

SDC's Quality Assessment of Decentralised Evaluations



Commissioned by the Evaluation and Controlling Unit of the Swiss Agency for Development and Cooperation (SDC)

Contents:

- I SDC Senior Management Response
- II Evaluators' Final Report

Bern, June 2023

Management Response to the Quality Assessment of Decentralised Evaluations

The response to the quality assessment of decentralised evaluations states the position of SDC's Foundations and Quality Division (GQ) on the recommendations made by the quality assessment. The Evaluation and Controlling Specialist Service (EC) commissioned an external quality assessment of SDC's decentralised evaluations. The assessment (1) looked at the quality of the evaluations and created a baseline of the current quality; (2) assessed how SDC staff in charge of decentralized evaluations were using the support currently available to them (toolkits, Project Cycle Management Course (PCM), etc.) and how it could be improved; (3) assessed the support needs of SDC staff for managing evaluations and the quality of the available support; and (4) assessed the terminology of the evaluations and proposed, where applicable, new criteria and terminology.

Between 2018 and 2020 the Swiss Agency for Development and Cooperation (SDC) mandated more than 150 decentralised evaluations of which almost two thirds (92) were used to calculate the success rate. In order to assess the quality of the evaluations, the evaluation reports and the Terms of Reference (ToR), a sample of 60 evaluations were scrutinized by a document analysis. The sample included 48 evaluations that were used for the success rate and 12 evaluations that did not meet the criteria of EC and were therefore not used for the success rate. In addition, the utility of the evaluations for SDC staff was investigated by an online survey with 133 respondents and interviews with ten selected survey participants.

Appraisal of the Assessment

The Quality Assessment (hereafter Assessment) was conducted by a team of independent experts in accordance with international standards. During the process there was an intense exchange between EC and the team of consultants. EC notes that the assessment sets high standards and uses stringent criteria which the assessment does not always achieve itself. Furthermore, there was no comparison to international benchmarks or to the performance of similar agencies. The report provides a large number (39) of detailed recommendations without prioritisation. The report produced a large body of findings. They are considered to be useful inputs for the definition of measures to strengthen the quality of decentralised evaluations in SDC.

The findings will serve as an input to

1. the revision of SDC's evaluation guidelines,
2. the revision of the How-to Note Evaluation (together with the Specialist Service Quality Assurance and Internal Digitalisation, QS),
3. the optimisation of training courses to SDC staff (together with QS), and
4. EC strategic reflection on its roles and responsibilities in the new structure after the reorganisation of SDC in 2022.

Each measure defined in the response to the quality assessment falls under at least one of these four areas.

The Assessment rated the quality of decentralised evaluations conducted by SDC as high with respect to the Terms of Reference, the Executive Summary, and the Use and Usability and medium with respect to Introduction and Context, Evaluation Findings, and Conclusions and Recommendations. The Assessment judged the quality of the methodology used in decentralised evaluations to be in need of improvement or inadequate in around 72% of

evaluations reviewed. SDC is taking this finding seriously. Further analysis shows that the low score stems from weaknesses in the description of the evaluation design and the quantitative elements of the methodology. Qualitative elements, as exemplified through the high rating (85% very good, good or satisfactory) of the sources of evidence, are robust. SDC has been aware of gaps in its quantitative data collection. To address these issues, it has started in a concerted effort the collection of results data through Aggregated Results Indicators (ARI) and Thematic Results Indicators (TRI) and the digitalisation of its collection system through the Results Data Management (RDM) project. In addition, as described in the response and measures to each recommendation, SDC will 1) endeavour to strengthen evaluation knowledge and skills throughout the institution and 2) define more clearly minimum standards for decentralised evaluations. SDC will further put a focus on the adequate description of employed methodologies in decentralised evaluations.

Main Findings

The overall findings of the assessment are as follows.

Terms of Reference

- Most ToRs are appropriate and complete.
- The Inclusion of evaluation criteria, the description of the evaluation object, and the description of the responsibilities, deliverables and schedule are rated exceptionally well.
- Descriptions of the rationale and purpose, as well as of the scope of the evaluation are assessed as rather weak.

Executive Summary

- The writing style is mostly rated good or very good.
- Information on evaluation design and lessons learnt is often missing.

Introduction and Context Analysis

- Only about half of the introductions are of satisfactory quality.
- It is not always clear how the context analysis relates to the evaluation.
- More than two thirds of reviewed evaluations do not provide a clear Theory of Change or Results Model.

Evaluation Methodology

- Information on evaluation design, evaluation matrix and data collection plan is often missing.
- The data collection instruments are not always presented.
- The data analyses applied and the limitations of the evaluation are often not discussed.
- A mixed-method approach is rarely applied.

Evaluation Findings

- Data basis is mostly presented sufficiently.
- Discussion of Relevance, Effectiveness and Sustainability is mostly adequate.
- Findings, conclusions and recommendations are not always separated.
- Discussion of Impact and Efficiency are often insufficient.

Conclusions and Recommendations

- Conclusions are mostly logically connected to findings.
- Recommendations derive from conclusions.
- Recommendations are often not sufficiently actionable.
- Lessons learnt are missing in half of the reports.

Implementation and Process Use

- SDC staff is mostly comfortable with defining the deliverables of an evaluation, preparing the reference documents, and specifying the budget, the process and the timeline.
- SDC staff reported that it is challenging to assure the quality of the Approach Paper and the Terms of Reference, to specify the methodology, to define the evaluation questions, and to recruit suitable evaluators.

Limiting Factors

- Some evaluation findings are insufficiently comprehensible. In some instances, recommendations lack clarity, persuasiveness and actionability.
- Some evaluation analysis is superficial, inadequate and / or incomplete. The evaluators lack objectivity and understanding of the intervention.
- Project stakeholders tend to resist adopting changes. Regulations are not supportive.
- External circumstances change or deteriorate and make recommendations obsolete.

Support Needs

- Staff most frequently use support through written guidance followed by personal counselling.
- The How-to Note Evaluation is little known.

Terminology

- The term *Evaluation* is used in a rather broad sense.
- In order to qualify as an evaluation, the following minimum criteria must be met: 1) a well-developed analytical framework, 2) the dimensions of analysis must be clearly defined, 3) minimum methodological standard must be applied, and 4) independence of evaluators facilitates the achievement of a sound and objective assessment.

Recommendations

The table below summarises the Assessment's recommendations as well as the response and measures proposed by EC.

Bern, June 14, 2023



Jean-Luc Bernasconi, Head of Foundations and Quality Division

Recommendations, response and measures

A transversal measure is to include Evaluation in the standard PCM training (virtual and in person). QS and EC will cooperate in this endeavour.

	Recommendation	Response	Justification and measures	Deadline
Terms of Reference				
1	Provide a more realistic budget and timeframe.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. In addition, the recommendations will be integrated into the PCMi course and there will be awareness raising in the PCM courses. <i>[EC, QS]</i>	2024
2	Include information about the reasons for the evaluation and its intended use and users.			
3	Provide more details about the scope of the evaluation.			
4	Stipulate qualitative and quantitative methods.			
Executive Summary				
5	Advise evaluators to include information about the rationale of the evaluation, its design and methodology as well as a summary of the lessons learnt.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2024
6	Make clear that the Executive Summary has to be comprehensible by its own.			

	Recommendation	Response	Justification and measures	Deadline
Introduction and Context Analysis				
7	Request considering SDC, international, national and regional strategies and policies.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2024
8	Request connecting the context analysis specifically to the intervention's goals and activities.			
9	Check for completeness of the evaluation object's description, including a clear and concise results model (also referred to as Theory of Change or Logframe).			
Evaluation Methodology				
10	Request evaluators to be explicit about the evaluation design and methodology.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2024
			Minimum standards for evaluation defined in evaluation guidelines. <i>[EC]</i>	2023
11	Make the presentation of an evaluation matrix and data collection instruments mandatory.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2024
12	Ask for a discussion of the limitations and challenges of the evaluation and its results.			
Evaluation Findings				
13	Request an attribution/ contribution analysis.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2024
14	Clarify the content of each OECD DAC criterion.			

	Recommendation	Response	Justification and measures	Deadline
Conclusions and Recommendations				
15	Advise evaluators to prioritize recommendations and provide guidance on who should do what and when.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2024
16	Make it mandatory for all evaluation reports to include a lessons learned section.			
Improving Implementation and Process Use				
17	Develop and implement measures to strengthen evaluation capacities of SDC staff.	Agree	These measures are to be coordinated with the effort of a continuing education concept currently being considered by SDC. Inclusion of evaluation material in the PCMi course. Awareness raising on evaluation materials in the PCM courses. <i>[EC, QS]</i>	2024
			Consider feasibility of introducing an online course on evaluation targeted at SDC staff. <i>[EC, QS]</i>	2024
18	Regularly inform about available sources of information.	Agree	EC will present the main findings of the Quality Assessment in all operational divisions of SDC and in the QS Network meeting. <i>[EC]</i>	2023
			Regular information on evaluations, available platforms, and SDC guidance materials will be increased using the QS newsletter. <i>[EC, QS]</i>	2023
19	Develop and maintain systematically evaluator rosters.	Disagree	Project evaluations are decentralised and often use local consultants in an effort to support local capacities and draw on local knowledge. A centralised roster is considered not productive in this setting. EC is aware of the fact that the definition of relevant qualifications for an evaluator can pose a challenge in this context and will define minimum standards for evaluators.	2024
			Minimum qualifications to be expected from an evaluator will be included in the 'How-to Note'. <i>[EC, QS]</i>	

	Recommendation	Response	Justification and measures	Deadline
20	Provide project management with guidelines with selection criteria for evaluators and advise them in assessing the quality of a proposal adequately.	Partially agree	EC agrees that minimum standards for evaluators must be included in the existing guidelines ('How-to Note'). EC does neither have the mandate, nor the capacity to advise and review the recruitment process of operational units. For measures, see recommendation #19.	
21	Plan sufficient time for preparing, commissioning, managing, and quality assuring the evaluations, and for translating recommendations in real action.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2024
22	Promote Inception Reports.			
23	When planning an evaluation elaborate on the purpose and objectives of the evaluation for different stakeholders.			
Limiting Factors				
24	Introduce a capitalization strategy for decentralized evaluations.	Partially agree	Under the current International Cooperation (IC) Strategy 21-24, SDC has introduced measures to strengthen evidence-based decision making. This includes the systematic consultation of studies, project experiences and evaluations during the project inception phase. Further, SDC's Project Cycle Management (PCM) tools and trainings emphasise the importance of monitoring, evaluation and learning (MEL) as tools for the feedback cycle from planning to implementation, steering and new planning. Furthermore, management of thematic knowledge is in the responsibility of thematic networks. SDC will continue to foster the development of evaluative thinking as part of evidence-based decision making. <i>[EC, QS]</i>	ongoing

	Recommendation	Response	Justification and measures	Deadline
25	Develop an obligatory reference framework for the implementation of decentralized evaluations and the utilization of their results.	Partially agree	With the SDC Evaluation Guidelines and the 'How-to Note' such a framework for the implementation of decentralized evaluations and the utilization of their results already exists. EC will review these documents in light of the findings of the Quality Assessment and the ongoing parliamentary (PVK) review. In this context, EC will examine the introduction of a tier model approach for project evaluations as used by other donors. A tier model approach bases evaluation requirements on the risks of the project in question. <i>[EC]</i>	Guidelines: 2023 How-to Note: 2024
26	Emphasize the Management Responses.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2024
27	Focus on usability of the evaluation process and the evaluation results during the inception, implementation and dissemination of an evaluation.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. In particular, EC will review the template ToRs for evaluations included in the 'How-to Note' and review the section on usability, if necessary. <i>[EC, QS]</i> Consider recommending the use of CLP and peer format in decentralised evaluations. <i>[EC]</i>	2024 2023
Support Needs				
28	Jointly revise and update the existing tools supporting the planning and implementation of decentralized evaluations and distribute them among SDC project management.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2024
29	Foster regional Evaluation Capacity Development for local evaluators.	Partially agree	EC is engaging in the strengthening of capacity development of local evaluators through its existing and ongoing support to the World Bank hosted Trust Fund 'Global Evaluation Initiative', GEI. EC plans to continue this involvement.	

