



Government of the Netherlands



IOB evaluation quality criteria

> Introduction

- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > 14. Sampling strategy*
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Introduction

This document presents the latest version of the evaluation quality criteria IOB uses to monitor and assess the quality of evaluations and their Terms of Reference (ToRs). The criteria are organised around different stages of the evaluation process. Each criterion is explained and briefly illustrated with an example. The intended readers of this document are:

- Evaluators;
- M&E staff of partners responsible for commissioning, managing and assessing the quality of external evaluations;
- MFA Policymakers and NL Embassy staff responsible for commissioning and managing evaluations and for assessing the quality of evaluation reports of projects and programmes.

In principle, these evaluation quality criteria are applicable to any evaluation that is intended to assess to what extent an intervention has achieved or is achieving its intended objectives (ex-post evaluations and mid-term reviews). The criteria are less applicable to ex-ante evaluations, scenario analysis or feasibility studies. Although most examples presented here come from development cooperation that often uses clear theories of change (ToCs) at the start of an intervention, the criteria are also applicable in other areas of foreign policy, for example in diplomacy, if clear prior theories are lacking and so theory must be constructed at the time of evaluation.

- Introduction
- **Reading guide**
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Reading guide

The column on the left and the list on page 5 present a brief list of the 26 evaluation criteria, grouped according to the various stages of the evaluation process. Criteria can be relevant in more than one stage, but for the sake of readability we have grouped them as follows:

Criteria 1-9

These criteria can help in formulating ToR for the evaluation. They refer to the quality control of the evaluation, the description and background of the project or programme, the evaluation's objective and scope and the evaluation questions. The evaluator should critically reflect on the ToC presented in the ToR (criteria 4 and 5), and may reconstruct this in the inception report. The evaluation's ToR should have a preliminary methodology, that can be assessed using criteria 10-17 below.

Criteria 10-17

These criteria can be used for judging a technical proposal and for reviewing and commenting on an inception report. They focus on the methodological quality. The commissioner may request prospective evaluators to propose their methodology in their tender. The appointed evaluator is then expected to develop the full methodology in the inception report and, if necessary, the evaluation's proposed methodology can then be adjusted.

Criteria 18-24

These criteria can be used to assess the quality of the draft report under the assumption that the ToR and inception report have already been assessed. The criteria focus on the quality of the methodology used and on the conclusions. At this stage, it is often no longer possible to adjust the implementation of the evaluation, but it remains possible either to augment the analysis or to reformulate conclusions in light of certain methodological limitations.

Criteria 25 and 26: these criteria can be used for assessing the final evaluation report, under the assumption that the draft report has already been assessed.

Criteria 1-26

For assessing the overall quality of the final evaluation report and the evaluation process.

Section 4 explains each evaluation criterion and elaborates on the specific conditions that should be met for the criterion to be scored as *good*, *adequate* or *inadequate*. Examples are drawn from practice.

- Introduction
- **Reading guide**
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Reading guide

Many of the criteria are interrelated and sometimes one aspect of an evaluation may affect several criteria. For example, if the only source of information is the perception of directly involved stakeholders, the scores for criteria 14 and 17 will be negatively affected.

It is important to note that the criteria have been developed mainly for assessing individual project evaluations, not for broader policy evaluations.

When assessing the overall quality of the final evaluation report and the evaluation process, IOB recommends that *at least 23 of the 26* evaluation criteria are scored as 'adequate' or 'good'. In addition, there are 13 knock-out criteria. If an evaluation scores 'inadequate' on one of these 13 criteria, the evaluation as a whole should be regarded as inadequate. The knock-out criteria are: 2, 4, 5, 10, 11, 13, 14, 15, 17, 20, 21, 22 and 23. In this document, the knock-out criteria have been asterisked (*).

- Introduction
- Reading guide
- **The list of evaluation criteria**
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

The list of evaluation criteria

Quality control

- 1. Reference group
- 2. Independence of evaluators*

Description and background

- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*

Objective and delimitation

- 6. Evaluation's objective
- 7. Scope of the evaluation

Questions

- 8. OECD DAC evaluation criteria
- 9. Evaluation questions

Methodology

- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*

- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*

Results and conclusions

- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions

Usefulness and readability

- 25. Usefulness of the recommendations
- 26. Readability

- Introduction
- Reading guide
- The list of evaluation criteria
- **1. Reference group**
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Quality control of the evaluation

1. Reference group

This group generally comprises the commissioner of the evaluation, a representative from the implementing organisation evaluated and members with both thematic and evaluation experience, including at least one independent member. The role of the reference group is to assure evaluation quality and independence. It advises the commissioner on the ToR, the selection of evaluators, the elaborated methodology (inception report) and the draft evaluation report. Comments and advice from the reference group should be seriously considered by the evaluation team.

Good 	This criterion can be scored as good when: a. There is a reference group with at least one independent member; AND b. The reference group has been asked for advice during the different stages of the evaluation (draft ToR, inception report, draft final report).
Inadequate 	This criterion must be scored as inadequate when: a. There is no reference group; OR b. None of the members of the reference group are independent and external; OR c. The reference group has not been asked for advice in the different stages (ToR, inception report, draft final report).

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- **2. Independence of evaluators***
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Quality control of the evaluation

2. Independence of evaluators*

The evaluators and affiliated organisations must not have been involved in the design or implementation of the intervention (project, programme, policy) under evaluation and must have no interest in the evaluation's outcome.

Good



This criterion may be scored as adequate when:

- a. None of the evaluators have been involved in the design or implementation of the intervention; AND
- b. None of the evaluators is affiliated with one of the organisations active in the consortium responsible for the design or implementation of the intervention, or has been affiliated with one of the organisations during the implementation of the intervention; AND
- c. None of the evaluators has in the past worked for the Ministry of Foreign Affairs and has been responsible for formulating policy or setting up the programme that has led to the intervention under evaluation; BUT
- d. Programme staff have facilitated contact between external facilitators and beneficiaries and active stakeholders. Programme staff may accompany the external evaluators during field visits. Staff can help in making the necessary introductions, but must not have been present during interviews, nor should they have played an active role in sampling or case selection.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- **2. Independence of evaluators***
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Quality control of the evaluation

2. Independence of evaluators*

Inadequate 	This criterion must be scored as inadequate when: a. At least one of the evaluators has been involved in the design or implementation of the intervention; OR b. At least one of the evaluators is affiliated with one of the organisations active in the consortium responsible for the design or implementation of the intervention, or has been affiliated with one of the organisations during the implementation of the intervention; OR c. At least one of the evaluators has in the past worked for the Ministry of Foreign Affairs and has been responsible for formulating policy or setting up the programme that has led to the intervention under evaluation; OR d. The report does not mention whether or not evaluators were independent.
--	--

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > **3. Context of the intervention**
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > 14. Sampling strategy*
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Description and background of the intervention

3. Context of the intervention

This can include the national, sector and political context and explains the rationale of the intervention. Note that the benchmark for assessing progress (i.e. baseline data) need not be quantitative; it may be qualitative.

Good 	This criterion can be scored as good when: Detailed baseline data are presented on the project result indicators; AND there is a clear context and problem analysis.
Adequate 	This criterion may be scored as adequate when: - Baseline data are not presented, but very detailed problem and context analysis has been done, which forms the rationale for the intervention; OR - In the absence of baseline data, there is an explicit strategy to measure progress and to (re)construct the baseline situation.
Inadequate 	This criterion must be scored as inadequate when: There are no baseline figures on project result indicators; AND there is no problem analysis.

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > **4. Description of the intervention***
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > 14. Sampling strategy*
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Description and background of the intervention

4. Description of the intervention*

The ToR should preferably include a ToC; if not, they should include an intervention logic, policy strategy, or result chain. The evaluator may need to reconstruct a ToC, using whatever is available in project documentation but he or she must critically reflect on the ToC from the evaluator's point of view and include both this reflection and the ToC in the methodology section of the inception report. In some cases, the reconstruction of a theory is part of the evaluation itself.¹

Good 	This criterion can be scored as good when: - The specified ToC (or reconstructed ToC), intervention logic, or result chain is meticulously presented and takes account of intermediate steps between activity, output, outcome and impact; AND - Important external factors that can affect results at different levels in the result chain are discussed; AND - All assumptions about the relationships presented are mentioned, and their potential effects at different levels in the result chain are described in detail.
Adequate 	This criterion may be scored as adequate when: - The specified ToC (or reconstructed ToC), intervention logic or result chain is presented step by step and distinguishes between activities, outputs and outcomes. The result chain should make sense and should not omit important steps or factors; AND - The most important assumptions about the relationships presented are mentioned (e.g. between output and outcome and between outcome and impact).
Inadequate 	This criterion must be scored as inadequate when: The specified ToC (or reconstructed ToC), intervention logic or result chain is absent, or directly links the implemented activities to the results at outcome or impact level without distinguishing intermediate results and assumptions.

¹ The evaluator should be careful not to use the same set of observations for both the reconstruction of a theory and for validating this theory. The evaluator should separate these two phases in the evaluation or ensure there are several iterative phases.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- **4. Description of the intervention***
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Description and background of the intervention

4. Description of the intervention*



Example Good:

(a) *The Netherlands wishes to influence EU decision-making in favour of Dutch interests (impact). In order to do so, the Netherlands strengthens bilateral relationships and forms coalitions with other EU Member States (outcome), which requires improved policy dialogue with the government and societal actors to find common ground and generate joint action directed towards EU partners (outputs). The Dutch embassies in EU Member States are tasked with identifying priority themes on which the Netherlands and the relevant Member State are, or can become, likeminded and could take action as coalition partners (activities). Consequently, the Dutch embassies promote dialogue with the Member State's governments and societal actors to find common ground and generate joint action directed towards EU partners (activities).*

(b) External factors include:

- The openness of other Member States about their interests and concerns;
- The willingness of other Member States to form a coalition with the Netherlands;
- The reputation of the Netherlands as perceived by EU Member States, e.g. based on earlier consideration by the Netherlands of the other Member States' interests;
- The perceived likelihood that the Netherlands can support other Member States' interests in other EU decisions.

