluations

When: Evaluation performed towards the end of the period of implementation of the project.

Although end-term evaluations don't allow for changes to the current project, they can still provide valuable insights for planning or ongoing follow-up phases. A key benefit of end-term evaluations is that they offer a full assessment of effectiveness, which is usually central. Since the project is still ongoing, it is possible to have full access to implementers, partners and beneficiaries, which is a challenge for ex-post evaluations. For end-term evaluations it is crucial to determine how important it is to establish a causal relationship between the project and its outcomes (refer to Figure 3 below). While it may be possible to assess the likelihood of sustainability and impact when follow-up phases are planned, in some cases this assessment may be too uncertain to pursue.

Box 7: Is an End-term Evaluation the right tool?

- ✓ The main purpose is learning, steering and/or accountability.
- ✓ Provide insights for the next phase or gather evidence for organisational learning.
- ✓ Assess DAC criteria of Effectiveness, Efficiency, Relevance and/or Coherence.
- ✓ Focus on assessing outcome achievement.

Box 8: Minimum standards End-term Evaluations

- ☐ Conduct the end-term evaluation during the last months of project/programme implementation or within six months after completion.
- ☐ Take stock of what is already known. Then, consider: What information is needed to inform decisions for the next project phase? What does SECO-WE aim to learn? What do the various evaluation stakeholders need to know? → evaluation questions
- ☐ Determine the importance of assessing the causal relationship between the project and its outcomes (DAC effectiveness criterion) (refer to Figure 3).
- ☐ If opting for an impact evaluation: (a) consult the decision tree in Annex VIII, and (b) assess the feasibility of the evaluation (evaluability assessment).
- ☐ Decide if there is sufficient information to evaluate the DAC criteria for Impact and Sustainability. If not, it's better to exclude these criteria.
- ☐ Start planning early, ideally at the beginning of the project.
- ☐ Consult WEQA for feedback on the Terms of Reference (ToR), and ideally on the inception and final reports as well.

2.2.3 Ex-post Evaluations

When: Evaluation of a project after it has been completed.

At SECO-WE ex-post evaluations should be conducted around 3 years after conclusion of the project (component).

Ex-post evaluations are particularly useful for assessing impact and sustainability, making these two DAC criteria their focus. The impact criterion covers both the intended and unintended effects of the project. Assessing unintended effects is challenging and requires creative, open methods to capture outcomes that go beyond what is outlined in the project's logframe.[10]

The sustainability criterion needs to be clearly defined. A USAID (2021) discussion note[11] highlights the importance of specifying what sustainability means for each project. Which elements or aspects of the project should be sustainable? Annex II exemplifies how a SECO-WE ex-post evaluation defined sustainability. Broadly, sustainability can be divided into two categories, both often addressed in a single evaluation.

[10] For example, by using the framework presented in "Foreign Aid and Its Unintended Consequences" by Dirk-Jan Koch (2024) or by using open-ended M&E methods, such as Most-significant Change or Outcome Harvesting.
[11] USAID (2021). Discussion Note: Ex-Post Evaluations.

Table 2: Two categories of sustainability (USAID, 2021)

Sustainability of project intervention	Sustainability of project effects

• Is the project still ongoing? Does service delivery continue?
• Focus: Outputs and activities.
• SECO example questions:
  • How would you assess the operation of the municipal water supply utility (staff, condition, and maintenance of facilities)? (*KfW 2018: Ex-post Evaluation of the Pisco Water Supply Rehabilitation Project in Peru*)
  • Does the Supreme Audit Institution in Tajikistan perform as an independent, professional external audit body and is fully operational? (*ZHAW 2018: Ex-Post Evaluation of the Tajikistan Supreme Audit Institution Project*)
	• Are the effects lasting?
	• How did the project contribute to these lasting effects?
	• Focus: Outcomes and impacts.
	• Question from SECO examples:
	  • What is the incidence of water-borne diseases? (*KfW 2018: Ex-post Evaluation of the Pisco Water Supply Rehabilitation Project in Peru*)
	  • Did the project (strengthening the Supreme Audit Insitution) contribute positively on the efficiency and transparency in resource use as well as induce accountability? *(ZHAW 2018: Ex-Post Evaluation of the Tajikistan Supreme Audit Institution Project)*

Early planning for ex-post evaluations is essential, ideally starting at the beginning of the project. Ensuring data and interviewee availability as early as possible is crucial. It is advisable to include a clause in the contract with the implementing partner for this purpose.

For larger, complex ex-post evaluations, it is highly recommended to carry out an evaluability assessment early on to assess feasibility and prepare for the evaluation. The level of complexity of an ex-post evaluation will also depend on how crucial it is to establish a causal relationship between intervention and change (see Figure 3 below).

Box 9: Is an ex-post Evaluation the right tool?

- ✓ The main purpose is accountability and/or learning.
- ✓ Gather evidence for organisational learning.
- ✓ Assess DAC criteria of Sustainability and Impact.
- ✓ Focus on assessing impacts (and outcomes).

Box 10: Minimum standards Ex-post Evaluations

- ☐ Conduct ex-post evaluations about 3 years after the project ends, once all phases are complete.
- ☐ Focus on assessing DAC Sustainability and Impact criteria.
- ☐ Define what "sustainability" means for the project.
- ☐ Start planning early, ideally at the beginning of the project, and
- ☐ ensure data and interviewee availability early on (include this in the contract) and until the ex-post evaluation takes place.
- ☐ Determine the importance of assessing the causal relationship between the project and its impacts (DAC effectiveness criterion) (refer to Figure 3).
- ☐ If opting for an impact evaluation: (a) consult the decision tree in Annex VIII, and (b) assess the feasibility of the evaluation (evaluability assessment).
- ☐ Consult WEQA for feedback on the Terms of Reference (ToR), and ideally on the inception and final reports as well.