	Recommendation	Response	Justification and measures	Deadline
30	Introduce an obligatory on-the-job online training program.	Partially agree	EC will consider how to reach as many SDC staff as possible. This measure is to be coordinated with the effort of a continuing education concept. EC will present the main findings of the Quality Assessment in all operational divisions of SDC and in the QS Network meeting. <i>[EC]</i> EC will consider whether to introduce online webinars or e-learning courses for decentralised evaluations. <i>[EC]</i>	2023 2024
31	Establish a mentoring network or regional focal points.	Disagree	SDC directorate considers that with the reorganisation of SDC in 2022, SDC has strengthened its decentralised nature. Hence, also for evaluation purposes, SDC will rely on its decentralised system at the country level.	
	Terminology			
32	Providing an analytical framework (evaluation matrix, data collection plan, or the like) should be made mandatory for every evaluation.	Agree	EC will define minimum standards in revised evaluation guidelines. EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2023 2024
33	Results in an evaluation report should always be structured according agreed analysis dimensions.	Agree	EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2024
34	Adhere to minimum methodological standards, e.g., theory-based approach, mixed-methods approach and check adequacy of proposed evaluation design.	Agree	EC will define minimum standards in revised evaluation guidelines. <i>[EC]</i> EC will review the 'How-to Note' Evaluation. During that review, EC will strengthen relevant sections where necessary. <i>[EC, QS]</i>	2023 2024
35	Let only independent consultants – i.e. not being involved in the planning and implementation of the project – conduct evaluations.	Partially agree	SDC's 'How-to Note' Evaluation clarifies that external evaluations must be carried out by persons or teams outside the organisation. In addition to this requirement, EC will define minimum standards in revised evaluation guidelines <i>[EC]</i>	2024 2023

	Recommendation	Response	Justification and measures	Deadline
36	Develop a guide defining the terminology which is to be applied within SDC to clearly distinguish between evaluations and other kinds of studies.	Partially agree	EC will not elaborate a standalone guidance on terminology. This guidance will be included in the revised evaluation guidelines. [EC]	2023
37	Only call studies, which fulfil the above-mentioned demands as evaluations.	Partially agree	EC will review the evaluation guidelines clarifying the terminology. However, EC will consider whether to declare such a terminology as mandatory or not. Imposition of a terminology would be hard to enforce and could, potentially, hamper trust-building. [EC]	2023
Systemic Recommendations				
38	Adapt SDC's evaluation guidelines and introduce a coherent evaluation system. <ul style="list-style-type: none"> • Conduct decentralised evaluations under the responsibility of the project management, supported by QS. • Sample of centralized evaluations (external, independent, final/ex-post). • Introduce Inception phases. • Meta Evaluations of decentralized evaluation reports to assess quality of evaluations, every 3 years. 	<p>Agree</p> <p>Disagree</p> <p>Agree</p> <p>Agree</p>	<p>Status quo.</p> <p>SDC directorate considers that with the reorganisation of SDC in 2022, SDC has strengthened its decentralised nature. Hence, project evaluations will remain decentralised.</p> <p>See identical recommendation 22.</p> <p>Mandate regularly (e.g., every 3 to 4 years) an external quality assessment of decentralised evaluations. [EC]</p>	<p>ongoing</p> <p>Ongoing (2028/26)</p>

	Recommendation	Response	Justification and measures	Deadline
	<ul style="list-style-type: none"> Evaluation synthesis of all evaluations (sectors, topics). Conduct selected impact evaluations. Comprehensive capitalization strategy. 	<p>Agree</p> <p>Partially agree</p> <p>Partially agree</p>	<p>Mandate regularly (e.g., every 3 to 4 years) an evaluation synthesis. [EC]</p> <p>EC has promoted impact evaluations through its <i>Impact Award</i> impulsion programme. The programme has been concluded successfully in 2023. In future, impact evaluations will be mandated by project management where appropriate and useful. The necessary information and guidance will be highlighted in the guidance and training materials on evaluation.</p> <p>See identical recommendation 24.</p>	Ongoing (2024)
39	Strengthen the evaluation function within SDC.	Disagree	SDC directorate considers that the evaluation function is adequately staffed for the size of SDC. During the reorganisation of SDC in 2022, the number of staff have been held constant or reduced. Strengthening the evaluation function is not deemed opportune.	

II Evaluators' Final Report

Commissioned by the Evaluation and Controlling Unit of the Swiss Agency
for Development and Cooperation (SDC)

Authors:

Dr. Stefanie Krapp,
Center for Continuing Education, University of Bern
stefanie.krapp@unibe.ch

Julie Ngo, CEval GmbH
j.ngo@ceval.de

Dr. Stefan Silvestrini, CEval GmbH
s.silvestrini@ceval.de

Contact

Center for Continuing Education, University of Bern
www.zuw.unibe.ch
CEval GmbH, Saarbrücken
www.ceval.de

June 2022

Content

Content.....	I
Executive Summary.....	II
1. Introduction.....	1
1.1 Background.....	1
1.2 Object and scope of the Quality Assessment.....	1
1.3 Objectives of the Quality Assessment.....	2
2. Methodology.....	3
3. Key results from the Quality Assessment of the evaluation reports.....	3
3.1 Terms of Reference.....	4
3.2 Executive Summary.....	4
3.3 Introduction and Context Analysis.....	5
3.4 Methodology.....	5
3.5 Evaluation Findings.....	6
3.6 Evaluation Conclusions and Recommendations.....	7
4. Use and usability of decentralized SDC project and program evaluations.....	7
4.1 Implementation and Process Use.....	8
4.2 Usability of Decentralized Evaluations.....	10
4.3 Use of Evaluation Results.....	11
5. Support needs and quality of support.....	14
6. Terminology.....	15
7. Annexe.....	

Executive Summary

Between 2018 and 2020 Swiss Development Cooperation (SDC) mandated more than 150 decentralised evaluations of which almost two thirds (92) were used to calculate the efficiency rate¹ of SDC. SDC's Evaluation and Corporate Controlling Division (E+C) decided to commission a Quality Assessment of these evaluations, in order to assess the validity and reliability of the evaluation results and their actual use and to improve its evaluation system overall. This Quality Assessment focuses on the methodological quality and on the utility of decentralized SDC evaluations in order to:

- 1) assess the quality of the evaluations and create a baseline of the current quality,
- 2) assess how SDC staff in charge of decentralized evaluations are using the support (toolkits, PCM training, etc.) currently available and how it can be improved,
- 3) assess the support needs of SDC staff for managing evaluations and the quality of the available support, and
- 4) assess the terminology of the evaluations and if necessary, propose new criteria and terminology.

In order to assess the quality of the evaluations, a sample of 60 evaluation reports and their Terms of Reference (ToR) were scrutinized using document analysis. In addition, the utility of the evaluations for SDC staff was investigated through an online survey with 133 respondents and interviews with ten selected survey participants.

The Quality Assessment was conducted between September 2021 and February 2022 by a team composed of Julie Ngo (CEval GmbH), Dr. Stefanie Krapp (ZUW, University of Berne), and Dr. Stefan Silvestrini (CEval GmbH) as team-leader. This report contains the main findings of the Quality Assessment. The findings are presented to the SDC E+C division. The Quality Assessment should contribute to revealing the strengths and weaknesses of current evaluation practice and support its improvement with particular focus on the added value of its results.

The report of this Quality Assessment contains two volumes. In volume one, the experts present their overarching conclusions and derive specific recommendations for action. Volume two contains the detailed findings from the analysis of the sample of SDC decentralized evaluation reports, the online survey, and interview results.

In addition to the two volumes, the key results are summarized in this Executive Summary under main strengths and weaknesses of SDC's decentralized evaluation reports and frame conditions for their implementation and use. The subsequently presented main recommendations have been selected according to their institutional relevance and systemic potential. Guidance for action to eventually implement those recommendations is provided in the subchapters of the main report.

Decentralized evaluation reports

Strengths

- ✓ The experts assess the **ToR** as most appropriate and complete.
- ✓ The experts conclude the **executive summaries** are of good quality with deficiencies in their comprehensibility and consistency with the main report.
- ✓ The experts assess the criterion **effectiveness** as the most adequately discussed among all the OECD-DAC criteria.

¹ The efficiency rates look at the quality of the projects and programs based on evaluation results. It is calculated to provide a comparable benchmark for the efficiency of the SDC over time.

- ✓ The experts state that all evaluation reports contain **conclusions and recommendations** and these conclusions and recommendations mostly refer to the findings of the evaluation.
- ✓ Most survey respondents are quite satisfied with the evaluation reports they received and consider them as **useful for their office**.

Weaknesses

- ✓ The experts consider the evaluations' **context analyses** being improvable. Furthermore, they state that many reports lack sufficiently outlined ToC/results model of the evaluation object.
- ✓ The experts assess the quality and comprehensiveness of the evaluations' **methodologies** as in need of improvement.
- ✓ According to the expert's judgement, the evaluators struggle most with the OECD-DAC criterion **impact**. Unintended impacts are often not discussed at all. The criteria coherence has only been applied in some evaluations because it was just recently introduced.
- ✓ The experts see room for improvement of the **conclusions and recommendations** in terms of prioritization, time-boundness and actionability.
- ✓ The experts state that only about half of the analyzed reports provide **lessons learned**.
- ✓ The majority of the survey respondents see shortcomings of the evaluation report in the **length of the evaluation report and in the applicability and the foundation of recommendations**. Most interview respondents assume that evaluators might tend to go the easy way and collect data selectively. Accordingly, evaluation reports contain information that is already known by the project staff.

Frame conditions for implementing decentralized evaluations

Strengths

- ✓ The majority of the survey respondents positively perceive the **competencies of evaluators, as well as well as their communication and management skills**. This speaks for a good collaboration between evaluators and those commissioning and managing the decentralized evaluations.
- ✓ The majority of the survey respondents perceive their **own support for the evaluation** as sufficient regarding time invested and logistical as well as technical and methodological support. Most tasks associated with the preparation of an evaluation are reported to be handled well.
- ✓ As regards the utility of the evaluation, the survey results show **lesson learnt as most relevant outcomes**, followed by providing legitimacy.
- ✓ QS already provides a variety of **support measures** for setting up and managing decentralized evaluations, like guidelines and trainings. Survey respondents mainly use support measures in the form of written guidance, followed by personal counselling.

Weaknesses

- ✓ For the majority of the survey as well as interview respondents, the **quality assurance of the Terms of Reference** is most challenging, particularly specifying the methodology, defining the evaluation questions, and finding suitable evaluators.
- ✓ The interview respondents have strong consensus that the **impartiality and independence of the evaluators** are at times threatened. They state that the project management often tends to work with evaluators they already know and this influences at times the recommendations.
- ✓ The majority of the interview respondents emphasize **evaluators' deficiencies** regarding their analytical skills, sectorial expertise and regional experience.

- ✓ In contrast to the results of the survey, most interview respondents claim **insufficient allocated time for different tasks in preparing and managing evaluations** as they often underestimate the time needed. This reduces the quality of the ToR, negatively impacts the whole evaluation process, and limits the potential of evaluations as an instrument for evidence-based decision-making.
- ✓ Most interview respondents see **personal guidance as more valuable for supporting the planning and management of evaluations** than written material. This implicates the dependency on others and an already established network of colleagues who could be approached for advice, which limits the possibilities of new staff in the country/region.

Use of decentralized evaluation reports

Strengths

- ✓ The majority of the survey as well as interview respondents **use recommendations mainly for planning new interventions, improving intervention approaches, and for increasing the acceptance or legitimacy of foreseen changes**. Thereby the reports serve the direct purpose of use on project level.
- ✓ The majority of the survey as well as interview respondents state adaptations of the interventions, changes in attitudes, awareness, and ownership of intervention staff, partners and stakeholders as **main changes** occurring due to an evaluation.

Weaknesses

- ✓ As the **recommendations are mainly used for the project itself** (see above), the experts conclude that uses beyond the project level do not seem to play a major role and learning and use for other projects falls short.
- ✓ As also **changes arising from the use of evaluation results** are limited to the intervention itself (see above), the experts conclude that changes with a broader impact are missing.
- ✓ Survey respondents see the **main reasons for limited use of evaluation results** in the shortcomings of the recommendations, incompleteness of analysis, impartiality of evaluators, resistance of stakeholders to adopt changes, and unsupportive regulations. Some interview respondents see systemic issues rooted in SDCs organization culture as impeding factors for evaluation use.
- ✓ The experts conclude that the **potential of evaluations** for institutional learning, decision-making beyond the perspective of planning a new phase, and accountability is not fully exploited, if the report is not even distributed nor discussed beyond the project staff and a Management Response is not drafted or implemented.

Institutional recommendations

The following recommendations have been selected according to their institutional relevance. They are based on various, more detailed recommendations geared towards action in each subchapter of the main report:

- ✓ QS and E+C should jointly develop and implement measures to **develop evaluation capacities of SDC staff**.
- ✓ QS should introduce a **capitalization strategy for decentralized evaluations**.
- ✓ SDC should develop an obligatory **reference framework** for the implementation of decentralized evaluations and the utilization of their results.
- ✓ SDC leadership should foster **regional Evaluation Capacity Development**.
- ✓ SDC should implement **M&E focal points** (in HQ, regional, sectoral).

- ✓ For establishing an adequate terminology, E+C and QS should jointly develop a **user guidance on different types of evaluation, criteria for defining an evaluation** and defining what is not an evaluation (study, assessment, analysis, etc.).