(c) The main assumptions about the external factors are:

- Member States' perception of the Netherlands is such that they are willing to consider engaging with the Netherlands in coalitions;
- Member States in a coalition with the Netherlands are unlikely to renege on their commitment to the Netherlands after engaging in dialogue with other Member States that the Netherlands is not in dialogue with.

The main assumptions within the result chain are:

- The Netherlands has identified issues interesting to both the Netherlands and its potential coalition partners;
- Policy dialogue does indeed result in coalitions;
- The coalitions are strong enough to influence EU decision-making;
- The EU decisions do indeed impact on Dutch economic or other interests).

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- **4. Description of the intervention***
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Description and background of the intervention

4. Description of the intervention*



Example Adequate:

(a) The Netherlands wishes to influence EU decision-making in favour of Dutch interests (impact). In order to do so, the Netherlands strengthens bilateral relationships and forms coalitions with other EU Member States (outcome), which requires improved policy dialogue with the government and societal actors to find common ground and generate joint action directed towards EU partners (outputs). The Dutch embassies in EU Member States are tasked with identifying priority themes on which the Netherlands and the relevant Member State are, or can become, likeminded and could become coalition partners (activities). The embassies therefore promote dialogue with the Member State's governments and societal actors to find common ground and generate joint action directed towards EU partners (activities).

(b) The main assumptions within the result chain are:

- *The Netherlands has identified the right issues that are of interest to the Netherlands and to its potential coalition partners;*
- *Policy dialogue does indeed result in coalitions;*
- *The coalitions are strong enough to influence EU decision-making.*



Example Inadequate:

The Netherlands wishes to influence EU decision-making in favour of Dutch interests (outcomes). Therefore, the Netherlands initiates dialogue with other Member States to find common positions on EU decisions (activities).

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > **5. Validation of assumptions***
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > 14. Sampling strategy*
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Description and background of the intervention

5. Validation of assumptions*

If, prior to the evaluation, the commissioner of an evaluation reflects on the ToC (which may be a policy, or project strategy or result chain) and assesses the underpinning assumptions, the commissioner could be guided to relevant evaluation questions for the ToR.

Otherwise, when performing the evaluation, the evaluator should not take the project ToC for granted, but should validate the ToC assumptions; these may relate to cause–effect relationships within the result chains, to the context, or to broader world views on development. The evaluator uses the evaluation results and broader literature (or literature reviews) to reflect on the validity of the ToC and adjusts or reconstructs the ToC if necessary.

Good 	This criterion can be scored as good when: - ToC assumptions have been tested against high quality literature (preferably systematic reviews) before the evaluation; AND - There is a clear strategy for testing the underlying assumptions as part of the evaluation methodology; AND - In the final evaluation report: assumptions have been tested, using the evaluation results.
Adequate 	This criterion may be scored as adequate when: - There is a clear strategy for testing the underlying assumptions as part of the evaluation methodology; AND - In the final evaluation report: assumptions are tested, using the evaluation results and a review of existing evaluations and literature.
Inadequate 	This criterion must be scored as inadequate when: - No assumptions are mentioned; OR - No strategy for testing the assumptions has been presented; AND - The project logic or ToC has not undergone critical reflection.

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > **6. Evaluation's objective**
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > 14. Sampling strategy*
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Objective and delimitation of the evaluation

6. Evaluation's objective

What the evaluation results will be used for must be clearly stated. There may be several objectives and it helps to distinguish:

- a. A knowledge objective (knowing whether objectives have been reached, for accountability; knowing what works, knowing how it works, knowledge for learning); this can be translated into questions about knowledge, whose answers will yield conclusions.
- b. An action objective (recommending what to do); this can be translated into policy questions whose answers will yield recommendations. These objectives will also determine when the evaluation results are needed, e.g. for a subsequent programme phase or new policy.

Good 	This criterion can be scored as good when: The objective of the evaluation is clearly mentioned and it is clear who will use the evaluation results and for what purpose.
Adequate 	This criterion may be scored as adequate when: The objective of the evaluation is mentioned, but it remains unclear what the results of the evaluation will be used for (thus, there is no distinction between knowledge objective and action objective).
Inadequate 	This criterion must be scored as inadequate when: a. The purpose of the evaluation remains unclear; OR b. Only an action objective is mentioned; OR c. The implicit objective is only to demonstrate the effectiveness, rather than to investigate it.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- **6. Evaluation's objective**
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Objective and delimitation of the evaluation

6. Evaluation's objective



Example Good:

- Knowledge objective: *'The objective of this evaluation is to determine and explain the effectiveness of programme X after five years of implementation.'*
- Action objective: *'Insight into the effectiveness of policy programme X and the underlying reasons for this level of effectiveness should enable policy makers to decide whether or not it should be continued, and if it is continued, what improvements can be made.'*



Example Adequate:

- 'The objective of this evaluation is to determine the effectiveness of policy programme X after five years of implementation.'*



Examples Inadequate:

- Unclear: *'The goal of this evaluation is to ascertain the contributions of policy programme X in a changing environment.'* (Contribution to what? Which changes in environment?)
- Only action objective: *'The objective of this evaluation is to gain input for deciding on the continuation of policy programme X.'*

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- **7. Scope of the evaluation**
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Objective and delimitation of the evaluation

7. Scope of the evaluation

It must be clearly stated what part of the intervention, expenditure, period, or even what part of the ToC is of interest for this evaluation.

Good 	This criterion can be scored as good when: a. The evaluation period is clearly defined; AND b. The geographical focus of the evaluation is clear; AND c. If there are various parallel result chains and various result levels, it is clear which result chains are covered in the evaluation and up to what level (e.g. outcome or impact).
Inadequate 	This criterion must be scored as inadequate when: a. The evaluation period is not clearly defined; OR b. The evaluation's geographical focus remains unclear; OR c. If there are various parallel result chains and various result levels, and it remains unclear which result chains are covered in the evaluation and up to which level (e.g. outcome or impact).

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- **8. OECD DAC evaluation criteria**
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation questions

8. Choice of OECD DAC evaluation criteria

It may turn out that given the evaluation objectives and limitations, not all the evaluation criteria (relevance, effectiveness, efficiency, impact sustainability and coherence) are needed. This will be reflected in the evaluation questions. (See [OECD DAC evaluation criteria 2019](#)). In addition to the OECD evaluation criteria, there are other cross-cutting issues that the evaluation may want to evaluate: examples include gender, inclusiveness, poverty reduction and climate smartness. If these are included, they should be defined in the ToR and translated into appropriate evaluation questions.

Good 	This criterion can be scored as good when: A reasoned choice has led to the selection of OECD DAC evaluation criteria and cross-cutting topics that will be used in the evaluation questions.
Inadequate 	This criterion must be scored as inadequate when: - It is unclear which of the OECD DAC evaluation criteria and cross-cutting topics have been used in the evaluation and why; OR - It is clear that the OECD DAC evaluation criteria have not been used when formulating the evaluation questions; OR - The evaluation criteria and cross-cutting topics are mentioned but this does not result in an organised and structured set of evaluation questions. Often, the evaluation criteria and cross-cutting topics have been mechanically translated into a set of universal evaluation questions but have not been customised for the specific intervention or for the objective of the evaluation.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- **9. Evaluation questions**
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation questions

9. Evaluation questions

The evaluation questions should be clear and follow logically from the intervention under evaluation, the evaluation objective and scope, and the chosen evaluation criteria. Evaluation questions should not be too general or vague, nor should they be too numerous and too detailed, as then the evaluation risks losing focus. The evaluation questions should be realistically ambitious and take account the evaluation's limitations (time, travel, budget, available information).

Good 	This criterion can be scored as good when: - The evaluation questions have a clear focus and are organised in a set that follows logically from the evaluation's objective, scope and the chosen OECD DAC evaluation criteria; AND - There is little to no overlap in the questions and there are not too many evaluation questions.
Adequate 	This criterion may be scored as adequate when: - There is some focus in the evaluation questions. The questions largely follow logically from the evaluation's objective, scope and evaluation criteria; AND - There is some overlap between evaluation questions; OR - The evaluation questions go beyond the evaluation objective or scope.
Inadequate 	This criterion must be scored as inadequate when: - There is no clear focus in the evaluation questions. The questions do not logically follow from the evaluation's objective, scope and evaluation criteria; OR - The combined answers of all evaluation questions do not provide sufficient information to answer the main evaluation question; OR - The evaluation questions are not formulated in a testable manner.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- **9. Evaluation questions**
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation questions

9. Evaluation questions



Example of good research questions (OECD DAC criterion relevance):

*'To what extent were the activities aligned with the needs of the beneficiaries during the implementation of the project?
Has the project been able to respond to changing needs as a result of the COVID-19 pandemic?'*



Example of a good research question (OECD DAC criterion relevance):

'To what extent were the activities aligned with the needs of the beneficiaries?'



Example of an unfocused research question (OECD DAC criterion relevance):

'To what extent was the project relevant?'