2.3 From Process to Impact Evaluations

Evaluations differ based on the level of results they focus on. At one end, process evaluations look at how inputs are turned into activities and outputs. At the other end, impact evaluations examine the causal relationship between a project's activities and its outcomes or impacts.

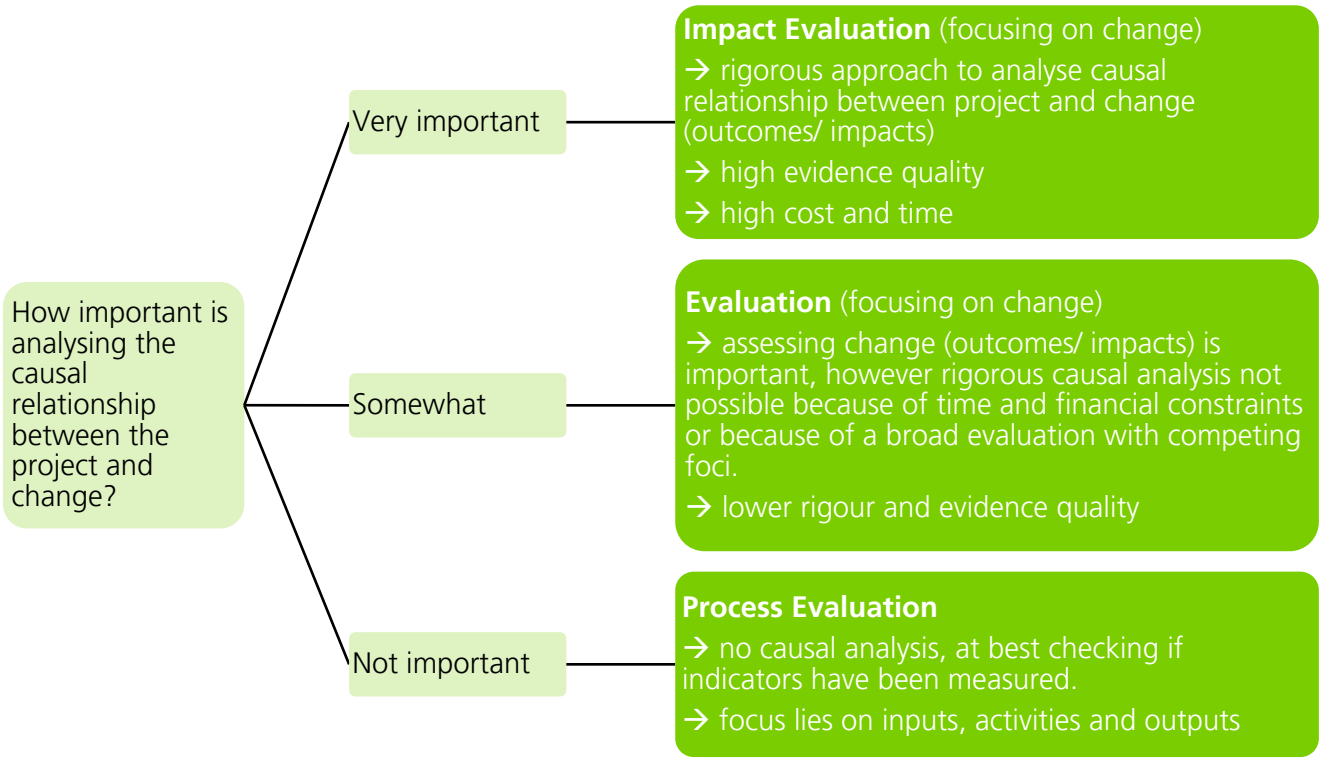
In practice, many evaluations incorporate elements of both types. Impact evaluations tend to be more costly (usually CHF 100'000 and more), time-consuming, and require a higher level of expertise due to the complexity of establishing clear causal relationships. However, in some cases, it may not be necessary to invest in rigorously proving these links, especially if the causal relationships are already evident or there are only a few well-known external factors influencing the outcomes. In such instances, a comprehensive impact evaluation might not be required, and the evaluation may lie somewhere between the two types (see Figure 3). In case an Impact evaluation is envisaged, consult the decision tree in Annex VIII.

Causality *is the relationship between cause (project) and effect (change, outcomes, impacts).*

Evaluations applying causal analysis are important *because they allow to determine whether the project was responsible for the observed change or whether it happened due to contextual or other factors. Causal analysis can also provide explanations of how and why a project contributed to change. Impact evaluations perform causal analysis.*

For example, an Impact Evaluation of the ***Social Impact Bonds*** *project in Colombia (phase I) provides causal evidence that short-term training and employment intermediation provided by the project increased the probability to get formal employment by 16% six months after training. The evaluation controlled for external factors like labour market conditions and socioeconomic characteristics.*