Above all, the authors of this report also suggest a more systemic recommendation to better meet the structural constraints as found in this Quality Assessment:

- ✓ E+C jointly with QS should **adapt SDC's evaluation policy and introduce a coherent evaluation system** focusing on leveraging the utility of evaluations for different learning purposes:
 1. **Decentralized evaluations** under the responsibility of the project management, supported by QS:
 - The evaluations should be obligatory at the end of a phase before a new phase starts.
 - The main objectives would be accountability to show what has been reached so far, and the use of the lesson learnt for the planning of the second phase.
 - The evaluations should be based on a standardized format so that evaluation reports are more comparable.
 2. Every two years a sample of **centralized external independent final/ex-post evaluations** of projects or under the responsibility of E+C:
 - The main objectives would be to receive more evidence on effectivity, impact, and sustainability.
 3. For both, decentralized and centralized evaluations, an **inception phase** should be applied whenever suitable.
 4. A **Quality Assessment of decentralized evaluation reports should be carried out** every 3 years under the responsibility of E+C for improving the quality of the evaluations.
 5. An **evaluation synthesis of all evaluations according to sectors/topics** should be conducted under the responsibility of E+C every 2 years.
 6. **Impact evaluations** for selected projects or programs of high importance for SDC, applying new approaches, or with intended high-level impacts under the responsibility of E+C:
 - The main objective would be to establish prerequisites right at the beginning of an intervention to later adequately measure impact.
 - Respective interventions should be supported in the set-up of their impact orientation from the beginning (Theory of Change, design, monitoring system, target and control groups, etc.), as well as in the data collection and analysis.
 - E+C could commission teams of external evaluators for setting up, accompanying and conducting the impact evaluations.
 7. A comprehensive **capitalization strategy** comprising all evaluations.
- ✓ In order to be able to eventually implement the above recommendation with systemic reach, the **evaluation function** within SDC should be strengthened. This implies the recognition of evaluation as important source for decision-making by the leadership, the extension of E+C's mandate and the provision of more evaluation staff to E+C.

1. Introduction

The present report contains the key results from the Quality Assessment of 60 decentralised SDC evaluations. Between 2018 and 2020 Swiss Development Cooperation (SDC) mandated more than 150 decentralised evaluations of which almost two thirds (92) were used to calculate the efficiency rate of SDC. In order to assess the quality of the evaluations, evaluation reports and their Terms of Reference (ToR), a sample of 60 evaluation reports, were scrutinized using document analysis. In addition, the utility of the evaluations for SDC staff was investigated through a full-population online survey and interviews with ten selected survey participants.

The present report starts with a brief description of the evaluation's background (Section 1.1), object and scope (Section 1.2) and the objectives (Section 1.3), followed by summary of its methodology (Chapter 2). Chapter 3 summarises the key results of the documents analysis, whereas Chapter 4 focusses on use and usability of SDC's decentralized evaluations, and Chapter 5 on support needs and the quality of provided support. The report ends with a brief outline of the terminology of SDC's decentralized evaluations and criteria to be applied thereto (Chapter 6). In the annexes, a list of the analysed evaluations can be found as well as further documents consulted, a list of the interviewed persons, the interview guideline, further tables, a detailed outline of the methodology, and finally more detailed findings and figures for each of the assessed criteria.

1.1 Background

The Evaluation and Corporate Controlling Division (E+C) of the Swiss Agency for Development Cooperation (SDC) oversees SDCs comprehensive thematic and institutional evaluations. The operational units are responsible for decentralised evaluations. The Quality Assurance Division (QA) provides support for planning and managing decentralised evaluations. In 2020 SDC's evaluation function was subject to an OECD-DAC Peer Review. Findings included a potential threat to credibility coming from the fact that E+C uses decentralized evaluations to calculate a performance rate for SDC's work and often uses the results of decentralized evaluations in its own studies but has no control over those evaluations. This means that it cannot vouch for the quality of that work, while using the findings reported. QA does have a role in this regard and has organized trainings and manuals to promote high-quality evaluations. However, with 80-100 decentralized evaluations being conducted each year, there is no mechanism to assure their quality.

The reviewers recommended to conduct a review of the quality of decentralized evaluations as part of E+C work program on a regular basis (e.g. every three to five years). The advantage of this approach is that it is much less resource-intensive than doing validations of each decentralized evaluation and could be used to highlight for the Management any quality issues with that work. Presumably this would lead to efforts to improve on any deficiencies (OECD DAC Evalnet Peer Review 2020). Therefore, E+C commissioned the CEval GmbH and the Center for Continuing Education at the University of Bern to conduct this Quality Assessment of decentralized evaluations between September 2021 and February 2022.

1.2 Object and scope of the Quality Assessment

Two major categories of evaluations exist within SDC (see SDC's [Evaluation Policy](#) for specifications): (1) commissioned by the SDC Directorate and managed by the Evaluation and Corporate Controlling Division (E+C) and (2) decentralised evaluations commissioned by Swiss representations or the operational units of the four SDC departments, and managed by operational units of the four SDC departments in the Swiss representations or at HQ. Support for planning and managing the decentralised evaluations is provided by the Quality Assurance (QA) Section.

The object of this Quality Assessment are decentralised evaluations which can comprise programme, project, cooperation programme, as well as impact evaluations. All types of evaluations are possible: external, internal, hybrid, self-(reviews), and Joint evaluations.

SDC does not impose a strict rule on when to conduct a decentralized evaluation. The time should depend on the context and the purpose for which the results will be used. "Evaluations should be commissioned whenever an (external) assessment for learning, steering and accountability purposes is desirable or necessary. Ideally, projects and programmes should be evaluated every four to eight years, which corresponds to every one or two phases" (How-to Note Evaluation, 2021, p.7).

Between 2018 and 2020 SDC mandated more than 150 decentralised evaluations of which almost two thirds (92) were used to calculate the efficiency rate of SDC. The efficiency rates look at the quality of the projects and programs based on evaluation results. It is calculated to provide a comparable benchmark for the efficiency of SDC over time.

Subject to the Quality Assessment is a total of 60 project and program evaluations (from hereon called interventions), consisting of

- a) a random sample of 48 evaluation reports from 92 decentralised, external mid-term and final evaluations of SDC, implemented between 2018 and 2020 and used for the calculation of the efficiency rate and
- b) a sample of 12 deliberately selected reports that are not used by SDC for calculating the efficiency rate.

External evaluation means that the evaluators are not involved in the programme or project or working for the FDFA. Midterm evaluations are performed towards the middle of the period of implementation of the intervention, and final evaluations at the end of an implementation phase. (How-to Note, 2021)

1.3 Objectives of the Quality Assessment

SDC`s E+C Section commissioned this Quality Assessment to assess the validity and reliability of the results of the selected SDC`s decentralised evaluations and their actual use. The Quality Assessment focuses on the methodological quality and on the utility of decentralized SDC evaluations in order to:

1. assess the quality of the evaluations and create a baseline of the current quality,
2. assess how SDC staff in charge of decentralized evaluations are using the support (toolkits, PCM training, etc.) currently available and how it can be improved
3. assess the support needs of SDC staff for managing evaluations and the quality of the support, and
4. assess the terminology of the evaluations and if necessary, propose new criteria and terminology.

In this regard, the Quality Assessment should contribute to revealing the strengths and weaknesses of current decentralised evaluation practice and support its improvement with particular focus on the added value of its results.

2. Methodology

The evaluations were assessed according to the following analysis dimensions, which follow the typical structure of an evaluation report, including its ToR:

1. Terms of Reference
2. Executive summary
3. Introductions and context analyses
4. Evaluation methodology
5. Evaluation findings
6. Conclusions and recommendations

For each dimension, a set of criteria was developed (see Appendix 2) according to which the evaluation reports were assessed. For assessing the use and usability of the evaluation by SDC staff, further criteria were developed including: (1) its process use, (2) the actionability and usability of its recommendations, as well as (3) the actual use of its results and lessons learnt. Finally, in order to assess the support needs of SDC staff, a set of questionnaire items was developed together with E+C (see Appendix 1).

The results of the Quality Assessment are based on two data sources: the evaluation reports, partially including their ToR, and SDC staff. While the reports underwent a qualitative document analysis, data from SDC staff were gathered through an online survey and additional interviews with selected survey participants.

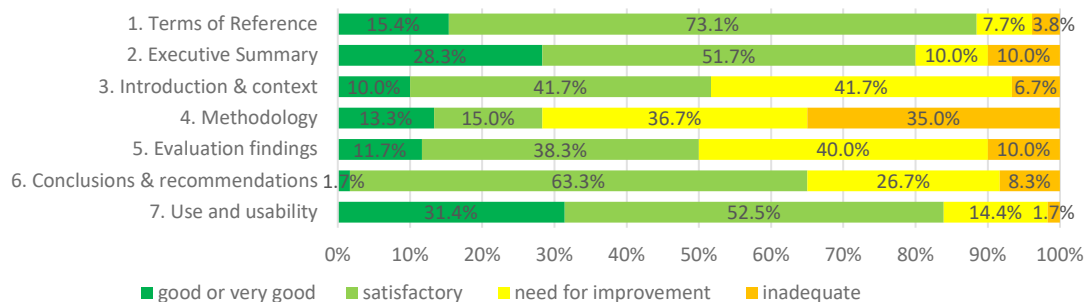
Further information about the methodology of the data collection and analysis can be found in annex 7.6.

3. Key results from the Quality Assessment of the evaluation reports

The presentation of the findings of the Quality Assessment follows the structure of the analysis protocol outlined in the method chapter, from the assessment of the ToR to the conclusions and recommendations (Sections 3.1 to 3.6). In the following, only the key results and recommendations are summarised. More detailed information on the assessment of the individual criteria can be found in annex 7.7 along with some figures illustrating the quantitative results.

The following figure shows the overall assessment of each criterion at a glance, whereby the scale is categorized as follows: $\leq 0,25 \rightarrow$ inadequate (orange); $> 0,25$ to $\leq 0,5 \rightarrow$ need for improvement (yellow); $> 0,5$ to $\leq 0,75 \rightarrow$ satisfactory (light green) and $> 0,75 \rightarrow$ good or very good (dark green).

Figure 1: Evaluation report assessment summary (for 1. N=26, for 2.-7. N=60)²



² Only for 26 of the 60 evaluations ToR were available in the annex of the reports.

3.1 Terms of Reference

The quality of the evaluations' Terms of Reference (ToR) is assessed according to eight criteria including the adequacy of the **context description**, the description of the **evaluation object**, its **rationale and purpose** of the evaluation, its **scope**, the **evaluation criteria/questions**, the **evaluation methodology**, the **responsibilities, deliverables and schedule** and finally, the **feasibility** of the evaluation.

Key results

At large, most ToR are appropriate and complete. With almost nine out of ten being of at least satisfactory quality, they score best among the six assessment categories of the document analysis. In contrast, only one ToR is deemed to be inadequate.

Taking a closer look at the individual criteria according to which the ToR are assessed, it turns out that the overall good rating of the ToR is explained by three main criteria which are rated exceptionally well: the adequate **inclusion of evaluation criteria and questions**, the **description of the evaluation object** and the **description of the responsibilities, deliverables and schedule**. All ToR manage to present evaluation questions and/ or criteria sufficiently well. Plus, most of the ToR also describe the evaluation object and the context of the evaluation in a satisfactory way. In contrast, the **feasibility of the evaluation** as stipulated in the ToR, the **descriptions of its rationale and purpose**, as well as of its **scope**, are assessed rather weak. Moreover, most ToR fail to describe adequately the **evaluation methodology**.

Recommendations

Overall, there is little room for improvement as regards the ToR. Nevertheless, there are some aspects that might be considered in the future in order to provide even better guidance to the evaluators:

- ✓ Provide a more realistic budget and timeframe for the evaluation, taking into account the effort required to answer all evaluation questions. While the How-to Note highlights the necessity of budgeting an evaluation (see Steps 8 and 12), it is not specified what in particular needs to be taken into account (e.g., time for development of an evaluation matrix, data collection instruments, data collection and analysis) in order to derive a realistic budget.
- ✓ Include information about the reasons why the evaluation is implemented or requested, as well as its intended use and users. This aspect is covered in the How-to Note on page 6 but only with regard to the overall purpose such as learning, steering and accountability. However, it can also make sense to outline the specific rationale for the evaluation (e.g., end of funding period, planning of a follow-on project) in order to get a better understanding of the evaluation results.
- ✓ Make sure that the scope of the evaluation is outlined well, e.g., by including information about the period of the intervention's implementation and its geographic areas that should be covered by the evaluation. The How-to Note actually mentions the necessity of scoping the evaluation object repeatedly (e.g., on page 12 under step 2).
- ✓ As a standard, the ToR should stipulate the combination of qualitative and quantitative methods as well as data triangulation. This aspect is apparently not covered in the How-to Note at all.