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- **10. Research design***
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

Note that the objective is to assess the evaluation methodology, i.e. the intended methodology described in the inception report, and how it was carried out, as presented in the evaluation report. For example, it must be ascertained whether intentions mentioned in the methodology section or in an inception report have been achieved; this should be done by scrutinising the evaluation's results and the conclusions it draws.

10. Research design*

The research design is clearly elaborated and shows how the research results will contribute to answering the evaluation questions. The design may consist of several quantitative and / or qualitative methods. If more than one method is used, the quality assessment must look at the individual methods and also the combination of methods.

- a. Quantitative methods** include three main research designs: survey, time series and experiment / quasi-experiment (see explanation under criterion 11).
- b. Qualitative methods** are usually mainly based on elements of the Case Study approach and the Grounded Theory approach. Methods include many research designs, some of which are more suitable for evaluating effectiveness and are less susceptible to bias (see further explanation under criterion 11).

Further explanation and elaboration of more detailed methodological choices are needed for both methods, and are highlighted in the criteria 11-15.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- **10. Research design***
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

10. Research design*

Good 	This criterion can be scored as good when: a. The research design has been clearly elaborated by showing what methodology or combination of methodologies have been used including a clear explanation of why they have been chosen; AND b. The chosen methodologies are appropriate for answering the research question, and it is clearly explained how they do so.
Adequate 	This criterion may be scored as adequate when: a. The research design mentions the methodologies that are used and provides an adequate explanation; AND b. The chosen methods are appropriate for answering the research question, even though an explanation for why they do so is lacking.
Inadequate 	This criterion must be scored as inadequate when: a. The research design is not clearly elaborated. Consider the following elements: – Methodologies are mentioned but not/poorly explained; – Only data-gathering techniques are mentioned (i.e. interviews, questionnaires, observations); – Only information sources are mentioned (i.e. type of respondents, documents); AND/OR b. The chosen methodologies are inappropriate for answering the research question, and/or it is unclear how they contribute to answering the research questions.

For criteria 11-15, a distinction is made between qualitative and quantitative methods, acknowledging that an evaluation often uses several methods.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- **10. Research design***
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

10. Research design*



Example of a research question and appropriate proposed methodology:

'How effective has the microfinance programme been, and what explains the level of effectiveness?'

In order to research the effectiveness of a microfinance programme for increasing household income and identify underlying reasons for this level of effectiveness, this research employs a mixed methods design. To determine the influence of microfinance on household income an experimental design will be implemented. This design will consist of before and after measurements in a treatment group and similar control group in order to isolate the effect of the microfinance programme and look for other explanatory factors. To understand how the microfinance programme works, and uncover explanatory mechanisms for the link between microfinance and household income, a case study of the programme will be conducted, looking at various aspects of the programme's ToC. Together, both methods will provide a complete picture of the effectiveness and underlying reasons for this level of effectiveness.'



Example of a research question and adequate proposed methodology:

'How effective has the microfinance programme been, and what explains the level of effectiveness?'

In order to research the effectiveness of a microfinance programme for increasing household income and identify underlying reasons for this level of effectiveness, this research employs a mixed methods design. To determine the influence of microfinance on household income an experimental design will be implemented. In addition, a case study on the programme will be conducted, looking at various aspects of the programme's ToC.'

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- **10. Research design***
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

10. Research design*



Example of a research question and inadequate methodology:

'How effective has the microfinance programme been, and what explains the level of effectiveness?'

Lacking methodological explanation:

'In this research we will distribute a questionnaire amongst users of the microfinance programme to determine the level of effectiveness of the programme. In addition we will conduct expert interviews and focus group discussions on the programme.'

Examples of inappropriate design choices:

- A single measurement survey to determine effectiveness.
- A quantitative approach to uncover the mechanisms responsible for how the programme works.
- A case study to determine the level of effectiveness.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- **11. Methods for effectiveness***
 - Qualitative methods
 - Quantitative methods
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

11. Methods for evaluating effectiveness*

The methods are appropriate to evaluate effectiveness in terms of attribution and / or contribution (if effectiveness is an evaluation criterion/question).² A challenge in evaluating effectiveness is to establish a causal relationship between the intervention (a project) and observed changes (outcomes). There will be other factors influencing the observed changes, so what part of these changes can be attributed to the intervention, or what was the contribution of the intervention to the observed changes?

- a. Qualitative methods** can substantiate a claim about the effect that the project has contributed to. All qualitative evaluation methods that can do so (i) formulate the cause–effect contribution question; (ii) reconstruct an intervention theory; (iii) formulate an alternative theory; (iv) collect data for both the intervention theory and alternative theory; (v) validate the theories step by step.

A good overview of qualitative evaluation methods is provided by [White and Phillips \(2012\)](#). They made an inventory of eight evaluation methods and distinguish four that are suitable for substantiating claims of effectiveness:

1. Realistic Evaluation;
2. Contribution Analysis;
3. Process Tracing;
4. General Elimination Methodology.

Using these four methods, White and Phillips created a general framework for qualitative evaluation. They also identified four qualitative evaluation methods that are less suitable for substantiating claims of effectiveness:

5. Most Significant Change;
6. Success Case Method;
7. Outcome Mapping;
8. Method for Impact Assessment of Programmes and Projects.

² This corresponds with internal validity: To what extent is there a causal relationship between, for example, outputs and outcomes? ([Vaessen et al., 2020, pp.11-12](#))

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- **11. Methods for effectiveness***
 - Qualitative methods
 - Quantitative methods
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

11. Methods for evaluating effectiveness*

More recently, Outcome Harvesting has gained popularity amongst practitioners and evaluators as a qualitative tool for monitoring and evaluation. IOB explicitly advises against using Outcome Harvesting as an independent, external evaluation method, because in practice this method is not appropriate to evaluate effectiveness and cannot validly establish the contribution of interventions to observed outcomes. In addition, the method is not in the spirit of several other evaluation quality criteria, specifically regarding the independence of evaluators (criterion 2), sufficient independent sources (criterion 17), triangulation (criterion 18) and avoidance of bias (criterion 19).

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- **11. Methods for effectiveness***
 - **Qualitative methods**
 - Quantitative methods
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

11. Methods for evaluating effectiveness*

Qualitative methods

Good 	This criterion can be scored as good when: a. The evaluator has first formulated a causal chain hypothesis before collecting data; AND b. The evaluator identifies possible other factors affecting the results (alternative hypotheses), before collecting data; AND c. The evaluator validates the causal chain step by step and also checks the effects of other factors.
Adequate 	This criterion may be scored as adequate when: a. The evaluator has first formulated a causal chain (or used an existing causal chain) as a hypothesis; AND b. To validate this causal chain the evaluator has then collected data; AND c. The evaluator has considered other factors (during and after the evaluator's data collection) that may have affected the results.
Inadequate 	This criterion must be scored as inadequate when: a. The causal chain between the intervention and the results is created after the evaluator's fact-finding (no separation between hypothesis and testing); OR b. Other factors that might have influenced the results were not considered; OR c. Results at outcome level are attributed directly to activities without a step-by-step validation along the causal chain.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- **11. Methods for effectiveness***
 - **Qualitative methods**
 - Quantitative methods
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

11. Methods for evaluating effectiveness*



Example Good:

The Netherlands applies a 'multi-bi approach' to influence EU decision-making and promote Dutch interests (see criterion 4).

- (a) *First, the evaluator formulated a hypothetical causal chain: identifying eligible priority issues by embassies and investing in dialogue with other EU Member States results in more policy dialogue (output), which improves mutual understanding of positions, increases like-mindedness and even increases formation of coalitions on priority issues (outcome). At impact level, coalition formation results in influencing EU decision-making successfully in terms of Dutch preferences. To achieve impact, new additional activities are required (negotiations with non-coalition Member States, EU institutions etc.).*
- (b) *In addition, before collecting data, the evaluator identified other factors to check whether:*
 - *Other Member States' policy positions were sufficiently close to the Dutch positions or were flexible enough for coalitions to be formed;*
 - *Dutch positions were flexible enough to join coalitions and the Netherlands was seen as an attractive coalition partner by other Member States;*
 - *Embassies have sufficient personnel and expertise to engage in dialogue and in the various Member States the selected priority issues are sufficiently coherent to build winning coalitions, or are too divergent;*
- (c) *The evaluator then collected data and conducted a step-by-step validation, showing that embassies had selected priority areas and intensified dialogue with partner governments and societal stakeholders. In a few cases, Member States were or became likeminded and became coalition partners. However, it was also found that positions and interests often differed too much to engage in coalition building, that Dutch positions were not flexible enough to make compromises in the process of coalition formation and that the coalitions that were formed were too weak to influence decision-making effectively. The evaluator was thus able to draw robust conclusions about the project logic and about other factors affecting the results.*

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- **11. Methods for effectiveness***
 - **Qualitative methods**
 - Quantitative methods
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

11. Methods for evaluating effectiveness*



Example Adequate:

- (a) *First, the evaluator formulated a hypothetical causal chain: identifying eligible priority issues by embassies and investing in dialogue with other EU Member States results in more policy dialogue (output), which improves mutual understanding of positions, increases like-mindedness and even results in coalitions being formed on priority issues (outcome). At impact level, coalition formation results in influencing EU decision-making successfully in terms of Dutch preferences. To achieve impact, new, additional activities are required (negotiations with non-coalition Member States, EU institutions, etc.).*
- (b) *The evaluator then followed a step-by-step validation, showing that embassies had selected priority areas and intensified dialogue with partner governments and societal stakeholders. In a few cases, Member States were or became likeminded and became coalition partners.*
- (c) *The evaluator checked possible other factors, and found that positions and interests often differed too much to engage in coalition building, and that the coalitions that were formed were too weak to influence decision-making effectively.
The evaluator thus tested the intervention logic while keeping an eye out for possible other factors.*



Example Inadequate:

- (a) *The evaluator did not base their evaluation on an intervention logic but instead asked diplomats from the Netherlands and other Member States why coalitions had or had not been formed. Diplomats attributed the formation of coalitions to the existence of like-mindedness. The position adopted by the Netherlands was sometimes perceived to be a minority position, and sometimes it was perceived to be a majority position.*
- (b) *The evaluator did not follow a project logic and did not consider factors that were not mentioned by stakeholders.*

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- **11. Methods for effectiveness***
 - Qualitative methods
 - **Quantitative methods**
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

11. Methods for evaluating effectiveness*

b. Quantitative methods can robustly substantiate a claim about the effect attributable to the project. The Maryland scientific methods scale distinguishes 5 levels:

1. One observation moment, after the project: comparison with and without project.
2. Two observation moments: comparisons before and after the project, without a control group.
3. Two observation moments: comparing before and after AND with and without project (double difference).
4. Two observation moments: comparing before and after AND with and without project (double difference, semi-experimental design), and correcting for other, external influences.
5. Two observation moments: comparing before and after AND with and without project (double difference); the participants are randomly assigned to a project: (randomised control group, experimental design).