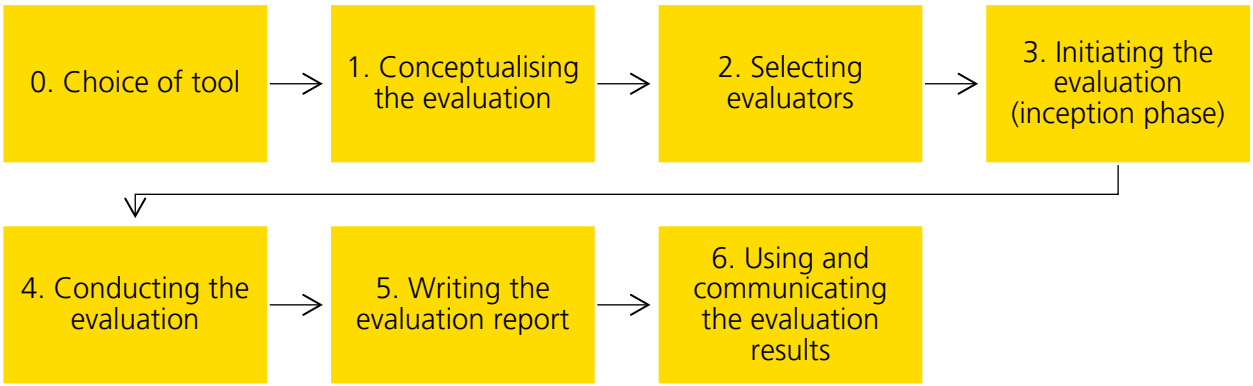
Figure 3: Decision tree on the importance of establishing causal relationships (own illustration).



3 Evaluation Process

The evaluation process can be structured in seven steps. The subsequent sections highlight each step.

Figure 4: Evaluation process flow chart



3.0 Step 0: Choice of analysis tool or process

Before starting, it is important to decide if an evaluation is the right tool for the purpose and goals. Numerous other options exist for learning, steering and accountability, such as audits, experience capitalisations, expert assessments, monitoring or third-party monitoring, political economy analyses, reports, and reviews (refer to Annex VI).

Evaluations answer focused evaluation questions using a sound research methodology and tap into multiple data sources. They can't be used to replace PCM processes like monitoring, risk management, or project appraisal. Consider whether the project can be properly evaluated. Using the evaluation decision tree (Annex VII) can help decide if an evaluation is the best choice.

➔ **Result of this step:** The right tool for the objective is selected.

Box 11: Support by WEQA

It is recommended to contact WEQA as early as possible when planning, designing, and implementing an evaluation. This should ideally begin when the project is approved. WEQA can help find the best solution for each case.

*At a minimum, **WEQA must be consulted on the Terms of Reference (ToR)**. A first meeting should take place before drafting the ToR to get advice on options and next steps. It is **recommended that WEQA reviews both the inception and final reports**.*

3.1 Step 1: Conceptualising the evaluation

Outline the key aspects of an evaluation in the Terms of Reference (ToR).

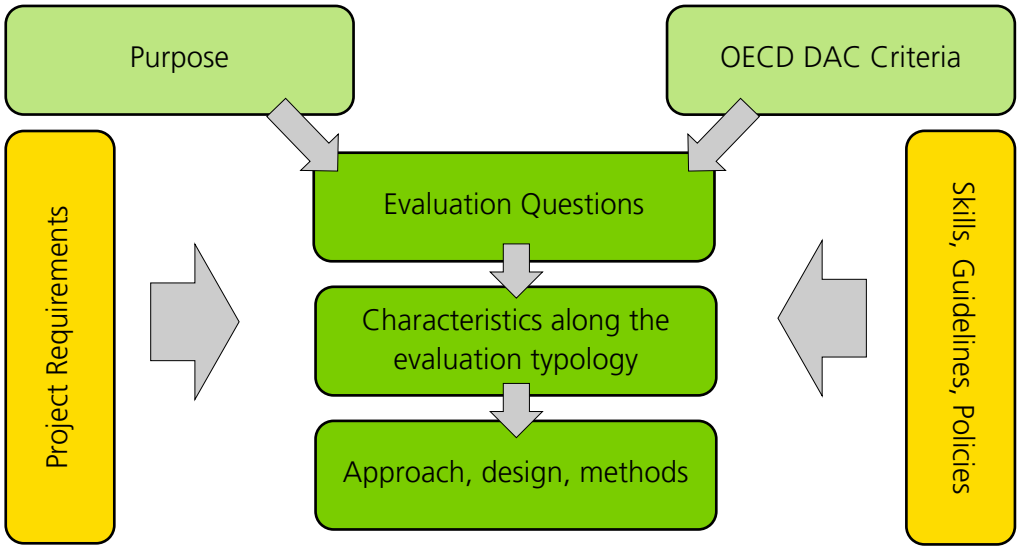
Table 3: Top 4 tips on conceptualising the evaluation

- ✓ *Dedicate ample time to thoroughly conceptualise the evaluation. Make use of the extensive project knowledge. This is the most time-intensive task for the commissioning body. If time is limited, it may be wiser to postpone.*
- ✓ *Involve relevant evaluation users in the conceptualisation process to ensure common ground and mutual understanding.*
- ✓ *Make sure the evaluation addresses a knowledge gap or learning need for the project and/or wider organisation, as this will enhance its usefulness. Steering and accountability needs should also be considered where relevant.*

✓ *Aim for focused and targeted evaluations. Broad evaluations often lack rigour and quality of evidence. Prioritise the most relevant DAC criteria and limit evaluation questions to a maximum of 10.*

The flowchart below provides an overview of the conceptualisation process. Once the purpose, objectives, and relevant DAC criteria are clearly defined, the next step is to formulate evaluation questions. These questions will shape the evaluation's characteristics, which in turn guide the selection of the most suitable approach, design, and methods.

Figure 5: Flowchart of conceptualising an evaluation (own illustration)



Reviewing the evaluation decision tree in Annex VII will offer further guidance on defining the key aspects of the evaluation.