3.2 Executive Summary

The quality of the evaluations' executive summaries is assessed according to two criteria, the **completeness of the executive summaries' content** and their **writing style**.

Key results

On average, the executive summaries are of good quality. Nevertheless, a closer look at the two criteria reveals a somewhat more differentiated picture. While the executive

summaries' writing style is mostly rated good or very good, only less than 4 out of ten can be regarded as sufficiently complete, with the evaluation design and the lessons learned missing most often.

Recommendations

As regards the completeness of the evaluations' executive summaries, we recommend the following:

- ✓ Advise the evaluators to include information about the rationale of the evaluation, its design and methodology as well as a summary of the lessons learned. This information will improve the comprehensibility of the report for 'external' readers who were not involved in the implementation of the project or the evaluation.
- ✓ Make clear that the executive summary should be a 'stand-alone' document, which has to be comprehensible without the need to look into the report or its annexes.

3.3 Introduction and Context Analysis

The quality of the evaluations' introduction and context chapters is assessed according to three criteria: the adequacy of the description of the **rationale and purpose** of the evaluation, of the **context analysis** provided and of the description of the **evaluation object**.

Key results

The introduction and context assessment shows a slightly different picture to the two previous analysis dimensions. Overall, it scores as the second lowest mean among the six assessment categories (but still better than the evaluations' methodology). Indeed, the ratings are quite distributed unevenly between the different criteria, with 'context analysis' criterion as the weakest, as it oftentimes does not clearly relate enough to the intervention. Overall, only about half of the introductions can be regarded as being of sufficient quality, while the other half features some room for improvement. Thereby, it has to be highlighted that a considerable share of reports lacks a sufficiently outlined ToC/results model of the evaluation object, which would enable the reader to contextualize the evaluation findings.

Recommendations

In order to improve the evaluations' introductions in terms of their completeness, the following recommendations can be given:

- ✓ Request the evaluators not only to refer to conditions on site when outlining the intervention context but also to consider international, national, and regional strategies and policies, as well as the Swiss Development Policy in particular.
- ✓ Emphasize the necessity to connect the context analysis to the intervention's goals and activities, instead of providing an expert assessment of the regional conditions at large.
- ✓ Check introductions for the completeness of the evaluation object's description, including its budget, implementation area(s), target groups and further stakeholders, relevant implementation arrangements (partner structure, task division, etc.). Particularly, request a clear and concise results model.

3.4 Methodology

The quality of the evaluations' methodologies as reflected in the methodology chapter of the evaluation reports is assessed according to six criteria: the adequacy of the **evaluation design**, the **sources of evidence**, the **data collection**, the **sampling**, the **data analysis**, and the **limitations and challenges**.

Key results

Scoring lowest among the six assessment categories, the methodology chapters leave some room for improvement as regards their quality and comprehensiveness. Only about 13% of the evaluation reports contain good or very good methodology sections. The finding is to be interpreted as critical since therewith also the validity and reliability of the evaluation

results has to be questioned. This assumed relationship is confirmed by the statistical analysis (see Table 1 in Annex 7.5), which shows a highly significant correlation between the assessment of the methodology and the assessment of the quality of the evaluation findings ($r = 0,545$, $p = 0,001$).

Recommendations

From our point of view, the evaluations' method chapters could be improved by considering the following recommendations:

- ✓ Request evaluators to be explicit about the evaluation design and methodology, including a justification of the sampling strategy, and a description of the data collection instruments and analysis methods. Clarify, what is to be understood under an evaluation design (in contrast to an evaluation approach or methodology).
- ✓ Scrutinize the evaluation design in view of its adequacy to the evaluation object and the purpose of the evaluation.
- ✓ Make the presentation of an evaluation matrix (analysis grid/protocol) in the annex of every evaluation report mandatory and provide a template for that. Also stipulate the provision of the final data collection instruments.
- ✓ Give clear advice how to document data sources not only in the method section but also during the discussion of the evaluation findings.
- ✓ Ask evaluators to discuss the limitations and challenges of the evaluation and their consequences on the validity and reliability of its findings.

3.5 Evaluation Findings

The quality of the evaluations' findings is assessed according to seven criteria comprising the adequacy of the discussion of the six OECD/DAC criteria **Relevance**, **Coherence**, **Effectiveness**, **Impact**, **Efficiency** and **Sustainability** in the reports, as well as the overall **strength of evidence** upon which the findings are based.

Key results

The assessment of the quality of the evaluation findings provides a quite heterogeneous picture. The ratings are distributed quite unevenly among the different criteria, with the discussion of the OECD/DAC criterion Effectiveness on average being assessed best and Coherence worst. However, in order to do justice, it has to be mentioned that as regards the latter, the majority of the evaluations under investigation was conducted before the revised set of criteria was published in December 2019 and thus, it is only referred to in about half of the cases. From the remaining criteria, impact is the criterion the evaluators struggled with the most. Thereby, a particular weakness can be found in unintended impacts being oftentimes not discussed.

Interesting to note is another very plausible finding from the statistical analysis (see Table 1 in Annex 7.5): the rating of the quality of the evaluation findings highly correlates with the rating of the quality of the thereof derived conclusions and recommendations ($r = 0,535$, $p = 0,001$). While this does not pose a major conclusion for the analysis, after all it speaks for its consistency, as the latter is based on the former.

Recommendations

For strengthening the documentation of findings in SDC's decentralized evaluations, the following recommendations can be made:

- ✓ Request evaluators to include references to data sources when discussing the findings, to make an effort to decide if an observed change can be attributed to the intervention under review, and to clearly separate findings from conclusions and recommendations.

- ✓ If the evaluation should follow the OECD/DAC criteria, request the evaluators to get acquainted with their definitions and according topics to be dealt with under the respective headings. Based on the findings from the document review, the following aspects should be given more attention:
 - Relevance: intervention's consistency with international and regional policies as well as with SDC's development cooperation policy
 - Effectiveness: intervention's contribution to observed changes at outcome level; disaggregation of findings by (target) groups
 - Impact: intervention's contribution to observed changes at impact level; unintended impact
 - Efficiency: production and allocation efficiency (i.e. transformation of inputs into outputs and outcomes); cost-benefit analysis
 - Sustainability: partner and target groups' capacities; intervention's strategies to mitigate risks to sustainability; including all three sustainability dimensions in the assessment

3.6 Evaluation Conclusions and Recommendations

The quality of the evaluations' conclusions and recommendations is assessed according to two criteria comprising their **reference to findings** and the **quality of recommendations**, as outlined in the respective chapters of the evaluation reports.

Key results

All evaluation reports contain conclusions and recommendations. Whereas these conclusions and recommendations mostly refer to the findings of the evaluation, their quality in terms of prioritization, time-boundness and actionability shows some room for improvement. This finding is quite critical insofar as usually conclusions and recommendations represent the most important parts of the report to its readers. Also, only about half of the reports provide lessons learned.

Recommendations

In order to strengthen the evaluations' conclusions and recommendations sections, we suggest the following:

- ✓ Advise evaluators to prioritize recommendations and provide guidance on who should do what and when.
- ✓ Make it mandatory for all evaluation reports to include a lessons learned section.

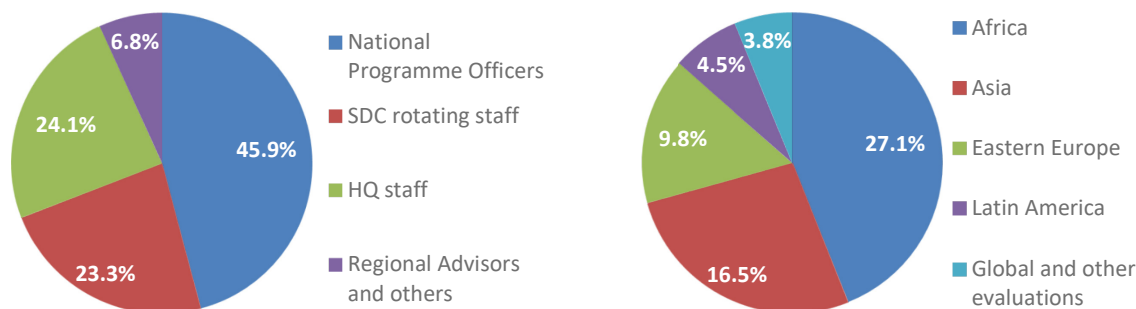
4. Use and usability of decentralized SDC project and program evaluations

The use and usability of decentralized evaluations is assessed according to three criteria: the **implementation and process use of the evaluations**, the **perceived usability of their findings and recommendations**, and their **actual use in practice**.

In contrast to the previous chapters, conclusions drawn in this and the following chapter are based on the results of a **semi-standardized full-population survey** and **interviews** with ten selected survey participants. The purpose of the combination of these two empirical data collection instruments was to achieve both statistical and contextual representativeness of SDC staff in terms of their position and regional distribution. While the survey yielded primarily quantitative data, the interviews provided further insights about experiences with the management of evaluations as well as motivations and justifications of attitudes towards particularly critical issues (e.g., the factual use of recommendations) and specific support needs.

The following figure shows the functions and regional distributions of the survey respondents:

Figure 2: Functions and regional distribution of respondents



4.1 Implementation and Process Use

For assessing the implementation of the evaluations and their process use, SDC staff was asked in the online survey and the interviews how challenging they perceive the **preparation of the evaluation**, how they assess their **collaboration with the evaluators** and the **evaluators' competencies** as well as how they rate the **adequacy of their own effort** for the evaluation.

Preparation of the evaluation

As it is the responsibility of the project staff in the country or regional offices to schedule, plan and prepare the decentralized evaluations, it can be positively stated that the tasks associated with the preparation of an evaluation are reported to be handled well by the survey respondents. These tasks include defining the deliverables of an evaluation, preparing reference documents, and specifying the budget or the process and timeline. Nevertheless, in accordance with the findings from the analysis which show especially severe shortages of the ToR, most challenging is the quality assurance of the Approach Paper or Terms of Reference, specifying the methodology and defining the evaluation questions. See Figure 10 in the annex.

Also finding suitable evaluators seems to be a major challenge. To address this challenge, the project management tends to work with evaluators they already know which might have implications regarding the impartiality and independence of the evaluators. Although SDC procurement rules clearly define different procurement formats depending on the size of the evaluation, it seems that procurement is sometimes avoided by the project management due to the additional efforts needed.

Apparently, the indicated main challenges of recruiting evaluators and writing ToR led to a loss of efficiency during the preparation phase of the evaluations and had implications for the whole evaluation process and the quality of the evaluation reports. Especially the threat to independence of the evaluators has to be taken very seriously, which will be further elaborated on in the following chapters.

Collaboration with the evaluators and evaluators' competencies

The severe methodological deficiencies of the evaluation reports as found in the analysis cannot be attributed to the findings of the survey regarding the competencies of the evaluators. Furthermore, particularly, the evaluators' transparency, sectoral and regional competencies are appreciated by the majority of survey respondents. What again has to be noted critically are some deficits regarding the evaluator's impartiality (see Figure 11 in the annex). Interestingly, the majority of the respondents of the interviews emphasize the lack of specific competencies of the evaluators they have worked with especially regarding their analytical skills, sectorial expertise and regional experience which rather confirms the findings from the analysis.

The evaluators' availability, responsiveness, as well as their communication and management of the evaluation are assessed positively which speaks for a good collaboration between evaluators and those commissioning and managing the decentralized evaluations. Only the evaluator's flexibility and the assistance provided to them show some deficiencies. See Figure 12 in the annex.

The comparatively positive assessment of the evaluators' methodological competencies by SDC staff in the survey somewhat contradicts the experts' assessment of the analysis. All in all, the findings of the evaluator's competencies are not fully conclusive, which (at least partially) may also be ascribed to the sample of surveyed and interviewed individuals. It might also be that in some cases, survey respondents are not fully capable of assessing the evaluation methodology as this requires respective in-depth knowledge.

SDC's staff own efforts

SDC project management needs to invest quite some time in preparing, commissioning and managing the decentralized evaluations as it is solely their responsibility when and which kind of evaluation they want to conduct. Since it is only one of many tasks project staff must fulfill, it is important to know if this is too burdensome. It can be positively stated that SDC's staff own support for the evaluation in terms of time invested, logistical support as well as technical and methodological support provided seems to be adequate (see Figure 13 in the annex), although the required time is often underestimated. Not surprisingly, the tasks for which most efforts are dedicated for the evaluation - writing/reviewing ToR, selecting evaluators, providing backstopping and quality assurance – coincide with the stated main challenges during preparation. Insufficient allocated time could limit the quality of the ToR, negatively impact the whole evaluation process and finally, the potential of evaluations as an instrument for evidence-based decision-making.