Level 5 is best suited for attributing results to a project, but it is rare and not always possible to apply in evaluations. Level 4 is a commonly used good quantitative method. Levels 1 and 2 are generally not the preferred methods for making effect claims, and evaluators should be encouraged to aim at least for level 3 and preferably for level 4. Under certain strict conditions, evaluations below level 4 can be seen as just good enough, although in practice this is rare. Whether level 1, 2 or 3 is adequate depends on the evaluation subject and context, especially on whether the following assumptions hold true: (i) that without the project nothing would change over time, and (ii) that the control group is similar to the project group before the start of the project.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- **11. Methods for effectiveness***
 - Qualitative methods
 - **Quantitative methods**
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

11. Methods for evaluating effectiveness*

Quantitative methods

Good 	This criterion can be scored as good when: a. Level 4. The design considers the change that may happen without the project and the differences between participants and non-participants. b. Level 5. Experimental design (randomised control group): project participants are randomly assigned to a project intervention and to a control group.
Adequate 	This criterion may be scored as adequate when: a. Level 1 + two validated assumptions. One observation moment, comparing with and without project, assuming that the control group was similar when the project started and that nothing would have changed without the project. The evaluator has convincingly validated these assumptions and demonstrated the evidence; OR b. Level 2 + one validated assumption. The study design compares the project group before and after the project, without a control group, and assumes that without the project nothing would have changed. The evaluator has convincingly validated the assumption and demonstrated the evidence; OR c. Level 3 + one validated assumption. The study design compares a project group with a control group before and after the project, assuming that these two groups are similar. The evaluator has actively validated the assumption.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- **11. Methods for effectiveness***
 - Qualitative methods
 - **Quantitative methods**
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

11. Methods for evaluating effectiveness*

Inadequate



This criterion must be scored as inadequate when:

- a. Level 1 + one invalidated assumption. One observation moment, comparing with and without project, assuming that the control group was similar when the project started and that nothing would have changed without the project. The evaluator has not validated these assumptions; OR
- b. Level 2 + invalidated assumption. The study design compares the project group before and after the project, without a control group, and assumes that without the project nothing would have changed. The evaluator has not validated these assumptions; OR
- c. Level 3 + invalidated assumption. The study design compares a project group with a control group, assuming that these two groups are similar.



Example Good:

(a) A rural electrification project that aimed (among other things) to increase the income generated at the household level. To assess this, the evaluators combined a before–after comparison (comparing baseline with endline) and a with–without comparison (intervention village versus control village), assuring that these two groups are comparable (using covariates or matching techniques).



Example Adequate:

(a) A rural electrification project that aimed (among other things) to increase the amount of fresh food stored in household fridges. The evaluation must validate the following two assumptions: (1) there was no electricity before the project and that nobody else would have introduced electricity in the area and (2) the two groups are comparable, e.g. in terms of the general economic situation that might contribute to the amount of food people have stored.

If both assumptions are validated then not including a baseline may be seen as adequate.

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > **11. Methods for effectiveness***
 - > Qualitative methods
 - > **Quantitative methods**
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > 14. Sampling strategy*
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Evaluation methodology

11. Methods for evaluating effectiveness*



- (b) *A rural electrification project aimed, among other things, to increase the number of hours children do their homework. The hypothesis is that provision of electricity has resulted in children studying after dark. If the evaluation can validate and demonstrate that there were no other factors (e.g. money-earning activities) that affected the amount of time spent on homework, then not having a control group may be seen as adequate.*
- (c) *A rural electrification project in which the selection of participants is not affected by characteristics of the participants, but e.g. by geographical location. This may happen when a project rolls out its activities over different districts over the years. In this case, it may be assumed that the project participants do no better or worse than the control group (of participants that have not yet been included in the project).*



Example Inadequate:

- (a) *A subsidised farm equipment project in which farmer participants obtained higher crop yields than non-participant farmers. It is unclear whether the assumptions are valid because even before the project the participating farmers may have had higher crop yields, or they were more motivated and active and would have obtained higher crops even without the project.*
- (b) *A subsidised farm equipment project. Farm production has increased by 30%, and the evaluator attributes this to the project by comparing baseline data at the start of the project with endline data 3 years later. The assumption that production could not have changed in 3 years without the project cannot be validated. In fact, upon inspection it appears that production outside the project area has increased as well. Therefore, the changes in the project group area cannot be attributed to the intervention.*
- (c) *A subsidised farm equipment project. Baseline and endline data has been gathered from farmers for both the intervention group and the control group. However, it seems likely that farmers participating in the project are younger and more entrepreneurial than those who don't participate, partly because the project was selective (farmers had to pay some cash up front), and partly because risk-adverse farmers were not keen to join the project. The assumption that participants are similar to non-participants cannot therefore be validated, and the measured differences between the groups over time cannot be attributed to the project.*

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- **12. Methods for efficiency**
 - Qualitative methods
 - Quantitative methods
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

12. Methods for evaluating efficiency

The methods are appropriate to evaluate efficiency (if efficiency is an evaluation criterion/question). The evaluation needs to specify what aspect of efficiency is considered.³

a. Qualitative methods: e.g. organisational efficiency, assessment of demonstration of leverage effects and scaling.

b. Quantitative methods: e.g. calculation of cost effectiveness, timeliness of implementation, overhead costs.

³ The OECD DAC evaluation criteria for efficiency consider only the aspects (i) cost effectiveness and (ii) operational efficiency, but there are more aspects of efficiency.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- **12. Methods for efficiency**
 - **Qualitative methods**
 - Quantitative methods
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

12. Methods for evaluating efficiency

Qualitative methods

Good 	This criterion can be scored as good when: - The description of the methodology starts with an explanation of how efficiency is considered and ascertained, and this approach is appropriate for the evaluation questions; AND - The evaluation has applied this methodology correctly.
Adequate 	This criterion may be scored as adequate when: In the methodology section it is explained how efficiency is considered and ascertained, but the evaluator explains that because of a lack of data, the methodology was inappropriate for addressing the evaluation question in the ToR.
Inadequate 	This criterion must be scored as inadequate when: Although the ToR has a question on efficiency without further specification, when describing the methodology the evaluator does not specify how efficiency was ascertained. The results section on efficiency summarises ad hoc observations that indicate organisational efficiencies or inefficiencies.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- **12. Methods for efficiency**
 - **Qualitative methods**
 - Quantitative methods
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

12. Methods for evaluating efficiency



Example Good:

- (a) *In the evaluation of the Dutch multi-bi approach to influence EU decision-making, the ToR defined efficiency in two ways / two questions: (i) does the relative investment by Dutch diplomacy match the relative importance of the EU decision, given the probability of influencing that decision? (ii) were Dutch activities timely to influence EU decision-making?*
- (b) *The evaluator compared several cases – some successful, some unsuccessful – of Dutch influencing of EU decisions. The evaluator found that some cases received the attention they deserved, given their importance and the chances of success, but the evaluator also found that certain cases received too much attention (staff time), given the very low chances of success. The evaluator found that the identification of important issues and potential coalition partners was timely, but that the dialogue for coalition forming and adopting joint positions was often too late: other EU Member States had already cemented their positions.*



Example Adequate:

Although the ToR suggested cases be compared, taking account of the probability that influencing would succeed, the evaluator could not find a way of collecting data on the chances of success. Instead, the evaluator considered only the second question on the timelines of Dutch activities and was transparent about the choices made.