Table 4: Tips on formulating evaluation questions

The evaluation questions should:[12] ✓ Directly link to the purpose of the evaluation. ✓ Avoid combining two questions in one. ✓ Avoid yes/no and leading questions. ✓ Be limited in number (maximum of 10) and focus on specific issues. ✓ Be feasible to answer at the project's current implementation stage. ✓ Be answerable using empirical evidence.	
Bad Practices: a) To what extent is the project effective in meeting its objectives? → too broad b) Do vocational training providers provide adequate support? → too vague c) Were the learnings we made useful? → leading and unclear	Good Practices: a) To what extent did the project increase youth employment (intended outcome) in targeted regions? Did employment outcomes vary by region or gender? → evaluates effectiveness and causal relationship b) Does the training and technical support for vocational training providers meet the needs of participants and government stakeholders? → evaluates relevance c) What broader lessons can our organisation draw from the past five years of implementing our approach to support bottom-up accountability mechanisms in country X? → learning capitalisation question

[12] USAID. Evaluation Questions Checklist.

If relevant, establish a steering committee with the partner organisation or select SECO-WE colleagues to join the evaluation team as peer members to enhance capitalisation. Relevant internal and external stakeholders must be informed and involved. Dissemination strategies throughout the evaluation process should be defined. They will ensure knowledge is transferred to other projects and the evaluation benefits the wider organisation.

➔ **Result of this step**: The evaluation is clearly outlined in the ToR, ensuring it is easy for both evaluators and stakeholders to understand.

3.2 Step 2: Selecting evaluators

Prepare an externally sharable version of the ToR and follow the required procurement procedures.

To assess offers, it is recommended to request a tentative evaluation matrix (refer to Annex IX). This matrix clarifies how the evaluation questions will be addressed, along with any limitations. It helps the commissioner assess whether the evaluator has fully understood the ToR and is realistic in the proposed approach.

It is essential that the evaluation team has a good balance of contextual (local), methodological and thematic skills. If there's uncertainty, prioritise methodological expertise to ensure the quality of the evaluation.

Some operational units and WEQA have ongoing backstopping mandates (or Project Support Partnerships PSP) through which evaluators can be selected.

Once evaluators have been selected elaborate a contract. The ToR must be integral part of the contract.

➔ **Result of this step**: Qualified and competent evaluators are selected. Contract signed.

3.3 Step 3: Initiating the evaluation / inception phase

Evaluators conduct an initial document review. During the inception phase, evaluators clarify scope, coverage, evaluation questions, approach, design, methods, field visits, and timeline, submitting these in the form of an inception report. As part of the inception report, evaluators should submit the (updated) evaluation matrix (refer to Annex IX). The inception report is discussed in a meeting[13] and finalised thereafter. Refer to Annex IV for a checklist to assess the quality of the inception report.

➔ **Result of this step**: A shared and documented understanding of the evaluation's purpose, scope, final evaluation questions, deliverables and workplan, as well as approach, design and methods is established in the final inception report.

3.4 Step 4: Conducting the evaluation

Evaluators review documents, gather data, conduct interviews and focus group discussions, visit project sites, and consult other relevant sources. They then perform preliminary data analysis and present their findings in a debriefing (or capitalisation) meeting, focusing on ensuring the findings are factually accurate and that stakeholders understand and can relate to them.

➔ **Result of this step**: Clear, well-documented findings are produced, shared, and understood, with feedback provided by the responsible section.

[13] To minimise the carbon footprint, improve stakeholder engagement across various locations, and optimise resource efficiency, virtual meetings are often the preferred option.

3.5 Step 5: Writing the evaluation report

Draft Evaluation Report

Using the findings and feedback from the debriefing meeting, the evaluators develop conclusions and, from those, recommendations. The draft evaluation report includes the context, evaluation questions, evaluation methodology (approach, design and methods), limitations, findings, conclusions, and recommendations. Evaluators must also submit the completed evaluation matrix (refer to Annex IX). Refer to Annex V for report length and structure and to Annex IV for a checklist to assess the quality of the evaluation report.

The draft report is then discussed in a meeting[14], with a focus on factual accuracy, clarity, and concise, actionable recommendations. Recommendations should indicate their priority and specify the entity responsible for their implementation. The report is not subject to negotiation, as it represents an independent perspective. Factual errors should be rectified. SECO-WE may agree or disagree, and it is acceptable to have differing views.

Final Evaluation Report

Based on the feedback received on the draft evaluation report, the evaluators prepare the final evaluation report.

Note: The Assessment Grid is not required as part of the report; it will be completed by the WEQA backstopping team.

➔ **Result of this step**: A factually accurate, clear, and easy-to-read report is produced, with actionable, addressed (indicating the responsibility for implementation), and prioritised recommendations.

3.6 Step 6: Using and communicating the evaluation results

Management Response

Following the final evaluation report, the responsible section prepares a Management Response. This response includes an assessment of the evaluation, outlines the position on each recommendation (agree, partially agree, disagree), and defines actions to address the recommendations, including deadlines and assigned responsibilities.

Dissemination

The evaluation report, along with the management response, will be published. The report must not contain confidential or privacy-sensitive information. It should be sent to WEQA (incl. management response) for inclusion in SECO-WE's evaluation inventory and publication on ARAMIS[15]. Additional dissemination activities, such as brown bag lunches, workshops, presentations, webinars, or mailing lists, are strongly encourage. They ensure using the full organisational learning potential of the evaluation.