Recommendations for improving the implementation and process use of decentralized evaluations

- ✓ **QS and E+C should jointly develop and implement measures to develop evaluation capacities of SDC staff:** Although, M&E is part of the overall PCM and SDC staff is required to do the online PCMi training course, more efforts should be laid on training of fundamentals of evaluation to gain knowledge about types of evaluations, evaluation planning and management, quality assurance, and follow-up. While it may not be required to convey in-depth methodological skills, SDC staff involved in setting up and managing decentralized evaluations should be enabled to write good ToR and distinguish a good from a bad evaluation. This is most urgent for new staff without any experience in evaluation.
- ✓ **QS should develop and implement measures for continuously strengthening evaluation capacities of SDC staff:** SDC staff should be sensitized to keep up and further strengthen their evaluation capacities, for instance by means of continuous refresher courses, which could be provided through short online formats on specific topics (e.g., every 2 years).
- ✓ **QS should regularly inform about available sources of information:** As many respondents claim not to be fully aware of already available material and guidelines related to evaluation, QS should make sure that the country offices or SDC project management directly receive these materials and be reminded on a regular basis. This could be done by sending once or twice a year an "Evaluation Brief/Evaluation Flash" including links to relevant material, training possibilities (internal and external; onsite and online) and news concerning evaluation within SDC.
- ✓ **Systematic maintenance of evaluator rosters.**
- ✓ **The project management should put more efforts in selecting adequately qualified evaluators:** Project managers commissioning decentralized evaluations should set up clear criteria particularly for assessing the methodological qualification, sectorial

expertise, and regional experiences of the evaluators to be commissioned for a decentralized evaluation. Although procurement processes might be extensive, the project management should deliberately consider calls for tenders to find suitable evaluators. This would also be a measure to avoid commissioning the same evaluators over and over again to prevent lack of independence and impartiality of the evaluators.

- ✓ **QS should provide the project management with guidelines with selection criteria for evaluators and advise them in assessing the quality of a proposal adequately.**
- ✓ **The project management needs to plan sufficient time for preparing, commissioning, managing, and assuring quality of their evaluations, and for translating recommendations into action.**
- ✓ **Promotion of Inception Reports:** As the evaluation practice of other donor organizations clearly shows, Inception Reports highly contribute to a better quality of evaluations. Although it might not be advisable to instruct the project management to implement IR as a mandatory deliverable due to different purposes of decentralized evaluations, the advantages of this tool should clearly be communicated. The purpose of Inception Reports, respective templates and explanations are already included in the How-to-Note Evaluation, but not widely recognized or used yet. QS should therefore include the topic in trainings and further promote whenever possible. The evaluation should be divided into an inception phase and a 'main' evaluation phase. Therewith, the scope for adapting evaluation designs and methodologies can be maintained. In a worst-case scenario, consider terminating the evaluation early if the suggested design does not fulfill minimum quality criteria.

4.2 Usability of Decentralized Evaluations

For assessing the usability of evaluations, SDC staff was asked about its **satisfaction with the evaluation report**, the **actionability of recommendations**, their **usefulness for particular purposes** and the overall **added value of the evaluation**.

Satisfaction with the evaluation report

Taking into consideration the rather critical assessment of the analysis, it is surprising that most survey respondents are quite satisfied with the evaluation reports they received. Only the length of the evaluation report, the applicability and the foundation of recommendations were rated somewhat less positive (see Figure 14 in the annex). These findings clearly contradict the rather mediocre expert rating of the decentralized evaluation reports, as well as the assessment of most interview respondents. What is worth mentioning is the fact that although the analysis shows severe shortcomings of the reports, those commissioning the evaluation consider the evaluation report as useful for their office. Some explanations in that regard can be drawn from the interviews: Evaluators might tend to go the easy way and collect data selectively, accordingly, evaluation reports contain information that is already known by the project staff. The project management and other stakeholders might at times influence the formulation of recommendations, which goes a long with the above stated possible implications of lack of impartiality and independence of the evaluators. It can thus be concluded once more that there is an urgent need of Evaluation Capacity Development among SDC staff.

Actionability and usefulness of recommendations

Due to the rather critical analysis assessments of the quality of the recommendations in the decentralized evaluation reports, it could be concluded that the recommendations are not perceived as very useful and their transferability to action is rather limited. Although the interviews confirm this assumption, the vast majority of the survey respondents do not at all. In accordance with their positive assessment of the whole report, they see the recom-

mendations as valid and usable, as well as the elaboration and alignment of recommendations to particular actors. Only the prioritization of recommendations leaves room for improvement. See Figure 15 in the annex.

Yet, what is evident and important to note is that the perceived usefulness of recommendations for particular purposes is limited to two main aspects: First, for planning new interventions and improving intervention approaches as well as developing an already running intervention further, and second, for increasing the acceptance or legitimacy of foreseen changes. Improving SDC strategies or increasing aid effectiveness at large does not seem to play a major role (see Figure 15 in the annex). Not surprisingly, the actual use of the recommendations as explained further below, further confirms this perspective.

Added value of the evaluation

A likewise heterogeneous picture yields the question about the added value of the evaluation at large. Survey respondents see the highest added value of an evaluation in providing transparency about the project's achievements. Learning and use for other projects does seem to fall short. All in all, the usability of evaluations is recognized, although to a different extent regarding generating knowledge, enabling organizational learning, providing for accountability, as well as providing the basis for decision-making (see Figure 17 in the annex).

Recommendations for the usability of decentralized evaluations

- ✓ See the above recommendations on more efforts selecting adequately qualified evaluators and develop and further strengthen evaluation capacities of SDC staff
- ✓ **Already in the planning stage, the project management should elaborate on the purpose and objectives of the evaluation for different stakeholders.** This requires thinking about who else could have an interest in the findings and consult them beforehand. Guiding question: "Who is using the evaluation results for what?" Partners, other stakeholders in the country or region, other projects, SDC country office or HQ, etc. In that sense, the report can have more value added, the recommendations can be better differentiated, and the use of the evaluation can be broadened.

4.3 Use of Evaluation Results

This section focuses on the use of evaluation results, starting with SDC staff's feedback about **what they use recommendations for** and which **further evaluation outcomes** are relevant to them. Furthermore, the **limiting factors** are discussed as well as the **options available to facilitate the use of evaluations**.

Use of recommendations

SDC project management is not only responsible for setting up and managing decentralized evaluations, but it also falls under their responsibility to use the results by transferring recommendations into concrete action. The analysis attests shortcomings in the formulation of recommendations, so it could be assumed that they are not very user-friendly. Nevertheless, as stated above, the ones commissioning the evaluation declare the usability of the recommendations. Looking at the actual use of the evaluations' recommendations, the survey and the interviews provide a clear picture, which is also in line with the *perceived usability* as discussed in the previous section: Recommendations are mainly used for planning new interventions and improving intervention approaches, further development of a current intervention, and preparing a formal Management Response. Therefore, new topics and ideas (e.g., for activities, products), good practices and lessons learnt from the evaluated project, as well as suggestions on project management and steering are most appreciated. In contrast, for legitimizing change processes, disseminating the evaluation report and improving SDC strategies recommendations appear to be somewhat less important (see Figure 18 in the annex).

Once again it can be concluded that the actual use of recommendations is rather limited if the focus remains on deciding how to continue with a project, particularly when entering a new phase. If the report is not even distributed nor discussed beyond the project staff and a Management Response is not drafted, it is very likely that the potential of evaluations for institutional learning, decision-making beyond the perspective of planning a new phase, and accountability is not fully exploited.

Relevant evaluation outcomes

Adding to the above-mentioned uses of recommendations derived from decentralized evaluations, lessons learnt are by far the most relevant outcomes for the survey respondents followed by providing legitimacy, for instance pointing out in which areas the project was particularly successful, showing beneficiaries' perspective, confirming goal achievement, or providing information about expectable sustainability (see Figure 19 in the annex).

Real changes due to evaluations seem to be limited to the intervention itself: adaptations of the interventions, changes in attitudes, awareness and ownership of intervention staff, partners and stakeholders, and increased interest by other donors to collaborate and replicate activities. Changes with a broader impact are missing and improving SDC strategies are apparently realized only rarely.

Limiting factors for the use of evaluation results

According to the survey respondents, the use of evaluation results, including the implementation of their recommendations, is hampered by shortcomings of the recommendations, incompleteness of analysis, impartiality of evaluators, as well as resistance of stakeholders to adopt changes and unsupportive regulations. Interesting to note is that the survey respondents do not assess recommendations per se as insufficient when asked about the quality of the report (see above), but when directly related to their use, they come to a different conclusion. Same refers to the impartiality of evaluators as stated several times above.

What is striking is that the interviews yield another perspective on impeding factors, which are not so much related to the evaluations themselves but rather conceived as systemic issues rooted in SDC's organizational culture. They refer to fear of failure, lacking interest in learning, and missing commitment by the leadership to use evaluation as a tool for institutional learning and development. Interestingly, the latter was also mentioned by the survey respondents. These impressions suggest that evaluation is insufficiently anchored within the organization, and the organization culture impede a strengthened institutionalization of evaluation and evaluative thinking within the organization.

Facilitation of use

Taking the above-mentioned factors for limiting the use of evaluation into account, it is not astonishing that change in attitude, and usability as the focus in each evaluation stage are seen as most supportive factors for enabling use of evaluations. Furthermore, both, survey as well as interview respondents suggest the establishment of a learning and knowledge sharing platform to foster the use of evaluation. These suggestions mirror to a certain extent the required support as discussed in chapter 5. Interestingly, once again, the survey respondents demand a fundamental change in SDC's attitude towards evaluation, as its current state limits the use of evaluation results. Interviewees further emphasize the necessity of strengthening Management Responses and the capitalization of evaluation results.

Ultimately, this should lead to an organization-wide perspective on decentralized evaluations which is designed to generate information that helps the project management to make better decisions, be accountable, and support SDC's ongoing development as a learning organization.

Recommendations on the use of evaluation results

- ✓ **SDC should introduce a capitalization strategy for decentralized evaluations:** In order to provide for a supportive organizational environment, the general attitude towards learning (not only from evaluations) and further strategic development is suggested to be further developed towards an “evaluative thinking”³ in the organization. This can be done by introducing a capitalization strategy including various formats on lesson learnt from decentralized evaluations across SDC’s departments (HQ and in the field) and with partner organizations. The respective focus could vary, like discussing the results and what can be learned for different purposes or showing good practices (e.g., how evaluation results were used to improve aid effectiveness). Of further value for SDC could be a more intensified exchange with SECO on specific evaluation results (in a region or in a sector), presenting and discussing evaluations with the Swiss Evaluation Society, other donor organizations via the OECD-DAC EvalNet (e.g., through joint evaluations) or via more active engagement within GEI, on international platforms like 3IE, IPDET, the European Evaluation Society and other regional and national evaluation associations.
- ✓ **E+C jointly with QS should develop an organization-wide evaluation strategy:** This strategy should include timelines and milestones outlining what SDC wants to achieve with evaluations within a given period (e.g., the next five years) and anchored in the current reform process. This would include all kinds of evaluations conducted on behalf of SDC, not only the decentralized ones. Based on the already existing workplan of E+C, such a strategy should include:
 - tangible results and indicators like evaluation quality improvement by X percent,
 - the establishment of an online results-based monitoring framework for SDC projects and programs, which is being considered under the Results Data Management Project,
 - the implementation of a sector- and/or region-wide evaluation synthesis which aggregates results from different evaluations, so that overarching conclusions can be deduced,
 - as well as the standardized analysis of decentralized evaluations.
- ✓ **QS should develop an obligatory reference framework for the implementation of decentralized evaluations and the utilization of their results:** Clear criteria should be set out for deciding on when to conduct a decentralized evaluation, for what purpose and which type of evaluation to select. Furthermore, the framework should specify who (including top management) has to be involved at which stage of the evaluation, how project/ program management should proceed during the drafting of the Management Response, and when planning follow-up measures. Eventually, such follow-up measures should be closely monitored by QS and documented for later verification of the implementation of recommendations and their effectiveness.
- ✓ **The project management should emphasize the Management Response** of their decentralized evaluations and use it for discussions with partners, other stakeholders as well as the respective HQ staff. The latter should value decentralized evaluations for learning and accountability and make themselves available for discussing the Management Response and supporting the follow-up.

³ Evaluative thinking is defined as critical thinking applied in the context of evaluation, motivated by an attitude of inquisitiveness and a belief in the value of evidence, that involves identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective taking, and informing decisions in preparation for action (Buckley, J. et al., 2015, Defining and teaching evaluative thinking. American Journal of Evaluation 36, 3). It can be seen as is a constant state-of-mind within an organization’s culture and all its systems.

5. Support needs and quality of support

The focus of the Quality Assessment is also to assess the adequacy of the current support provided to SDC staff for planning and managing decentralized evaluations and to provide recommendations for improving it. Therefore, the **received support** and its perceived quality as well as **further needed support measures** are investigated. Once again, the drawn conclusions are based on the perceptions and information by the survey and the interview respondents.