Example Inadequate:

The evaluator looked at organisational processes, without clear questions or indicators, and afterwards interpreted and judged that organisational efficiency was low due to high staff rotation, unfilled vacancies and poor handing over of tasks.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- **12. Methods for efficiency**
 - Qualitative methods
 - **Quantitative methods**
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

12. Methods for evaluating efficiency

Quantitative methods

Good 	This criterion can be scored as good when: - The description of the methodology explains how efficiency is considered and ascertained, and this approach is appropriate for the evaluation questions; AND - The evaluator has applied this methodology correctly.
Adequate 	This criterion may be scored as adequate when: - In the description of the methodology it is explained that lack of data made it impossible to evaluate efficiency in accordance with the question posed in the ToR, so an acceptable alternative definition and method was proposed, using the available data; AND - The evaluation has applied this alternative method correctly.
Inadequate 	This criterion must be scored as inadequate when: The ToR request efficiency be evaluated but without further specification. In the description of the methodology it is not specified how efficiency is considered either. The results section on efficiency summarises ad hoc observations that indicate organisational efficiencies or inefficiencies.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- **12. Methods for efficiency**
 - Qualitative methods
 - **Quantitative methods**
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

12. Methods for evaluating efficiency



Examples Good:

1. *Organisational efficiency (a) has been defined as timeliness (or delays) of project implementation. (b) The evaluation found that a three-year project had planned to train 2000 farmers each year, but training only started in the third year with 500 farmers, and an extension has been requested.*
2. *Efficiency of output delivery (a) has been defined as the unit costs of a training (10 half-day training sessions) per trainee. (b) A comparison with benchmarks from literature about similar training found that the project training was relatively efficient.*
3. *Cost effectiveness (a) has been defined as the monetary value of outcomes, compared with project costs. (b) The project invested USD 100 per farmer on training and a quality control system, resulting in each farmer gaining a net additional profit of USD 50 in each year of the project.*



Example Adequate:

The ToR stipulated that the cost effectiveness of a scheme to train farmers be evaluated, but evaluators were unable to collect data on the monetary value of the results, because the full results (the increased farm revenue) will not become apparent for some years. Therefore, the evaluators focused on comparing the unit costs (per trained farmer) with benchmark costs for farmer training.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- **13. Indicators or result areas***
 - Qualitative methods
 - Quantitative methods
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

13. Indicators or result areas*

The indicators or result areas are appropriate to capture the planned results across the different levels in the ToC.⁴

- a. Qualitative methods:** result areas and processes, including assumptions that are part of the ToC, are defined at and between different levels in the ToC (e.g. output, outcome, impact; context and other assumptions) and are appropriate for assessing the planned results. Generally, qualitative result areas are more open descriptions than SMART indicators used in quantitative research. When this is the case, the evaluator should still make them as specific as possible during the evaluation.
- b. Quantitative methods:** indicators are defined at different levels in the ToC (e.g. output, outcome, impact; context and other assumptions). Indicators should be SMART and appropriate for measuring the planned results.

⁴ This corresponds with construct validity: To what extent is the element that we have measured a good representation of the phenomenon we are interested in? (Vaessen et al., 2020, pp. 11-12)

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- **13. Indicators or result areas***
 - **Qualitative methods**
 - Quantitative methods
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

13. Indicators or result areas*

Qualitative methods

Good 	This criterion can be scored as good when: a. Result areas are well described by making abstract results more concrete, and they reflect the different result levels between interventions and final results in the ToC; AND b. On a case-by-case basis, the result areas of the overarching ToC have been further refined into measurable results of a local project-specific ToC, partly during project implementation, partly even during the evaluation itself, and often in discussion between evaluator and stakeholders.
Adequate 	This criterion may be scored as adequate when: Result areas describe at least three levels in the ToC: (1) intervention, (2) intermediate result and (3) project result of interest;
Inadequate 	This criterion must be scored as inadequate when: The result areas used reflect too large a step in the ToC;

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > **13. Indicators or result areas***
 - > **Qualitative methods**
 - > Quantitative methods
- > 14. Sampling strategy*
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Evaluation methodology

13. Indicators or result areas*



Example Good:

- (a) To evaluate the Dutch multi-bi approach to influence EU decision-making, the evaluator used a reconstructed programme ToC with the following result areas: EU decisions compared with the desired Dutch decision (impact); coalitions (number, weight) with Member States taking a joint position (outcome); Dutch policy dialogue with Member States (number / weight of Member States, level of mutual understanding – output); identification of priority topics around which coalitions could be formed (matrix of Member States, subjects, potential agreement – activities).
- (b) The ToR require the evaluator to analyse a number of carefully selected cases of Dutch policy influencing. For each case or lobby trajectory, the evaluator describes concrete result areas. A good description is given of the EU decision, and it is compared with the various scenarios of other decisions and the decision considered desirable by the Netherlands (impact). The different positions of all Member States are described and explained (external factors). The actual coalitions (with and without the Netherlands) are described and explained and are compared with the coalitions desired by the Netherlands. The policy dialogue with different Member States and other stakeholders is described, etc.



Example Adequate:

The evaluator defined three levels in the ToC: EU decision in line with Dutch interest (impact); coalitions with EU Member States and formulation of joint position (outcome); dialogue with other EU Member States and identification of joint interests (output).



Example Inadequate:

The evaluator defined two levels: the EU decision compared with Dutch interest (impact), and policy dialogue with other Member States (activity).

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- **13. Indicators or result areas***
 - Qualitative methods
 - **Quantitative methods**
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

13. Indicators or result areas*

Quantitative methods

Good 	This criterion can be scored as good when: a. Indicators have been linked to different (intermediate) levels in the ToC (or result chain) between intervention and results; AND b. Indicators are linked to other parts in the ToC, or to non-project factors that are likely to affect the project results; AND c. Indicators are SMART (the above examples are SMART): – Specific (see a and b); – Measureable: can be measured unambiguously; – Attainable: expected to change during the project intervention; – Relevant: say something about the intended results; – Time-bound: it is clear when the indicator should be measured and when results should be achieved.
Adequate 	This criterion may be scored as adequate when: a. Indicators are linked to at least 3 levels: project intervention, intermediate result and project result of main interest; AND b. The most relevant other factors outside the scope of the project are included if they are likely to affect the project results; AND c. Indicators are SMART, at least specific and measurable.
Inadequate 	This criterion must be scored as inadequate when: a. Indicators are linked to only two levels in the ToC (or result chain); OR b. No indicators are included of factors that are likely to affect project results; OR c. Indicators are not SMART.

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > **13. Indicators or result areas***
 - > Qualitative methods
 - > **Quantitative methods**
- > 14. Sampling strategy*
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Evaluation methodology

13. Indicators or result areas*



Example Good:

(a) *An SRHR advocacy programme (without service delivery) that intends to reach and organise youth and provide awareness campaigns.*

In the ToC one of the result chains is reflected by 8 indicators: number of youth reached through awareness programme; number of youth clubs formed; frequency at which these youth clubs meet; frequency of contact between members; number of SRHR information campaigns channelled through these youth clubs; adolescents' knowledge of SRHR; attitudes towards contraceptives; contraceptive use amongst adolescents in the preceding 12 months.

(b) *Percentage of youth indicating they can discuss SRHR topics with parents; % of youth indicating SRHR information is provided through schools; number of youth-friendly health providers in the area; access to contraceptives.*



Example Adequate:

(a) *An SRHR advocacy programme (without service delivery) that intends to reach and organise youth and provide awareness campaigns. Indicators: (1) number of youth clubs formed; (2) attitudes towards contraceptives; (3) contraceptive use.*



Example Inadequate:

(a) *An SRHR advocacy programme (without service delivery) that intends to reach and organise youth and provide awareness campaigns. In the ToC one of the result chains is reflected by only two indicators: (1) number of youth clubs formed and; (2) contraceptive use.*

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- **13. Indicators or result areas***
 - Qualitative methods
 - **Quantitative methods**
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

13. Indicators or result areas*

(b) There is no indicator for the access to contraceptives.

(c) Examples of non-SMART indicators: an inclusive and youth-friendly society, a more liberal culture, enhanced rule of law, a vibrant local economy, land degradation, etc.

Examples of indicators that are not appropriate to measure the intended result: measuring the number of youth that attended a training to make statements about strengthening gender equality; measuring the land area under a recommended practice to draw conclusions on soil degradation.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- **14. Sampling strategy***
 - Qualitative methods
 - Quantitative methods
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

14. Choice of sample, cases and information sources*

The choice of sample, cases and information sources (e.g. countries, projects, organisations and persons) is justified.⁵

The sampling strategy or case selection should minimise selection bias.

Selection bias: the selected sample in the study is not representative of the target population, but it correlates (positively or inversely) with project effectiveness and thereby undermines the generalisability of the findings. Self-selection of participants is an example of selection bias: successful and more entrepreneurial beneficiaries may be more inclined to voluntarily participate in an evaluation of a youth employment programme.

- a. Qualitative methods:** Well justified choice of the selection of cases and / or qualitative sample (based on strategic, theoretical or practical considerations), number of cases (internal validity, saturation), and discussion of the limitations.
- b. Quantitative methods:** Well justified choice of sampling strategy (e.g. random, stratified), (type(s) of respondents, external validity), sample size (power calculation, response rate) and discussion of the limitations.

⁵ This corresponds with external validity: To what extent can we generalise findings to other contexts, people, or time periods? (Vaessen et al, 2020, pp. 11-12)

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > **14. Sampling strategy***
 - > **Qualitative methods**
 - > Quantitative methods
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Evaluation methodology

14. Choice of sample, cases and information sources*

Qualitative methods

Good 	This criterion can be scored as good when: a. The evaluator has set criteria for whom to interview, and has asked project staff for information about organisations and/or for lists of beneficiaries, which the evaluator has used to make the selection; AND b. Self-selection of respondents is greatly reduced. In addition to including the actively involved stakeholders, the evaluator makes a planned additional effort to include others who participate less actively, have stopped participating or have never participated; AND c. The selected cases follow logically from theoretical and/or practical selection criteria, and are appropriate to answer the research question (e.g. in relation to contexts or types of organisations); AND d. There is a justification of the sampling strategy and case study selection, an explanation of how it was attempted to reduce bias as much as possible, and a discussion of the limitations and possible bias in the evaluation report.
Adequate 	This criterion may be scored as adequate when: a. Project staff and evaluator have had exploratory discussions about the criteria for selecting organisations and persons to interview, but the final selection has been made by the evaluator; AND b. Self-selection of respondents is reduced: Besides including the actively involved stakeholders, the evaluator makes an effort to include others that participate in the project less actively; AND c. The selected cases capture most of the variety of the programme and hence are sufficient for valid conclusions to be drawn; AND d. There is a justification of the sampling and case study selection, and a discussion about the limitations and possible bias in the evaluation report.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- **14. Sampling strategy***
 - **Qualitative methods**
 - Quantitative methods
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

14. Choice of sample, cases and information sources*

Inadequate



This criterion must be scored as inadequate when:

- a. Project staff have determined the information sources, potentially resulting in biased findings; OR
- b. Self-selection of respondents lead to bias; OR
- c. The selected cases do not represent the project as a whole; OR
- d. In the evaluation report there is no justification of the qualitative sampling strategy and case study selection, nor is there a discussion about the limitations and bias of the sampling and case study selection.