Follow-up

The Programme Manager monitors the implementation of defined actions for as long as operational needs require. Progress can be tracked until a specified threshold, such as 80% completion of measures, is achieved.

➔ **Result of this step**: The evaluation results are published, accessible to relevant stakeholders, and used for steering, learning, and/or accountability. Defined measures are implemented.

[14] To minimise the carbon footprint, improve stakeholder engagement across various locations, and optimise resource efficiency, virtual meetings are often the preferred option.
[15] Evaluations submitted to WEQA in 2024 and beyond are published through ARAMIS.

4 Annexes

Annex I. Chronology of Versions

Version	Who	What
1.0 (ca. 2013)	Unknown, undated	First version.
2.0-draft (23-10-2024)	Malacarne, Nicola (WEQA)	Draft of second version elaborated.
2.0-draft (01-11-2024 - 31-01-2025)	Ferrini, Lia (WEHU); Hürzeler, Basil (WEQA); Kaenzig, Raoul (WEIN); Orjales, Carlos (WEMU); Schneider, Johannes (WEQA); Thönen, Reto (SDC); Walz, Janine (WEIF)	Consultation.
2.0 (01-02-2025)	Schneider, Johannes (WEQA)	Approval of version 2.

Annex II. Example of Sustainability Definition

Sustainability Definition from Sustainable Cacao Production Programme (SCPP) II used for the Ex-post Impact Evaluation (2025):

- Market-oriented supply chain actors (i.e., cocoa trading companies, certificate holders, farmers extension workers) in the Indonesian cocoa supply chain continue to provide training and support services to farmers while providing a price incentive, and premium incentives and market access for higher quality fermented beans.
- Farmers in Indonesia continue to grow cocoa in the face of a volatile world market price for cocoa beans and their tendency to switch to other crops.
- SCPP supported farmers continue to be part of traceable and sustainability certified supply chains.
- Lessons learned from SCPP have been transferred and applied to other relevant commodity sectors in Indonesia.
- Local government agricultural extension agents and agronomist use their knowledge from the SCPP training to continue assisting cocoa farmers.

Annex III. Extended Evaluation Typology

Unless stated otherwise, the definitions in double quotation marks ("…") are sourced from OECD DAC (2023)[16].

Objects of Analysis	**Project evaluation:** "Evaluation of an individual project designed to achieve specific objectives within specified resources and implementation schedules, often within the framework of a broader programme, examining its relevance, coherence, effectiveness, efficiency, impact [and/or] sustainability."
	Evaluation of a project component: Evaluation of a specific component of a project designed to achieve specific objectives within specified resources and implementation schedules. Impact evaluations often concentrate on individual project components due to their ability to deeply analyse and elucidate the causal mechanisms involved.
	Programme evaluation: "Evaluation of a set of projects, combined to attain specific global, regional, country, or sector development objectives."
	Thematic evaluation: "Evaluation of a selection of projects, all of which address a specific sustainable development priority or topic, that cuts across countries, regions, and sectors. Note: Often thematic evaluations will examine a strategic approach or priority topic across a variety of projects. An example would be evaluating the extent to which the rights of people with disabilities were advanced across a portfolio of projects in education, health and employment."
	Sector (programme) evaluation: "Evaluation of a cluster of projects within one country or across countries, all of which contribute to the achievement of a specific goal. Note: a sector includes development activities commonly grouped together for the purpose of public action such as health, education, agriculture, transport, etc."
	Evaluation synthesis: An evaluation synthesis aggregates findings, conclusions and recommendations from a series of evaluations (sometimes also referred to as meta-evaluation).
	Quality assessment: A quality assessment judges the quality of one or several evaluations and their adherence to standards and established good practice in evaluation (sometimes also referred to as meta-evaluation).
Timing	**Ex-ante evaluation:** "An evaluation that is performed before the implementation of a project."
	Real-time evaluation: "A process that provides immediate (independent) evaluative evidence, insights and feedback to inform decision-making, learning and implementation while the project is underway."
	Mid-term evaluation: "Evaluation performed towards the middle of the period of implementation of the project." It serves to make course corrections during implementation to (further) improve.
	End-term: Evaluation performed towards the end of the period of implementation of the project.[17]
	Ex-post evaluation: "Evaluation of a project after it has been completed." Ideally the evaluation takes place around 3 years after the project ends, once all phases are complete. A key focus should be put on assessing DAC Sustainability and Impact criteria.
Level of Result	**Process evaluation** assesses "internal dynamics of implementing organisations, their policy instruments, their service delivery mechanisms, their management practices, and the linkages among these." They analyse how inputs translate into activities and outputs.
	Impact Evaluation[18] assesses the casual relationship between project and change by using a research method rigorously. More specifically, an Impact Evaluation:

[16] OECD (2023). Glossary of Key Terms in Evaluation and Results-Based Management.

[17] SECO-WE definition.

[18] SECO-WE definition, inspired by Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). Broadening the range of designs and methods for impact evaluations.