Received support

Since the project management is solely responsible for planning and managing decentralized evaluations and only rarely have staff with evaluation expertise, available support plays a key role in implementing quality evaluations. QS already provides a variety of support measures for setting up and managing decentralized evaluations, like guidelines and trainings; the Field Handbook contains comprehensive information as well. The question is, if those measures are well-known and sufficient. Of course, the quality of an evaluation highly depends on the competencies of the evaluators, conducting the evaluation. However, as concluded various times in this report, the ones commissioning and managing the evaluation lay the ground for a successful implementation. They decide on the kind of evaluation, the selection of the evaluators, the ToR, the time allocated, and the process. Furthermore, the commitment by the management, partners, and other stakeholders plays a crucial role.

Obviously, there is a lot on the project management's shoulders. As reported by the survey respondents, they mainly use support in form of written guidance, followed by personal counselling. The Field Handbook as well as other publicly available guidelines and papers are well known, appreciated and widely used, in contrast to specific guidelines and tools, like the How-to Note Evaluation, which apparently is lacking in dissemination. As regards to personal support, it becomes evident that superiors and colleagues play a very crucial advisory role that their personal guidance is even seen as more valuable than the written material. Yet, this also implicates the dependency on others and an already established network of colleagues who could be approached for advice, which limits the possibilities of new staff in the country/region.

It can be concluded that support structures are in place but not sufficiently recognized by and aligned with the needs of SDC project staff in the countries. Apparently, SDC already has a multitude of tools at hand to support its staff for planning and managing evaluations, but some of these tools are not known and/or used by all respective colleagues, key documents such as the How-to Note Evaluation are not well disseminated, and available guidelines are not being followed by evaluators. While these observations may indicate a general overburdening of SDC project staff, it also suggests that the available tools may not fully match their needs. The following section sheds some light on other possible forms of support.

Useful support measures

A wide range of additional support measures for planning and managing an evaluation can be concluded from the survey and interview results and can be read in detail in the annex, yet, none of them strikes out as THE missing factor in the system. An online helpdesk, as well as online tutorials and virtual trainings seem to be more adequate as face-to-face trainings and further written manuals and guidelines. As time is scarce, it becomes evident that practical digital support measures are very useful for project staff involved in decentralized evaluations. This also includes an online toolbox with exemplary ToR, evaluation questions, report templates and criteria for assessing the quality of an evaluation, as well as online accessible repositories for (good and bad) evaluations and evaluators by region and sector.

Recommendations on support needs

- ✓ **QS and E+C should jointly revise and update the existing tools supporting the planning and implementation of decentralized evaluations and QS should distribute them among SDC project managements:** Based on a needs assessment – but also on the basis of the findings of this Quality Assessment – the existing guidelines should be checked for gaps and revised accordingly. New tools, like the above-mentioned ones, and further reference material (e.g. an evaluators' database, repository of good practice evaluations) should be introduced and actively promoted. Such active promotion should be amplified by (online) events, audio/video tutorials and the like.
- ✓ **SDC leadership should foster regional Evaluation Capacity Development:** As part of SDC's localization and capacity development strategies and mandates, SDC country and regional offices should engage in regional Evaluation Capacity Development. For instance, by providing national evaluators opportunities to participate in trainings and include them in evaluations with distinct roles. This does not mean that SDC should finance mandates, but rather contribute to the strengthening of local evaluation capacities, may they be mandated by SDC for a decentralized evaluation at a later point of time or not. Good practices already in place within SDC could serve as role models. The Global Evaluation Initiative (GEI), where SDC is a partner, provides possibilities of cooperation and support in that regard.
- ✓ **QS should introduce an obligatory on-the-job online training program** for everyone assigned to commission an evaluation with concrete advice on the kind of evaluation, how to write a ToR, on stakeholder involvement, communicating evaluation results, conducting evaluations in fragile contexts, etc. including refresher courses (e.g., every two years).
- ✓ **SDC should establish a mentoring network or regional focal points** with more experienced SDC staff that provide case-based support to evaluations for younger colleagues.

6. Terminology

As outlined in the introduction, another task of the Quality Assessment was to assess the terminology of the evaluations and make suggestions to revise it as necessary. When looking at the sample evaluations, it is striking that overall, the term evaluation is used in a rather broad sense. Thereby, the common ground of the studies is that they provide recommendations, may it be for the further planning of a project, revising intervention concepts or even – but somewhat less – further developing development cooperation strategies. While recommendations, and preceding conclusions, are key for any evaluation, there are more exigencies to this type of applied social science than that.

In general, the purposes of an **evaluation** are to provide knowledge, to contribute to (institutional) learning, to allow control and to make implementers accountable. Taking these purposes as the starting point for further deliberations, and considering that evaluation at large is to be understood as the assessment of an object (here: a project or program) by pre-defined measurable and commonly accepted criteria (indicators) using empirical methods, its constitutive elements become clear:

1. **Analytical framework:** Any decent evaluation is guided by an analysis grid (evaluation matrix, data collection plan, or the like) that operationalizes (i.e., makes measurable) the questions to be answered and informs about the data sources, data collection instruments, analysis methods, responsibilities, timelines, etc. Such a framework should thus be a mandatory element in every SDC evaluation, to be developed and approved during the inception phase of the assignment and documented in the annex

of the evaluation report. All data collection instruments should be based on this framework and ideally the results presentation should be structured accordingly.

2. **Coherent analysis dimensions:** The results in an evaluation report should be presented according to agreed analysis dimensions. While the OECD/DAC criteria as *the* internationally accepted standard categorization for Development Cooperation may serve well in many cases, they are not mandatory as such. Depending on the primary information needs and the point in time the evaluation takes place during the course of an intervention (ex-ante, during its implementation, at the end or ex-post), focus can be put on a selection of these or it can even be deviated from that framework and another be used instead, such as the ALNAP criteria for humanitarian aid projects.
3. **Methodological rigour:** No matter what the main focus of the study is, minimum methodological standards need to be adhered to in order to justify it as an evaluation. To this end, the application of a theory-based approach and the combination of different data collection and analysis methods, preferably including qualitative and quantitative ones, should be considered as a minimum standard. It should be further considered that assessing the effectiveness and impact of an intervention requires particular designs to be applied, such as a quasi-experimental design or a regression discontinuity design. Likewise, there is a limited choice of designs suitable for efficiency assessments, such as benchmarking or cost-benefit analysis. Since the application of such more sophisticated designs require specific qualifications, these need to be checked when selecting an evaluation team.
4. **Independence:** Although in principle evaluations can be conducted internally and externally, and more recently hybrid approaches are becoming increasingly fashionable, a sound and objective assessment of any object, requires a minimum amount of distance between the ones who are responsible for it and the one who assesses it. Therefore, we suggest to only consider studies, which are implemented by independent consultants (or consultancies), i.e. who were not involved in the planning or implementation of the project, as evaluations.

Recommendation for SDC evaluation terminology

- ✓ E+C and QS should jointly develop a guide defining the terminology which is to be applied within SDC to clearly distinguish between evaluations and other kinds of reports.
- ✓ We suggest to only call studies, which fulfill the above-mentioned demands as evaluations. Any study in which one or more of these exigencies is violated should be rather called a 'review' or, if it also includes non-evaluative content such as a needs assessment, stakeholder mapping or context analysis, simply a 'study'.

7. Annexes

Content

Content.....	1
7. Annex	2
7.1 List of evaluations.....	2
7.2 Further documents consulted	4
7.3 Persons interviewed	4
7.4 Interview guideline	5
7.5 Further tables	7
7.6 Detailed outline of the methodology.....	8
7.7 Detailed analysis of each criterion of the document review.....	10
7.7.1 Terms of Reference	11
7.7.2 Executive summary	13
7.7.3 Introduction and context	13
7.7.4 Methodology.....	14
7.7.5 Evaluation findings	16
7.7.6 Evaluation conclusions and recommendations	18
7.8 Detailed analysis of the survey and interview results	19
7.8.1 Implementation and Process Use	19
7.8.2 Usability of Decentralised Evaluations	22
7.8.3 Use of Evaluation Results.....	24
7.8.4 Support Needs and Quality of Support.....	27
7.9 References to appendices	30

7. Annex

This Annex contains lists of the analysed evaluation reports and further documents consulted, as well of the persons interviewed. Furthermore, the interview guideline is presented along with additional tables with statistical calculations. Due to its size, the online questionnaire is attached to this report as a separate document (Appendix 1).

7.1 List of evaluations

Ref/ No	Name of the evaluation (decentralised evaluations)	Year	Country/ Region	Document analysis	Survey
1.0 1	Evaluation report of the Promoting Market Oriented Skills Training and Employment programme in the Great Lakes Region phase II	2018	Africa Great Lake Region	x	x
1.0 2	PROYECTO BIOCULTURA Y CAMBIO CLIMÁTICO MISIÓN DE EVALUACIÓN DE MEDIO TERMINO	2018	Bolivia	x	
1.0 3	Consultoría para realizar el diseño y la ejecución de la evaluación al Proyecto "Ampliación de la capacidad de respuesta de la Registraduría Nacional del Estado Civil en el marco de la suscripción del acuerdo final para la terminación del conflicto y la construcción de una paz estable y duradera", cofinanciado por COSUDE	2018	Colombia	x	
1.0 4	Jinsha River Basin Integrated Water Resources and Risk Management under Changing Climate Project	2018	China/Tibet	x	
1.0 5	Mid-term Evaluation of the Project Titled "Improving Land Governance in the IGAD Region"	2018	Africa (IGAD Region)	x	
1.0 6	Evaluation of the GPMD' Migration and Development Engagement on Decent Work in the Middle East, South Asia and Sri Lanka	2018	Middle East, South Asia and Sri Lanka	x	x
1.0 7	Mid-term Assessment Mission SURAFCO III Final Report	2018	Laos	x	
1.0 8	External Project Review Contribution to improved access to water and sanitation services in Bekaa Valley	2018	Lebanon	x	
1.0 9	Projet d'Appui aux Communes Urbaines du Mali (PACUM) / Phase I (1.3.2013 – 31.12.2018)	2018	Mali	x	x
1.1 0	RAPPORT D'ÉVALUATION INTERMÉDIAIRE / PAFA / PROGRAMME D'APPUI AUX FILIÈRES AGROPASTORALES DE SIKASSO	2018	Mali	x	x
1.1 1	"Fortalecimiento e Innovación académica-científica en Reducción de Riesgo ante Desastres y Adaptación al Cambio Climático en la Universidad Centroamericana (UCA), año 2013-2017"	2018	Nicaragua	x	
1.1 2	Seed and Knowledge Initiative End of Phase Evaluation	2018	Southern Africa	x	x
1.1 3	External Review of the Prevention of Domestic Violence project in Tajikistan	2018	Tajikistan	x	
1.1 4	Revue externe à mi-parcours Programme de soutien au processus électoral en Tunisie- phase 1	2018	Tunisia	x	
1.1 5	Uzbekistan External Review of SDC's Rural Water Supply and Sanitation Program Phases I – IV (2007- 2018), with Emphasis on Phase IV (2013 - 2018)	2018	Uzbekistan	x	x
1.1 6	External Review: Livestock Development in the South of Armenia Project (Phase 4); Livestock Development South- North Project (Phase 5)	2019	Armenia	x	
1.1 7	Évaluation de la phase 1 du Programme de contribution au FADeC (2016-2019)	2019	Benin	x	x
1.1 8	Evaluación externa del proyecto Vida sin Violencia (VsV) de la Cooperación Suiza en Bolivia, Fase I	2019	Bolivia	x	