Example Inadequate:

- (a) *Example 1: the project staff prepared a list of villages, organisations and interviewees for the evaluator.*
Example 2; the staff helped select which specific results of the programme should be evaluated.
Example 3: the project staff prepared a selection of success stories for the evaluator to validate.

- (b) *While visiting a community, a consultant asked those present to participate in a spontaneous focus group discussion, resulting in the focus group comprising only the older men who were not working in the field during the day.*

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > **14. Sampling strategy***
 - > Qualitative methods
 - > **Quantitative methods**
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Evaluation methodology

14. Choice of sample, cases and information sources*

Quantitative methods

Good 	This criterion can be scored as good when: a. The evaluator sets the criteria for case study selection. Project staff provide information about possible cases, and the evaluator makes the final selection; AND b. Self-selection of respondents is highly reduced; AND c. The sample size is based on a power calculation. This requires (a) one important indicator of interest, (b) the indicator's variance is known and (c) the minimum effect size that the evaluation should be able to find. This calculation gives the minimum sample (e.g. 600 households) needed to detect the effect ; AND d. The evaluation report contains a justification of the sampling strategy and case study selection.
Adequate 	This criterion may be scored as adequate when: a. The evaluator sets the criteria for sampling. Project staff provided information and there is an exploratory discussion about case selection. The evaluator must have made the final selection; AND b. Self-selection of respondents is reduced; AND c. The sample size was copied from similar studies that had shown significant effects; AND d. The evaluation report contains a justification of the sampling strategy and case study selection.
Inadequate 	This criterion must be scored as inadequate when: a. Project staff have played an active role in selecting the sample and case studies, potentially resulting in biased evaluation findings; OR b. The self-selection of respondents has led to bias; OR c. There is no consideration of sample size prior to data collection, or the sample size is too small to find the intended effects; OR d. The evaluation report does not contain a justification of the sampling strategy and case study selection.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- **14. Sampling strategy***
 - Qualitative methods
 - **Quantitative methods**
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

14. Choice of sample, cases and information sources*



Example Good:

(b) If the woman in the household was not available, the household was skipped and the enumerator moved on to a neighbouring household.



Example Adequate:

(b) If the woman in the household was not available for the interview, a second attempt was made the next day. If the woman is again not available, the questions for the women are skipped, and not answered by the man.



Example Inadequate:

(a) Guided by project staff, only the successful project villages close to the capital city were visited for the survey.
(b) A survey included questions that must be answered by women. If the woman in the household was not available for the interview, the male head of the household responded – but women would have responded differently.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- **15. Analyses***
 - Qualitative methods
 - Quantitative methods
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

15. Appropriateness of the analysis*

The analyses are appropriate for the chosen research design.⁶

- a. Qualitative methods:** the data analysis methodology is clear, appropriate for the research design and includes e.g. theory construction, coding, comparison of cases.
- b. Quantitative methods:** statistical analyses are appropriate for the research design, chosen indicators and sample size; comparisons are appropriate, e.g. difference in differences, analysis of variance, regression analysis, matching techniques.

⁶ This corresponds with data analysis validity: To what extent are data analyses methods applied correctly for drawing adequate conclusions? (Vaessen et al, 2010, pp11-12)

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- **15. Analyses***
 - **Qualitative methods**
 - Quantitative methods
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

15. Appropriateness of the analysis*

Qualitative methods

Good 	This criterion can be scored as good when: a. There is a clear analysis strategy which has been fully elaborated and applied. It shows how raw data leads to conclusions; b. Results from different information sources are presented transparently as described in the methodology section (note that information should not be traceable to individuals); AND c. Results are transparently presented along the cause–effect chain (or theory) and considering other factors. They are presented in the sequence expected from the methodology section, step by step, and there is a critical reflection on the assumed theory.
Adequate 	This criterion may be scored as adequate when: a. There is an analysis strategy which shows how raw data has been translated into conclusions, but it is not described in full detail, or it is difficult to see the link between data and conclusions; b. Results from different sources are aggregated when they converge, are presented separately when they don't converge and are presented as announced in the methodology section; AND c. Results are presented along the result chain and considering other factors; they follow the sequence expected from the methodology section and there is a discussion of the assumed theory.
Inadequate 	This criterion must be scored as inadequate when: a. There is no analysis strategy that shows how raw data has been translated into conclusions; b. The results do not transparently present the different information sources as presented in the description of the methodology; OR c. The results do not transparently present the information along the cause–effect chain (or theory) presented in the description of the methodology, and the conclusions have not been substantiated. There is no critical reflection on the theory; d. There is no analysis.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- **15. Analyses***
 - **Qualitative methods**
 - Quantitative methods
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

15. Appropriateness of the analysis*



Example Good:

- (b) *Farmers and staff appreciated training differently: farmers had the impression that government maintenance tasks were delegated to farmers, while government staff had the impression that farmers could easily apply acquired knowledge.*
- (c) *Farmers appreciated the training on irrigation maintenance they received the previous year, but field observations showed that tertiary irrigation had not yet been maintained. Apparently, the training did not sufficiently motivate farmers to carry out maintenance jointly. Some farmers who suffer less from water shortage blame the lack of water coming from upstream villages. Others who suffer more from water shortage blame their peers' unwillingness to carry out joint maintenance. Government extension staff had not visited the village for over a year and had not followed up to discuss this problem with farmers.*



Example Adequate:

- (b) *All agreed that the irrigation channels were poorly maintained and that some farmers receive too little water. Although some farmers point at too little water coming from upstream villages, other farmers blame their fellow villagers' lack of willingness to carry out joint maintenance.*
- (c) *In the evaluation report, training is discussed first, then joint maintenance, then the result in terms of water availability, and then the conclusion that the training was not sufficient to assure joint maintenance is presented and discussed.*



Example Inadequate:

- (b) *Although in the methodology section different information sources have been distinguished, e.g. opinions of farmers and government staff, field observations, the report sums up the results without linking them to the sources: 'the main constraints were lack of water flowing into the irrigation scheme, lack of maintenance, and lack of advice and training'.*
- (c) *'In spite of the training provided last year, farmers still experience a shortage of water in their fields'.*

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- **15. Analyses***
 - Qualitative methods
 - **Quantitative methods**
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

15. Appropriateness of the analysis*

Quantitative methods

Good 	This criterion can be scored as good when: - The presentation of results is as planned in the design; AND - The statistical analysis is as planned in the design, and the results confirm that the analysis is indeed appropriate.
Adequate 	This criterion may be scored as adequate when: - The presentation of the results is not in line with the chosen evaluation design, but this is explained in the report; OR - The statistical analysis draws correct conclusions but does not answer the initial research question.
Inadequate 	This criterion must be scored as inadequate when: - The statistical analysis is not in line with the chosen evaluation design; OR - The statistical analysis leads to spurious conclusions.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- **15. Analyses***
 - Qualitative methods
 - **Quantitative methods**
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

15. Appropriateness of the analysis*



Example Good:

- (a) *The design announced a difference in differences analysis (with–without project and before–after project comparison) and the results are indeed presented that way.*
- (b) *The statistical analysis shows whether or not there is an effect; if no effect was found, there is a discussion about the sample size and the minimum effect size that could be detected (power calculation).*



Example Adequate:

- (b) *'Although the means show a difference, the difference is not significant: the sample size was not large enough to measure a significant effect, so no conclusions about effectiveness can be drawn'.*



Example Inadequate:

- (a) *Although a double difference design (before–after project and with–without project comparison) was planned in the methodology section (or inception report), only before–after or only with–without results are presented and interpreted.*
- (b) *Example 1: 'Households have increased their income by 20%' – but no statistical analysis was done, or the result of the statistical analysis was not significant.*

Example 2: 'The project had no effect on income', but the sample was too small to find a significant effect size.

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > 14. Sampling strategy*
- > 15. Analyses*
- > **16. Evaluation matrix**
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Evaluation methodology

16. Evaluation matrix

An evaluation matrix summarises the methodology. This matrix shows how (i) evaluation questions are translated into (ii) sub-questions / indicators / result areas, and (iii) methodologies and (iv) information sources.

Good 	This criterion can be scored as good when: - A table gives an overview that shows for each evaluation question (i) indicators / result areas, (ii) anticipated evaluation methods, and (iii) information sources; AND - The proposed indicators / result areas and information sources are convincingly adequate to enable each evaluation question to be answered.
Adequate 	This criterion may be scored as adequate when: - There is an overview in the text or in a table that per evaluation question states which indicators or result areas will be used and which information sources will be used (the methods may be presented elsewhere, not in this overview); AND - The proposed indicators / result areas and information sources are likely to be sufficient to enable almost all evaluation questions to be answered.
Inadequate 	This criterion must be scored as inadequate when: - There is no overview either in the text or in a table, that shows how evaluation questions are 'translated' into (i) indicators or result areas, (ii) methods and (iii) information sources; AND - It is unclear whether the methodology is appropriate for answering the evaluation questions.