	- assesses the contribution[19] of a project to change (at outcome or impact level). It explains and assesses the plausibility of the causal relationship between project and change; - AND / OR it measures attribution[20] of intended change (at outcome or impact level) to the project. It demonstrates the presence of a causal relationship between the project and change; - it does this by using a (scientific) research methodology rigorously. - The OECD DAC criteria of effectiveness and impact are usually prioritised.
Approach	**Theory-based evaluation:** Evaluations based on an explicit theory of change or logical framework. To test the theory of change, theory-based evaluations often use deductive and inductive reasoning. Deductive reasoning starts with the assumed theory of change and formulates hypotheses that must hold true if the theory is correct. It then tests these hypotheses against observations made and data collected. It allows to *reject* or *not reject* the assumed theory of change. Inductive reasoning aims at identifying patterns in observations and data. From these patterns it will establish general conclusions, where possible. It tests the assumed theory of change by testing for alternative theories of change that explain the observed results.
	Utilisation-focused evaluation: Utilisation-focused evaluations identify and involve its intended user(s) and strive to be useful to them. The reasoning is, that if intended user(s) are involved in the process, they will provide important inputs, will better understand the reasoning, will have more ownership and hence will be more likely to act on the findings, conclusions and recommendations.
	Developmental evaluation: "An iterative, embedded approach to evaluation, designed to support learning particularly in complex or uncertain environments. It involves providing real-time, or near real-time, (independent) feedback to project staff, thus facilitating ongoing learning and enabling improvements during implementation."
	Participatory evaluation: "An approach in which partners (including target groups) work together and are actively involved in the evaluation including designing plans, collecting and interpreting data, documenting and using findings, and formulating conclusions and recommendations."
Design	**Case-based design**[21]**:** "Evaluation designs that use cases (or case studies) as their primary empirical material". - Within-case analysis involves examining the causal processes within a single case. This analysis is necessary to provide a detailed understanding of the specific mechanisms and variables at play in that case. It helps in identifying how particular projects have contributed to observed outcomes, which can clarify patterns and provide deeper insights into causal relationships. - Cross-case: Cross-case analysis entails comparing multiple cases to identify patterns of convergence and divergence across different contexts. This analysis synthesises findings from individual case studies to draw broader conclusions about causal relationships. It allows to recognise consistent patterns and variations, aiding in the generalisation of findings.
	Experimental design: Before the project starts, individuals/organisations/companies are randomly assigned to a group that will benefit from the project (treatment group) and a second group that will not (control group). Both groups are surveyed. The survey can be repeated during the project but must be repeated after the project ends. The difference between the two groups allows for the determination of a causal link between project and change (attribution of the project to change).
	Quasi-experimental designs use statistical methods to infer from individuals that do not benefit from the project and from those that do the causal link between project and change (attribution of the project to change). Examples of such designs include difference-in-difference, regression-discontinuity, statistical matching, time-series, instrumental-variables, and panel analyses.
	Pre-post-test designs assess change by comparing situation at beginning of project (baseline) with the actual situation (mid-term or endline) without having a control or comparison group.

[19] "The role or part played by a project, together with other projects, in bringing about an observed (or expected) result. The way(s) a project helps to advance towards a goal."
[20] "The ascription of a causal link between observed (or expected to be observed) changes and a specific project."
[21] Definitions inspired by: Raimondo, Estelle (2023). The Rigor of Case-Based Causal Analysis. Busting Myths through a Demonstration.

	Ex-post-facto designs are implemented as a one-off study during or after a project. They compare and consider existing differences and their relationship to change. Ex-post-facto designs aim at identifying the contribution of a project to change without having a baseline.
Intended Use	**Formative evaluation:** "Evaluation intended to improve performance or to inform planning of a subsequent phase, often conducted during the implementation phase of the project. Note: Formative evaluations may also be conducted for other reasons such as compliance, legal requirements or as part of a larger evaluation initiative."
	Summative evaluation: "A study conducted at the end of a project (or a phase of that project) to determine the extent to which anticipated outcomes were produced. Summative evaluation is intended to provide information about the worth or significance of the project."
Method	**Data collection:** "Methods used to identify information sources and collect information." Examples include interviews, surveys, observations, focus group discussions, expert opinions, case studies, literature search and using secondary data. Data collection methods result in different types of data, such as socio-demographic data for project region, GIS data, national statistics (census) or administrative data.
	Sampling: "Sampling is a process that enables information to be collected from a small number of individuals or organisations within a project, and then used to draw conclusions about a wider population. There are many different sampling methods. Quantitative analysis tends to require large, random samples. Qualitative analysis usually relies more on smaller, purposefully chosen samples."[22] - Probability sampling: "[…] selecting a truly random and statistically representative sample that will permit confident generalisation from the sample to a larger population. The purpose is generalisation."[23] - Purposeful sampling: "[…] selecting information-rich cases for study in depth. Information-rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the research […]"[24]
	Data analysis: "The aim of data analysis is to help turn raw data into knowledge, which can then be used for decision-making and other purposes. […] There are many different types of data analysis. These include quantitative, qualitative and participatory analysis."[25] Examples include statistical analysis, rating and scales, thematic analysis, narrative analysis, etc.

[22] INTRAC (2017). Sampling.
[23] Patton, M. (1990). Qualitative evaluation and research methods.
[24] Ibid.
[25] INTRAC (2017). Data Analysis.

Annex IV. Evaluation Report Checklist

→ *To be used by Programme Managers to assess quality of offers, inception reports, or evaluation reports.*

Criteria	*Offer*	*Inception report*	*Final report*
Is the offer/report easy to read and understand, with only minimal errors?	x	x	x
Does the report adhere to the stipulated structure and length?		x	x
Is the report suitable for publication, and does it exclude confidential or privacy-sensitive information? (Confidential or privacy-sensitive information are to be submitted in a separate annex.)			x
Does the report include a set of annexes that provide additional detail beyond the main report (e.g., specific methodological aspects, case studies, etc.)?			x
Is the evaluation approach, design, and methodology clearly articulated, including the use of an evaluation matrix (refer to Annex IX)?	x	x	x
Are the limitations of the chosen approach/design/method thoroughly explained?		x	x
Is the sampling strategy clearly described and justified?		x	x
Is the data analysis procedure detailed?		x	x
Do the findings logically emerge from the collected and analysed data? Are the findings supported by evidence?		x	x
Does the evaluation transparently address how it treats, values, triangulates, and assesses different types of evidence? Does it assess the quality or strength of the individual pieces of evidence?		x	x
Do the conclusions follow logically from the findings?			x
Do the recommendations logically stem from the conclusions? Are they clearly articulated, targeted at the intended users, and action-oriented?			x