1.1 9	External Evaluation Report: Project: "Strengthening the Role of Local Communities/Mjesne Zajednice in Bosnia and Herzegovina 2015–2019"	2019	Bosnia-Herzegovina	x	
1.2 0	Evaluation of the Swiss Contribution to the inclusion of Roma and other vulnerable groups "Health and Education for All"	2019	Bulgaria	x	x
1.2 1	Review CapaCITIES project, Phase 1	2019	Other	x	
1.2 2	Independent Mid-Term Review of the Low Carbon Cement Project (LCC) (Phase 2)	2019	Other	x	
1.2 3	External Evaluation Report: World Overview of Conservation Approaches and Technologies (WOCAT)	2019	Other	x	
1.2 4	Evaluación Externa del Programa Fortalecimiento de los Derechos Humanos y a las Auditorias Sociales (FDHAS)	2019	Honduras	x	x
1.2 5	EVALUATION OF THE SOMALIA RESILIENCE PROGRAM (SomRep) 2017-2019	2019	Horn of Africa	x	
1.2 6	Enhancing Youth Employment Project: final report	2019	Kosovo	x	
1.2 7	End of project external evaluation: Health Care waste Management with Hospital Acquired Infection Control project in Kyrgyz Hospitals (HCWM/HAI)	2019	Kyrgyzstan	x	
1.2 8	Appui Suisse aux Associations Professionnelles au Maroc", Phase II - Evaluation finale du projet	2019	Morocco	x	
1.2 9	Vocational Skills Development Program Myanmar - Report on the Mid-term Review	2019	Myanmar	x	x
1.3 0	Evaluation of the Swiss Contribution to the inclusion of Roma and other vulnerable groups "Reform Fund linked to the Roma Inclusion and other Vulnerable Groups"	2019	Romania	x	x
1.3 1	Mid-Term Review 2019: R4 Rural Resilience Initiative in Southern Africa (Phase II)	2019	Southern Africa	x	
1.3 2	Independent review of RisiAlbania - Phase 2	2020	Albania	x	x
1.3 3	Mid Term Review of the project "Modernisation Vocational Education and Training programme" in Croatia	2020	Croatia	x	
1.3 4	End of Project Evaluation of the project "Postharvest Management in Sub-Saharan Africa" - Final Report	2020	Mozambique/Benin	x	
1.3 5	Mid-Term Review of the Swiss Agency for Development and Cooperation Strengthening Land Governance in Myanmar Project - Final Report	2020	Myanmar	x	x
1.3 6	SDC/HA Emergency Cash Pilot For Earthquake Affected Households in Bubq, Albania - Final Report	2020	Albania	x	x
1.3 7	Mid-Term Review WaSH Support for Rohingya Refugees and Vulnerable Local Communities in Cox's Bazar District	2020	Bangladesh	x	x
1.3 8	Independent Terminal Evaluation - INCLUSIVE AND SUSTAINABLE LOCAL ECONOMIC DEVELOPMENT IN UPPER EGYPT (SOHAG) – PHASE 2 (HAYAT)	2020	Egypt	x	x
1.3 9	Post-Harvest Management in Sub-Saharan Africa – FAO/IFAD/WFP End of Project Evaluation - Final Report	2020	Burkina Faso, Uganda, DR Congo	x	x
1.4 0	External Review of the project Safe Resources Recovery and Reuse (RRR) - Final Report	2020	Peru, Uganda	x	
1.4 1	EXTERNAL EVALUATION of the initiative "El Agua nos une - SuizaAgua" 2016-2020	2020	Latin America	x	x
1.4 2	Evaluation Report of the Durable Solutions Initiative in Somalia	2020	Somalia	x	x
1.4 3	REVUE EXTERNE INTERMÉDIAIRE DES PROJETS - "RENFORCEMENT DES MOYENS TUNISIENS ÉTATIQUES ET NON-ÉTATIQUES DE RÉPONSE AUX URGENCES HUMANITAIRES LIÉES À LA MIGRATION"	2020	Tunisia	x	

1.4 4	Revue externe à mi-parcours - Projets de la Coopération suisse en Tunisie en matière de prévention de l'extrémisme violent (PEV)	2020	Tunisia	x	x
1.4 5	VALORACIÓN DEL PROGRESO DEL PROYECTO: FONDO DE AYUDA FINANCIERA A TERCEROS (FAFT) DE AYUDA EN ACCIÓN, NICARAGUA	2020	Nicaragua, Costa Rica	x	
1.4 6	Informe de evaluación del proyecto: "Derechos e integración: respuesta a la crisis migratoria de la población nicaragüense refugiada y solicitante de refugio en Costa Rica"	2020	Costa Rica	x	
1.4 7	FINAL EVALUATION OF THE SEEDS AND MARKETS PROJECT (ESWATINI, LESOTHO AND ZIMBABWE) - Final Report	2020	Eswatini, Lesotho, Zimbabwe	x	
1.4 8	EVALUACIÓN DE MEDIO TÉRMINO PROYECTO "MERCADOS INCLUSIVOS"	2020	Bolivia	x	

Ref/ No	Name of the evaluation (additional evaluations)	Year	Country/Region	Document analysis	Survey
2.0 1	Mid-Term Evaluation Report / Afghanistan Independent Human Rights Commission (AIHRC) / Strategic Plan 1393-1397 (2014-2018)	2018	Afghanistan	x	
2.0 2	Evaluation of the Ecological Organic Agriculture Initiative in Africa	2018	Africa	x	
2.0 3	Swiss Contribution Thematic Fund 'Security' Module 2 - Bulgaria Evaluation	2018	Bulgaria	x	
2.0 4	External Review: final report: "Accessible Quality Healthcare" (AQH) Project - implemented in Kosovo	2019	Kosovo	x	
2.0 5	Evaluación de Resultados del Proyecto Fondo de Capital Semilla - Fideicomiso y Perspectivas de Sostenibilidad operativa y financiera	2019	Bolivia	x	
2.0 6	INFORME DE EVALUACIÓN INTERMEDIA DEL PROGRAMA PRODEL	2019	Cuba	x	x
2.0 7	Peer Review Report: Strengthening Oversight Function and Transparency of the Parliament of Serbia Project 2015-2019	2019	Serbia	x	x
2.0 8	External Mid Term Review of the Policy Forum Strategic Plan 2017-2020 - Final Report	2020	Tanzania	x	
2.0 9	External Review of the International Centre for Asset Recovery (ICAR)	2020	Other	x	
2.1 0	External Review of Geneva Water Hub Programme (revised draft report)	2020	Other	x	
2.1 1	External Review of IUNC's Development Relevancy	2020	Other	x	
2.1 2	Gender Evaluation - final report	2020	Horn of Africa	x	x

7.2 Further documents consulted

OECD DAC Evalnet (2020): SDC Evaluation Function.

SDC (2021): How-to Note Evaluation. A toolkit for programme and project evaluations.

SDC (2020): SDC's Evaluations. Evaluations completed in 2020.

7.3 Persons interviewed

Note: In order to provide for anonymity, the interview references in the report are randomised and do not refer to their order below.

- ✓ Buerli, Markus, staff SDC
- ✓ Corderoy, Sylvain, staff SDC
- ✓ Hoch, Katri, staff SDC
- ✓ Jaggi, Martin, staff SDC
- ✓ Läubli, Ursula, staff SDC
- ✓ Oertle, Thomas, staff SDC
- ✓ Poretti, Fabrizio, staff SDC
- ✓ Ruedin, Laurent, staff SDC
- ✓ Anonymous, local employee SDC
- ✓ Weber, Séverine, staff SDC

7.4 Interview guideline

Introduction:

Some days ago, you have participated in an online survey on behalf of SDC regarding your experiences with SDC's project and/or program evaluations. We have selected 10 respondents to conduct a more in-depth interview, and you are one of them.

Thank you very much for your availability today.

My name is Stefanie Krapp. I am Head of the Evaluation Group at the Center for Continuing Education/University of Bern. Together with my colleagues from CEval, we have been commissioned a backstopping mandate for E+C and the Meta Evaluation is one task we are currently working on behalf of E+C. It is within this role that I am conducting the interview with you.

The objective of the interview is to better understand your assessments and get a deeper insight on your perspectives and demands with respect to evaluations. This information will further contribute to assessing the overall quality and usefulness of SDC's evaluation portfolio, and to derive recommendations for their future improvement.

The interview will take app. 30-45 minutes and will be handled strictly confidential.

Following information to be added before the start of each interview:

Interviewee:

Date of the interview:

Current position:

Name and kind of evaluation:

When did the evaluation take place:

1. Evaluation planning and management

1.1 In the questionnaire you indicated "..." as your biggest challenge regarding the planning and management of the evaluation. Can you explain a bit further please? (3.1)

1.2 You indicated that you did search for "..." support when preparing the evaluation, and that you did find the following support: "...". (3.2 – 3.5)

- ✓ If different, why did you not get the support you were looking for?
- ✓ Please elaborate a bit further on the usefulness of this support.
- ✓ What would be more useful/better help in preparing an evaluation (specific topics, how)? And why? Would that be guidelines or rather practical support, like trainings/workshops on evaluation, one-to-one support etc.?

1.3 You assessed the How-to Note as "...". Please explain a bit further. Is the How-to Note sufficient? Should the guideline be improved? In what sense? (3.6, 3.7)

OR 1.3 You indicated that you did not look for any particular support for preparing the evaluation. Why? Do you know the How-to Note? If yes, why did you not use it?

1.4 Ich: Check the answers of the items in 3.9 "your effort for the evaluation" (3.9)

✓ Please explain.

✓ What would it realistically have needed?

✓ What would you consider as appropriate in terms of time (max. amount of working days)?

✓ Which type of logistics and technical support should be provided?

1.5 You indicated that "... " should be provided in the future to support you in the planning and management of evaluations. How and why? Also: Reasons why some items have not been mentioned (i.e. specifically ask for the Help Desk). (3.10)

1.6 You indicated that you have been unsatisfied with the collaboration with the evaluator in the sense that "...". Would you please explain further? (3.12)

1.7 You indicated that you have been unsatisfied with some competences of the evaluator. Which competences in particular was the evaluator/the evaluation team lacking? And do you think this had an effect on the overall quality of the report/the usability of the results and recommendations? (3.13) How can SDC provide support to overcome this issue in the future?

2. Usability and use of the evaluation

2.1 You rated the final evaluation report regarding the aspects "...", and "... " as unsatisfactory. Please explain why. (4.1)

2.2 How do you assess the readability of the final evaluation report?

2.3 You rated the recommendations in the final evaluation report regarding the aspects "...", and "... " as unsatisfactory. Please explain why. (4.2)

2.4 What have you been using/are using the evaluation results/recommendations for? (4.3)

Or 2.4 You have indicated that you did not make any use of the evaluation at all, why not?

2.5 If you applied recommendations, did this have any value for the project/did anything change? Did it improve the project's effectiveness, efficiency, impact or sustainability? If so, please elaborate. If not, why not?

2.6 What would have been/would be necessary in order to increase the usability of the evaluation/concrete recommendations? Are there any hindering systems/structural factors?

3. Other

3.1 Are you familiar with SDC terminology (example: Approach Paper)? If no, what is missing?

7.5 Further tables

The following tables show the correlations of the analysis and survey aggregates. Correlation means that the values of two (or more) variables are associated with each other (i.e. if one changes, the other changes as well). The relationship is not necessarily causal. Its causality can, however, be postulated if e.g. one variable precedes the other (e.g. ToR are usually written before the Executive Summary) or if it is plausible (e.g. recommendations are more likely to be used if they are perceived as being useful).

Table 1: Correlations of document analysis aggregates

	ToR	Executive Summary	Introduction	Methodology	Findings	Conclusions & Recommendations	
ToR	r	1	,362*	0,160	-,380*	-,390*	0,026
	p		0,036	0,366	0,027	0,022	0,883
	N	34	34	34	34	34	34
Executive Summary	r	,362*	1	-0,057	-0,244	-0,164	0,098
	p	0,036		0,747	0,165	0,354	0,583
	N	34	34	34	34	34	34
Introduction	r	0,160	-0,057	1	0,093	-0,041	-0,160
	p	0,366	0,747		0,600	0,817	0,366
	N	34	34	34	34	34	34
Methodology	r	-,380*	-0,244	0,093	1	,545**	0,187
	p	0,027	0,165	0,600		0,001	0,290
	N	34	34	34	34	34	34
Findings	r	-,390*	-0,164	-0,041	,545**	1	,535**
	p	0,022	0,354	0,817	0,001		0,001
	N	34	34	34	34	34	34
Conclusions & Recommendations	r	0,026	0,098	-0,160	0,187	,535**	1
	p	0,883	0,583	0,366	0,290	0,001	
	N	34	34	34	34	34	34

* Correlation is significant at a level of 0,05 (two-sided).
** Correlation is significant at a level of 0,01 (two-sided).

Table 2: Correlations of survey aggregates

	Q3.1	Q3.12	Q3.13	Q4.1	Q4.2	Q4.6	Q4.7	
Q3.1: Preparation of the evaluation	r	1	,229*	,257**	,367**	,280**	,183*	,288**
	p		0,012	0,005	0,000	0,002	0,049	0,002
	N	126	119	118	120	119	116	118
Q3.12: Collaboration with evaluators	r	,229*	1	,471**	,572**	,497**	,548**	,483**
	p	0,012		0,000	0,000	0,000	0,000	0,000
	N	119	120	118	119	118	114	117
Q3.13: Competencies of evaluators	r	,257**	,471**	1	,565**	,528**	,480**	,407**
	p	0,005	0,000		0,000	0,000	0,000	0,000
	N	118	118	119	119	118	114	117
Q4.1: Satisfaction with evaluation report	r	,367**	,572**	,565**	1	,732**	,645**	,591**
	p	0,000	0,000	0,000		0,000	0,000	0,000
	N	120	119	119	121	120	116	119
Q4.2: Quality of recommendations	r	,280**	,497**	,528**	,732**	1	,577**	,655**
	p	0,002	0,000	0,000	0,000		0,000	0,000
	N	119	118	118	120	120	115	118
Q4.6: Usability of recommendations	r	,183*	,548**	,480**	,645**	,577**	1	,461**
	p	0,049	0,000	0,000	0,000	0,000		0,000
	N	116	114	114	116	115	116	115
Q4.7: Use of evaluation	r	,288**	,483**	,407**	,591**	,655**	,461**	1
	p	0,002	0,000	0,000	0,000	0,000	0,000	
	N	118	117	117	119	118	115	119

* Correlation is significant at a level of 0,05 (two-sided).
** Correlation is significant at a level of 0,01 (two-sided).