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > 14. Sampling strategy*
- > 15. Analyses*
- > 16. Evaluation matrix
- > **17. Independent information sources***
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Evaluation methodology

17. Sufficient independent information sources*

Besides sourcing information from project implementers, direct beneficiaries and other local stakeholders, the evaluator should also independently select and consult sufficient independent sources, e.g. the opinion of other experts or non-beneficiaries that can critically reflect on the intervention; objective observations; validated secondary data.

Good 	This criterion can be scored as good when: - A mix of dependent and independent information sources has been used, including external people who can provide critical reflection on the project; AND - Subjective information on perceptions has been complemented with objective information (data from secondary sources, own field observations, measurements); AND - The evaluator has used opportunities to add more information when possible and relevant, aiming for information saturation (snowballing, adding additional documents, interviews during the evaluation); - There is enough flexibility for the evaluator to adjust and add information sources during the evaluation, to reach saturation (extra time and support have been factored into the planning).
Adequate 	This criterion may be scored as adequate when: - A mix of directly involved (dependent) and independent information sources is used; AND - There is some flexibility to add information sources during the evaluation (squeezed in ad hoc, without the need for extra time).
Inadequate 	This criterion must be scored as inadequate when: - The main source of information is monitoring data collected by the implementing organisation; OR - The evaluation consulted no or very few independent informants that were not actively involved in project implementation or have an interest in the evaluation results; OR - Representatives of the implementing organisations were present during interviews or focus groups with beneficiaries; OR - There is no flexibility for the evaluator to adjust and add information sources during the evaluation, to reach saturation (no time, no possibility to re-plan).

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- **18. Triangulation**
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

18. Triangulation of results

This includes a comparison and critical reflection by the evaluator of results from different sources and results from different research methodologies, data collection methods (i.e. interviews, surveys, observations) and data sources (i.e. persons, documents, sites).

Good 	This criterion can be scored as good when: a. Sufficient different methodologies, data collection tools, and information sources have been used; AND b. The primary evaluation results are compared with other evaluations and a literature review (if available: a systematic review); AND c. The differences between results from different evaluation methods, data collection methods and information sources are discussed. In addition, the results from the evaluation and from other evaluations and literature are compared and discussed.
Adequate 	This criterion may be scored as adequate when: Where possible, different evaluation methods, data collection methods and information sources have been used. As a minimum: subjective perceptions are complemented by objective data (secondary data, observations); and opinions of project stakeholders are complemented by opinions of independent persons (experts not directly involved in the project, or non-participant farmers); AND a. At least some use is made of other evaluations or literature on similar interventions. As a minimum: a reflection on the results of earlier evaluations is included; AND b. Results from different evaluation methods, data collection methods and information sources are presented in a disaggregated way and then discussed.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- **18. Triangulation**
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

18. Triangulation of results

Inadequate



This criterion must be scored as inadequate when:

- a. Not enough different methodologies, data collection methods and information sources are used to validly answer the main evaluation questions; OR
- b. No use has been made of data from secondary sources, evaluations of similar interventions or broader literature reviews; OR
- c. There is no comparison and discussion about differences between results from different sources, research methods or data collection methods.



Example Good:

- (a) *A quantitative baseline–endline survey is combined with focus group discussions with farmers; farmer perceptions are combined with objective field observations; the opinions of project stakeholders are combined with the opinions of external experts.*
- (b) *Conservation agriculture practices recommended by the project did not result in improved farm production in the 3rd year of the project. A literature review confirms that this is what would reasonably be expected in the first few years.*
- (c) *Although farmers in focus group discussions were positive about conservation agriculture, field observations revealed that farmers only apply this on a very small test plot. The survey data show an unexpected yield decline, which is confirmed in the literature, where it is reported that the main benefits of conservation agriculture are not short-term yield gains but long-term yield stability and reduced costs of labour and input.*

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- **18. Triangulation**
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

18. Triangulation of results



Example Inadequate:

(a) *Example 1: only quantitative farm production survey data have been used, yet qualitative focus group discussions could have yielded valuable information as well.*

Example 2: only farmer perceptions have been elicited, yet field observations could have added valuable information as well.

Example 3: only project staff and direct beneficiaries have been interviewed about the project effects on farm production, yet external government staff, agricultural experts and non-participant farmers could have provided valuable information as well.

(b) *Based on a quantitative study of crop yield the evaluation has found that the conservation agriculture practices recommended by the project have not resulted in good yields. No literature on conservation agriculture practices has been reviewed to discuss what can be expected from applying these practices under the conditions in prevailing in the project area.*

(c) *'Farmers in focus group discussions were positive about the yield increase. Secondary yield data collected by the district agricultural department shows that yields did not increase' (no comparison or discussion).*

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- **19. Discussion of bias**
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

19. Discussion of bias

The evaluator has provided a critical reflection on different forms of bias (covering at least selection bias, respondent bias and evaluator bias). Note that this criterion does not reflect the extent to which bias has been effectively addressed (criterion 14), nor the extent to which bias has been taken into account when presenting the conclusions (criterion 23).

Selection bias: non-representative choice of cases and respondents (see criterion 14).

Respondent bias: the problem of 'courtesy bias', whereby a respondent tells you what they think you want to hear, is well established. In structured surveys, courtesy bias can affect both people's reported behaviour and self-reported outcomes. Courtesy bias has a clear relevance for qualitative interviews, for example when interviewing respondents about how influential a particular agency or programme has been in effecting a change. A related form of bias is that of 'social acceptability' or 'political correctness' bias, where people provide responses which reflect what they regard as being the socially acceptable thing to say. (White and Phillips, 2012).

Evaluator bias: it is commonly held that there may be biases pushing evaluators towards positive findings, the main one being 'contract renewal bias'. Another bias which may be even more influential is 'friendship bias'. If an evaluator has spent time with programme staff and has developed a good relationship with them, it becomes difficult to upset them with a critical report (White and Phillips, 2012). Other types of evaluator bias are world views or personal convictions about what works and what does not.

Good 	This criterion may be scored as good when: The evaluator discusses in detail the possible biases involved in the process of data collection and analysis.
Inadequate 	This criterion must be scored as inadequate when: a. There is no critical reflection on biases; OR b. There appear to be more biases in the process of data collection and analysis than discussed by the evaluator.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- **19. Discussion of bias**
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

19. Discussion of bias



Example Good:

The evaluator acknowledges and describes that she/he was only able to have discussions with a small group of women still actively participating in the project and was unable to contact women that had stopped participating. She/he notes that this may result in a positive bias, as it is unknown why other women have dropped out: they might have been unsatisfied.



Example Inadequate:

The evaluation acknowledges that only a small number of active women, present in the village during the evaluation, were contacted. However, although the evaluation report does present a high dropout rate, it does not mention that not contacting these women is a potential source of bias.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- **20. Transparent methodology***
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

20. Transparent methodology*

The description of the data collection and analysis is systematic, complete and transparent.⁷ In principle, if another evaluator would read the methodology description, the evaluator would be able to apply the same methodology (replicability).

Good 	This criterion may be scored as good when: There is a systematic, complete and transparent description of the data collection and analysis;⁸ the methodology section describes criteria for selecting respondents, As a result, another evaluator would be able to apply the same methodology.
Inadequate 	This criterion must be scored as inadequate when: There is no systematic, complete and transparent description of the data collection and analysis. As a result, another evaluator would not be able to apply the same methodology.

⁷ This corresponds with reliability: the transparency and replicability of the evaluation process (Vaessen et al, 2010, pp11-12)

⁸ It is possible for the detailed methodology to be included in an annex in the final evaluation report, or in a separate inception report.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- **20. Transparent methodology***
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

20. Transparent methodology*



Example Good:

The report annexes include guidelines (topics and guiding questions) for interviews and group discussions and – if applicable – survey questionnaires, criteria for the selection for respondents are presented, the analysis is described and can be recognised in the way results are presented.



Example Inadequate:

Example 1: for a survey, the topics are mentioned, but the questionnaire is not included.

Example 2: the criteria used to select respondents (for a survey or for interviews) are not clear.

Example 3: although the questionnaires and topics for a survey are described, it remains unclear which methods have been used to process the resulting data and arrive at the conclusions.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- **21. Limitations***
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

21. Limitations of the evaluation*

The evaluation's limitations are adequately discussed. The evaluator is self-critical and discusses the limitations of the study, including its reliability, internal and external validity, the relative contribution of the intervention and other external factors affecting the observed changes. Note that this criterion does not assess the limitations of an evaluation, but rather the acknowledgement, discussion and implications of limitations. Providing an evaluation's limitations have been made clear, it may still yield valuable information.