Annex V. Standard Format for Evaluation Reports

Evaluation reports are to be between 20- and 25-pages long (excluding title page, content page, acronyms and abbreviations, acknowledgements and annexes). They should align with the following structure:

- **Title page**
- **Content page**
- **Acronyms and abbreviations**
- **Acknowledgements**
- **Executive summary**
- **Introduction**

  Purpose and objectives of the evaluation

  Scope of the evaluation

  Evaluation questions

- **Description of the development intervention**

  Context of the intervention, including policy and institutional context

  Description of the intervention logic (Theory of Change or logframe) and the implementation arrangements

- **Methodology**

  Description of the evaluation process (e.g. stakeholder participation and consultation)

  Description of the evaluation approach (theory-based, participatory, developmental, utilisation-focused), design (case-based, experimental/quasi-experimental, statistical) and methods (sampling, data sources, data collection, data analysis), as well as their limitations.

- **Findings**

  Presentation and interpretation of the factual evidence in relation to the evaluation questions.

- **Conclusions**

  Assessment by the evaluators of the intervention results against the expected results (as identified at the planning stage or as reconstructed by the evaluators).

- **Lessons learned**

  Lessons that may have implication for the future of the development intervention or may be relevant for wider application.

- **Recommendations**

  Proposals for improvements for the client and users of the evaluation.

- **Annexes:**

  TOR

  List of stakeholders consulted

  Evaluation Matrix

Annex VI. Assessment tool and processes other than evaluation

An evaluation may not always be the right tool for your purposes. This non-exhaustive list provides examples of alternative assessment tools and processes.

Tool / process	*Description*	*Comparison with evaluation*
Audit	An independent, systematic, objective quality assurance assessment designed to document and improve the effectiveness of risk management, control and governance processes.[26]	An evaluation focuses on outputs, outcomes and impacts of projects, while audits focus on compliance as well as inputs and activities.
Experience capitalisation	Experience capitalisation refers to the transformation of knowledge into capital by those directly involved in a project to inform an institutional practice. It is a type of participatory learning process that is usually documented.[27]	An evaluation uses a research methodology and is based on multiple data sources (triangulation). An experience capitalisation is based on the combined experiences of practitioners and/or beneficiaries.
Expert assessment/ appraisal	Assessment or appraisal of a technical aspect of a project, usually conducted by a subject-matter expert.	An evaluation uses a research methodology and is based on multiple data sources (triangulation). An expert assessment is usually the informed opinion by one expert.
Monitoring or Third-Party Monitoring	A continuing process that involves the systematic collection or collation of data (on specified indicators or other types of information). Provides management and other stakeholders with indications of implementation progress, achievement of results, use of allocated funds and other intervention and context-related information.[28] If carried out by an external party, referred to as Third Party Monitoring.	An evaluation is more systematic and focuses on answering a set of questions using a research methodology. Evaluations usually use monitoring data as one source of information.
Political Economy Analysis	Political economy analysis (PEA) is a set of concepts, questions and tools helping to better understand the contexts in which projects operate. PEA helps to explain how change processes happen, why they can become blocked, and what development actors can effectively do to contribute to change.	A PEA is an evaluative analysis approach and has overlaps with evaluation. An evaluation usually puts more weight on the project and its achievements, while PEA emphasises contextual factors and how they influence a project.
Report	A structured document that provides a summary and appraisal, of activities, progress, outputs and outcomes, to a projects, programs, or policies. Reports are usually based on monitoring data.	While reports can be evaluative, they do not apply a research methodology.
Review	An assessment of the performance of a project, periodically or on an ad hoc basis.[29]	An evaluation is generally a more systematic and comprehensive assessment than a review. Reviews tend to emphasise operational aspects.[30]

[26] OECD (2023). Glossary of Key Terms in Evaluation and Results-Based Management.
[27] SDC (2009). Knowledge Management Toolkit.
[28] OECD (2023). Glossary of Key Terms in Evaluation and Results-Based Management.
[29] Ibid.
[30] Ibid.

Annex VII. General Evaluation Decision Table

Do you have any questions or doubts about the project?

↓

Do you have a question regarding:

Relevance, coherence, effectiveness, efficiency, sustainability, impact?	Monitoring system or monitoring data?	Compliance with project documents, contracts, etc.?	Project organisation, management and governance?	Thematic/technical issues?
↓	↓	↓	↓	↓
	Monitoring assessment. Third-party monitoring	Audit	Organisational or management review	Expertise/ backstopping by thematic/ technical expert

OECD DAC Criteria	**Timing**[31]	**Level**	**Approaches**	**Design**
Relevance	ex-ante, real-time, mid-term, end-term	Process evaluation, Impact evaluation	General approaches: - Theory-based - Utilisation-focused - Developmental - Participatory approach Causal inference approaches: - Counterfactual - Configurational - Generative	- Case-based (within-case and between-case) - Experimental - Quasi-experimental - Pre-post-test - Ex-post-facto - Statistical
Coherence	ex-ante, mid-term, end-term	Process evaluation		
Effectiveness	(ex-ante), real-time, (mid-term), end-term, ex-post	Process evaluation, Impact evaluation		
Efficiency	(ex-ante), real-time, mid-term, end-term	Process evaluation		
Sustainability	(mid-term), (end-term), ex-post	Process evaluation, Impact evaluation		
Impact	(end-term), ex-post	Impact evaluation		

[31] In brackets means an examination is only based on assumptions, forecast, likelihood.

Annex VIII. Impact Evaluation Decision Table[32]

Do you have any questions about the causal relationship between your project and change (at outcome or impact level)?

↓

Is your key evaluation question related to one of the following areas:

	To what extent can a specific (net) effect be attributed to the project?	**In which configuration did the project work best?**	**Why and how did the project make a difference?**	**What were the intended/ unintended, desirable/ undesirable consequences of the project?**
	↓	↓	↓	↓
	assessing attribution[33] through a counterfactual causal logic	assessing contribution[34] through a configurational causal logic	assessing contribution through a generative causal logic	assessing contribution through a participatory approach
	↓	↓	↓	↓
Requirements	- Can manipulate project - Possibility to quantify outcome/impact variable - Direct causal relationship - Sufficient numbers (beneficiaries, households etc.) for statistical analysis	- Comparable cases where a common set of causes and factors are present - Existing evidence for the relevance of these causes/ factors	- Theory of Change that allows identifying factors by the project and from the context leading to change	- Broad understanding of how the project and contextual factors lead to change
Approaches / Designs	- Experiments - Quasi-experiments - Statistical studies and econometrics, e.g. regression analysis - Hybrids together with case-based designs or participatory approaches	- Case-based designs to conduct cross-case analysis, e.g. Qualitative Comparative Analysis (QCA) - Statistical studies and econometrics, e.g. regression analysis	- Theory-based approaches: ▪ Contribution Analysis ▪ Process Tracing ▪ Realist evaluation - Participatory approaches	- Participatory approaches: ▪ Outcome Harvesting ▪ Most-significant Change

If your key evaluation question is NOT in one of these areas, an Impact Evaluation may not be suitable.

For a more elaborate choice of evaluation design and approach, work through Bond's (2016) Choosing Appropriate Evaluation Methods Tool.

---

[32] Inspired by: Stern, E. (2015). Impact Evaluation: A Guide for Commissioners and Managers. / Quadrant Conseil (2017). Impact Evaluation Approach Tree (retrieved 17.10.2024).

[33] "The ascription of a causal link between observed (or expected to be observed) changes and a specific project." According to OECD (2023). Glossary of Key Terms in Evaluation and Results-Based Management.

[34] "The role or part played by a project, together with other projects, in bringing about an observed (or expected) result. The way(s) a project helps to advance towards a goal." According to OECD (2023). Glossary of Key Terms in Evaluation and Results-Based Management.

Annex IX. Evaluation Matrix

→ *To be used by evaluator(s) as part of their offer/ proposal. Updated and re-submitted during inception and final reporting phases.*

→ *May be adapted according to the evaluation's / evaluator's needs and preferences. The essence of the tool – to establish a transparent link between evaluation questions, data collection/ analysis, limitations, and evidence quality – should remain visible.*

Name of Evaluation:	
Name of Evaluator(s):	

Evaluation Questions	Assessment Criteria, Indicators	Data Sources, Data Collection, Data Analysis	Limitations	Evidence Quality
- *What questions is the evaluation trying to answer?* - *Each evaluation question / sub-question should be specific, objective, neutral, measurable, and doable.* - *Each question / sub-question should be addressed in a separate row.* - *You can cluster your questions along the OECD DAC criteria, if it helps.*	- *Mention how the evaluation questions will be operationalised.* - *Which quantitative/qualitative indicators will be used?* - *Ensure key terms/ concepts from questions are defined.*	- *What information is needed to address each evaluation question? Where will you get them?* - *What data will you collect (primary data sources)? How will you collect the data?* o *Survey* o *Semi-structured interviews (e.g. Key Informant Interviews)* o *Focus group discussions* o *Observations* o *Sampling strategies: random sampling, purposeful sampling? How?* o *Sample size?* o *Which stakeholders will be represented in the sample? Which ones not?* o *etc.* - *Secondary data sources? (specify the sources)* o *Databases, Reports, Studies, etc.* - *Describe the data analysis techniques to be used.* o *Statistical analysis (e.g. regression analysis)* o *Descriptive analysis* o *Cost benefit analysis* o *Modelling* o *Mapping* o *Content analysis* o *Thematic analysis* o *Case studies* o *etc.* - *What controls will you apply to assure data quality during data collection and analysis?*	- *Is the question evaluable or are there concerns (evaluability)?* - *What are (potential) limitations and how will they affect the evaluation?* o *Data validity and/or reliability (e.g., due to sampling bias)* o *Data access* o *Security, confidentiality* o *Inability to generalise* o *etc.* - Be sure to address how these limitations will affect / have affected the evaluation.	- *What is the expected/ realised quality of evidence for the answers/ findings related to each evaluation question/ sub-question?* - *Use the following ratings; they may be adapted but should at least be based on two elements: (1) validity and reliability of data, and (2) triangulation, i.e. number of different data sources supporting the finding.* o *High: Finding is based on three or more valid and reliable data sources. (Or more than three data sources but only three considered sufficiently valid or reliable.)* o *Moderate: Finding is based on two valid and reliable data sources. (Or more but only two considered sufficiently valid or reliable.)* o *Low: Finding is based on only one data source. (Or on more but only one considered sufficiently valid or reliable.)*