Good 	This criterion can be scored as good when: There is an elaborate discussion about all of the limitations of the evaluation.
Adequate 	This criterion may be scored as adequate when: The evaluation briefly mentions the limitations of the evaluation and takes them into account sufficiently when presenting the findings and conclusions. If the external validity is not mentioned, the criterion may still be scored as adequate as long as the report does not generalise the findings beyond the cases studied cases or implies they are generalisable.
Inadequate 	This criterion must be scored as inadequate when: - There is no discussion about the reliability, validity or external factors that may have contributed to the evaluation's findings; OR - The evaluation's limitations are mentioned, but their implications are not sufficiently taken into account.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- **21. Limitations***
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Evaluation methodology

21. Limitations of the evaluation*



Example Good:

In the methodology section, two limitations are discussed:

1. *Only one of the project activities, provision of school meals, could be visited. Another, hygiene awareness at home, could not be included in the evaluation. The evaluation can therefore only draw conclusions for that part of the project.*
2. *One ambitious outcome, reduced child malnutrition, could not be established by the evaluation. The evaluation cannot therefore draw conclusions on reduced malnutrition.*



Example Inadequate:

From the methodology or results section, it becomes clear that only one of the project activities (provision of school meals) was visited, and that one outcome (reduced child malnutrition) could not be established. However, this is not discussed as a limitation.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- **22. Conclusions answer questions***
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Results and conclusions

22. Conclusions answer research questions*

Although the conclusions may be organised or grouped differently than the original research questions, in principle all research questions have been answered, or it is explained why they could not be answered.

Good 	This criterion can be scored as good when: All research questions have been answered. The conclusions may be organised differently than the original research questions.
Adequate 	This criterion may be scored as adequate when: Not all research questions have been answered but it is explicitly mentioned which research questions could not be answered and why (for example, limitations that emerged during the evaluation that could not have been foreseen).
Inadequate 	This criterion must be scored as inadequate when: Not all research questions have been answered and it is not explained why they have not been answered.

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > 14. Sampling strategy*
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > **23. Conclusions based on findings***
- > 24. Validation of conclusions
- > 25. Usefulness of the recommendations
- > 26. Readability

Results and conclusions

23. Conclusions based on findings*

- a. The results of each method are presented fully and transparently, to avoid unsubstantiated conclusions being drawn. Detailed results may be presented in an annex.
- b. The limitations and validity of the conclusions are discussed (in line with criterion 21).

Good 	This criterion can be scored as good when: a. All research findings from all methods are transparently presented, either in the evaluation report or in an annex; AND b. There is a (brief) reflection on the limitations of the evaluation; AND c. The conclusions presented are consistent with all research findings, and take the limitations of the evaluation into account.
Inadequate 	This criterion must be scored as inadequate when: a. Not all research findings from all methods are transparently presented, either in the main report or in an annex; OR b. Not all research findings are taken into account in the conclusions, but instead, certain results or findings are presented to fit the evaluator's narrative; OR c. The limitations of the evaluation are not taken into account when presenting the conclusions (see also criteria 19 and 21).

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- **23. Conclusions based on findings***
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- 26. Readability

Results and conclusions

23. Conclusions based on findings*



Example of a limitation that is discussed and taken into account in the conclusions:

For a large youth employment programme, young men and women received business training and investment capital to start a business – the amount of money depended on the quality of the applicant's proposal.

In the methodology section the limitations are discussed. 'The evaluation was unable to contact women that had not participated in the abovementioned youth employment programme. We do not know why they did not participate: were they not interested, or were they not offered project support? Moreover, we could not compare women that received business development support with women that did not receive this support, which would have helped in attributing effects to the programme.'

In the conclusions sections, these limitations are taken into account: 'Women that received business development support had successfully increased their market sales by the end of the project. However, it is not clear to what extent this can be attributed to the evaluated programme, because it remains unclear whether the project was inclusive to all interested women, or that only privileged or already entrepreneurial women were able to participate. Moreover, we do not know whether the support has had a positive spill-over effect on other women, or whether it has crowded out other women selling produce to the same market.'



Example of a limitation that is mentioned, but not taken into account in the conclusions:

For a large youth employment programme, young men and women received business training and investment capital to start a business – the amount of money depended on the quality of the applicant's proposal. After the programme had ended, the evaluators decided to perform only a few in-depth studies of some of the larger successful individual businesses.

In the methodology section, the evaluators concede that they incorporate selection bias not only by focusing on success stories, but also because they focused on the larger projects. This pragmatic approach, they argue, can help them in formulating certain conditions for success, and thus facilitates learning.

When discussing the findings and conclusions of the study, however, the evaluators still make claims about the effectiveness of the entire programme, even though the selection bias prevents them from generalising their findings for the entire group of beneficiaries.

- > Introduction
- > Reading guide
- > The list of evaluation criteria
- > 1. Reference group
- > 2. Independence of evaluators*
- > 3. Context of the intervention
- > 4. Description of the intervention*
- > 5. Validation of assumptions*
- > 6. Evaluation's objective
- > 7. Scope of the evaluation
- > 8. OECD DAC evaluation criteria
- > 9. Evaluation questions
- > 10. Research design*
- > 11. Methods for effectiveness*
- > 12. Methods for efficiency
- > 13. Indicators or result areas*
- > 14. Sampling strategy*
- > 15. Analyses*
- > 16. Evaluation matrix
- > 17. Independent information sources*
- > 18. Triangulation
- > 19. Discussion of bias
- > 20. Transparent methodology*
- > 21. Limitations*
- > 22. Conclusions answer questions*
- > 23. Conclusions based on findings*
- > **24. Validation of conclusions**
- > 25. Usefulness of the recommendations
- > 26. Readability

Results and conclusions

24. Validation of provisional conclusions

To substantiate the conclusions, the provisional conclusions have been discussed e.g. in a validation workshop with project implementers and independent experts, and compared with findings in earlier evaluations and broader literature. (Note that evaluators remain the independent authors of the conclusions, but project implementers may point out factual inaccuracies, add information, and comment on the usefulness and realism of the recommendations, see criterion 25.)

Good 	This criterion can be scored as good when: a. In addition to validating the findings through the reference group meetings, the evaluators have validated the findings with independent experts.; AND b. The provisional conclusions have been discussed with the project implementers prior to publication; AND c. The conclusions are clearly positioned in the broader context of findings from previous evaluations of similar interventions and, if available, of high quality research such as impact evaluations and systematic reviews. If the findings from the evaluation deviate from the existing knowledge base, in their findings and conclusions the evaluators must have reflected on the differences.
Adequate 	This criterion may be scored as adequate when: a. The provisional conclusions have been discussed with the project implementers prior to publication; AND b. The conclusions are compared with findings from previous evaluations of similar interventions and, if available, with high-quality research such as impact evaluations and systematic reviews.
Inadequate 	This criterion must be scored as inadequate when: a. The provisional conclusions have not been discussed with project implementers prior to publication.; OR b. The conclusions are not compared with previous evaluations of similar interventions or, if available, with high quality research such as impact evaluations or systematic reviews.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- **25. Usefulness of the recommendations**
- 26. Readability

Usefulness and readability of the evaluation report

25. Usefulness of the recommendations

Good 	This criterion can be scored as good when: a. The recommendations reflect the objectives of the evaluation (see criterion 6); AND b. The recommendations follow logically from the evaluation's conclusions and findings; AND c. The recommendations are realistic and match the sphere of influence of the user(s).
Inadequate 	This criterion must be scored as inadequate when: a. The recommendations are not in line with the stated evaluation objective (see criterion 6); OR b. The recommendations do not follow logically from the evaluation's findings and conclusions; OR c. There is no clear action perspective in the recommendations. The recommendations are unrealistic or do not match the sphere of influence of the user(s) of the evaluation; OR



Example Inadequate:

Example 1: an evaluation in a fragile country recommends increasing focus on physical well-being, because the study finds that human security is an important prerequisite for strengthening the rule of law. The actual well-being of the population, however, is beyond the sphere of influence of the project implementers.

Example 2: an evaluation finds that sexuality education can be an effective tool for influencing norms and values regarding sexual rights of minorities and recommends scaling up the activity without considering the increasingly restrictive political climate for such activities.

- Introduction
- Reading guide
- The list of evaluation criteria
- 1. Reference group
- 2. Independence of evaluators*
- 3. Context of the intervention
- 4. Description of the intervention*
- 5. Validation of assumptions*
- 6. Evaluation's objective
- 7. Scope of the evaluation
- 8. OECD DAC evaluation criteria
- 9. Evaluation questions
- 10. Research design*
- 11. Methods for effectiveness*
- 12. Methods for efficiency
- 13. Indicators or result areas*
- 14. Sampling strategy*
- 15. Analyses*
- 16. Evaluation matrix
- 17. Independent information sources*
- 18. Triangulation
- 19. Discussion of bias
- 20. Transparent methodology*
- 21. Limitations*
- 22. Conclusions answer questions*
- 23. Conclusions based on findings*
- 24. Validation of conclusions
- 25. Usefulness of the recommendations
- **26. Readability**

Usefulness and readability of the evaluation report

26. Readability of the report

The report is readable, consistent and includes a clear summary containing the evaluation's objective, evaluation questions, conclusions and recommendations.

Good 	This criterion can be scored as good when: a. The summary includes the (1) objective, (2) short policy description, (3) main evaluation questions, (4) main findings and conclusions that cover the main evaluation questions, (5) recommendations; AND b. In the main report, there is consistency between evaluation questions --> methodology --> results --> conclusions, and the conclusions answer the evaluation questions; AND c. The text is well written and unambiguous.
Inadequate 	This criterion must be scored as inadequate when: a. There is no summary; OR b. The summary does not include the (1) objective, (2) short policy description, (3) main evaluation questions, (4) main findings and conclusions that cover the main evaluation questions, (5) recommendations; OR c. In the main report, there is an inconsistency between evaluation questions and methodology, OR between methodology and results, OR between results and conclusions, OR between conclusions and evaluation questions; OR d. The text is not easy to read and is sometimes ambiguous.



Version March 2022

Contact at IOB

Caspar.Lobbrecht@minbuza.nl

Ferko.Bodnar@minbuza.nl