Caption: r: Correlation coefficient, p: level of significance, N: number of cases

7.6 Detailed outline of the methodology

For the document analysis an **assessment protocol** was developed and agreed with E+C. The protocol contains dedicated sets of assessment criteria (e.g. description of evaluation subject) and corresponding dichotomous indicators (e.g. description of intervention area, objectives, intervention logic) for assessing the quality of each chapter of the evaluation reports. According to the typical structure of these reports, the criteria are categorised as follows:

1. Terms of Reference
2. Executive summary
3. Introductions and context analyses
4. Evaluation methodology
5. Evaluation findings
6. Conclusions and recommendations

For all indicators, definitions were included in the protocol, as well as functions for calculating the aggregate value of each criterion.¹ Furthermore, guidance for an unambiguous assessment of the indicators was provided in form of brief definitions (e.g. the indicator “description of intervention period” is fulfilled, meaning it receives the value “1”, if start and end date of the intervention are mentioned, if start or end date or both are missing, it receives the value “0”). Whereas the overall structure and elements of the protocol base on an analytical framework developed at CEval and applied in a number of quality assessments before, its indicators and level of fidelity were adapted to the information needs of the SDC. Due to its size and format, the assessment protocol including all criteria and indicators is attached to this report as a separate document (Appendix 2).

In order to ensure **results traceability**, the documents were analysed with a software for qualitative text analysis (MaxQDA®). Thereby, the assessment protocol was used as a coding scheme according to which the reports were scrutinised for relevant content. The coded content was then assessed pursuant to the guidance of the protocol. For providing maximum inter-rater reliability, during the assessment a peer review was applied by which the team compared and if necessary discussed inconclusive assessments.

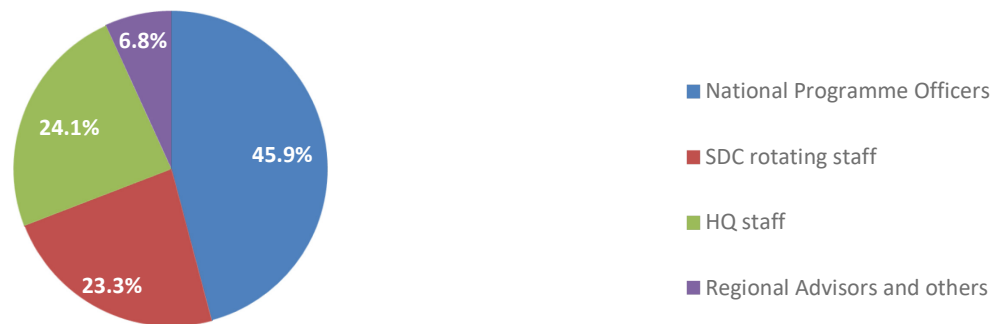
To assess the utility of the evaluations for SDC staff a **semi-standardised full-population survey** (27.10.-23.11.2021) and **interviews** with ten selected survey participants (December 2021) were conducted. The purpose of the combination of these two empirical data collection instruments was to achieve likewise statistically and contextually representativeness of SDC staff in terms of their position and regional distribution. While the survey yielded primarily quantitative data, the interviews provided further insights about experiences with the management of evaluations as well as motivations and justifications of attitudes towards particularly critical issues (e.g. the factual use of recommendations) and specific support needs (cf. interview guideline in Annex 0). Therefore, for the latter survey respondents were purposively selected on the basis of their working region, position and, in case, noteworthy assessments provided in the questionnaire.

For the survey, a questionnaire (cf. Appendix 1) was developed, focusing on the SDC staff's assessment of the preparation of the evaluation, its implementation, results and added value. Furthermore, questions on the quality of available and further required support were included. The survey was realised with an online software called OFB SoSciSurvey®. It was implemented as a personalised anonymous survey, meaning that each respondent received a personalised questionnaire whose entries, however, could not be traced back. This way it was possible to match survey and document analysis data without compromising the anonymity of the respondents. In order to select interviewees, the questionnaire allowed the respondent to voluntarily enter his/ her email address for this specific purpose.

¹ $\sum(I_1 + \dots + I_n) / n$; whereby I = Indicator, n = number of indicators.

A total of 592 SDC employees were invited to participate in the survey from which 133 valid questionnaires were received. This corresponds to a net return rate of 22.5%, which is an average value for such a type of survey. From these 133 responses, 34 were relating to evaluations that were also subject to the document analysis, which allowed comparing the experts' and SDC staffs' assessments. As the following figure shows, almost half of the respondents (45,9%) represented National Programme Officers and a quarter SDC rotating field staff (23,3%) and HQ staff (24,1%), respectively. The remaining 6.8% named themselves either as Regional Advisors or officers with special or several functions.

Figure 1: Functions of respondents



As regards the regional distribution of the evaluations the respondents referred to, Africa was represented with the largest share of a good quarter (27.1%), followed by Asia (16.5%) and Eastern Europe (9.8%). By contrast, clearly underrepresented were evaluations from Latin America (4,5%) and evaluations of global or other type of interventions (3,8%), as the following figure shows.

Figure 2: Regional distribution of evaluations respondents refer to



Eventually, more than a third of the respondents (38.3%) decided not to reveal to which evaluation her/ his responses refer to.

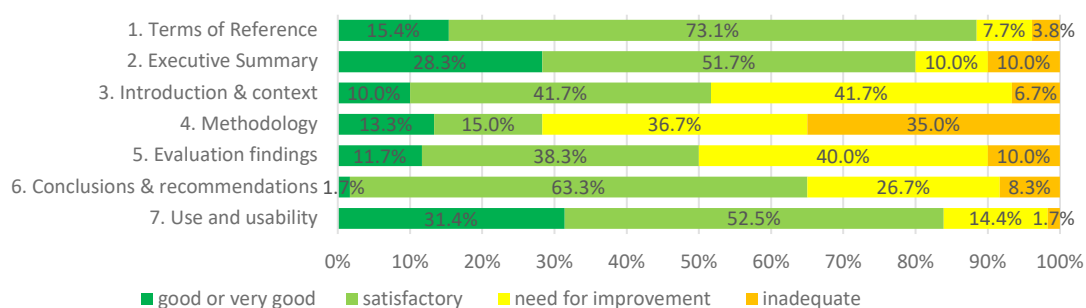
Data from the document analysis, the survey and the interviews were analysed **qualitatively and quantitatively**. The qualitative data analysis followed an inductive approach, as promulgated by e.g. Mayring². Quantitative data was analysed statistically, using e.g. frequencies (of cases with particular features), distributions (of shares of aggregated assessment criteria), arithmetic means³ (of overall category ratings) and correlations⁴ (for identifying relations between criteria). Due to the low number of cases and overall non-random selection of evaluations, inferential statistics were not applied.

Wherever meaningful, charts and tables are used in this report to substantiate and illustrate findings. Thereby, if not specified otherwise, for all charts relating to the document analysis n is 60 (i.e. number of reviewed evaluation reports), and for the survey n is 133 (i.e. number of survey participants). Further tables with full sets of frequencies and correlation test results can be found in Annex 7.5. An MS Excel® spreadsheet with the raw data from the document analysis and the survey, including the sources for all figures and tables in this report, is attached to this report as a separate document (Appendix 3).

7.7 Detailed analysis of each criterion of the document review

Overall assessment of the evaluation reports at a glance:

Figure 3: Evaluation report assessment summary (for 1. n=26, for 2.-7. n=60)⁵



The figure shows the aggregated assessments of all criteria according to the Quality Assessment protocol. Therefore, the values of all criteria were added, and the sum was divided by the number of criteria. Finally, the results were categorised as follows:

- ✓ $\leq 0,25 \rightarrow$ inadequate
- ✓ $> 0,25$ to $\leq 0,5 \rightarrow$ need for improvement
- ✓ $> 0,5$ to $\leq 0,75 \rightarrow$ satisfactory
- ✓ $> 0,75 \rightarrow$ good or very good

² Cf. Mayring, e.g. 2000 (cf. <http://www.qualitative-research.net/index.php/fqs/article/download/1089/2384>).

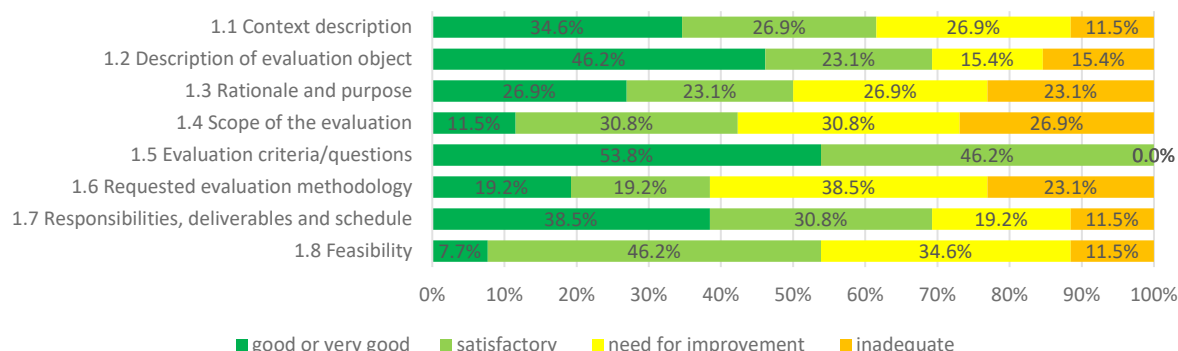
³ For easier comparability all criteria and category aggregates are normalised, i.e. converted in a standard scale ranging from 0 (worst) to 1 (best). Accordingly, all arithmetic means (AM) range from 0 to 1. In order to comprehend the explanatory power of the AM standard deviations (SD) are provided in parentheses. Thereby, it has to be considered that the lower the SD (little spread around the AM) is, the more consistent the ratings are, the higher (large spread around the AM) it is, the less consistent. As rule of thumb, an SD lower than about a quarter of the scale range (here: 0,25) indicates sufficiently consistence for a meaningful interpretation of the AM.

⁴ For identifying correlations, Pearson correlation coefficients were calculated. Moderate correlations ($0,5 < r < 0,7$) were reported only if they were highly significant at the level of $p \leq 0,01$ (two-sided test), strong correlations ($r > 0,7$) were reported if they were significant at the level of $p \leq 0,05$ (two-sided test).

⁵ Only for 26 of the 60 evaluations ToR were available in the annex of the reports.

7.7.1 Terms of Reference

Figure 4: Overall assessment of Terms of References (n=26)



Evaluation object and context

Most ToR contain an analysis of the current situation of the intervention area (84.6%). Some ToR describe the international and/ or national context (76.9%), and sometimes the Swiss development policies and/ or strategies (53.8%). Eventually, more than a half of the ToR's context descriptions (65.4%) refer adequately to the intervention.

Whereas overall the evaluation object is mostly well described, however, a third of them features room for improvement. A closer investigation of the individual indicators used for the assessment of this criterion reveals that the intervention objectives and its components are described in most of ToR (85% and 69% respectively). Besides, the time period of the intervention and its budget is also provided (76.9% and 61.5% respectively) and, sometimes the phases of the intervention are differentiated. Moreover, the current state of the intervention seems to be adequately described (58%). Whereas the implementation arrangement is also well specified (61.5%), the target groups/ stakeholders of the intervention is insufficiently or not at all depicted in 42 percent of the cases. It is also noticeable that while the scope of the intervention (intervention components, temporal delineation, etc.) is well defined, the description of the intervention area is lacking (46%). The weakest point of the evaluation object description is the absence of an intervention logic or results model (e.g. LogFrame, visualisation of results chain or model, description of the project's Theory of Change) in more than a half of the ToR.

Evaluation objectives and scope

At large, the objectives of the evaluation are adequately acknowledged. However, while all ToR communicate clearly the evaluation's objectives, its rationale (reasons why the evaluation is implemented or requested) is mentioned considerably less often, as is its intended use (61.5% respectively). The users and addressees of the evaluation results are mentioned in less than one third of the cases (30.8%).

In that regard it is worth to mention that some interviewees (Int 6,10) point out that it is most important to make very clear from the beginning why an evaluation is needed. Therefore, the ToR should lay out the focus and the purpose of the evaluation and which resources are needed. They argue that not each phase needs an evaluation, but that in SDC the evaluation type is chosen according to the 'appreciation' of the project/program manager and that he/ she decides by him/ herself which type of evaluation to carry out. Another respondent mentions that in one case the team, the country office and the HQ had different views on the purpose and objectives of the evaluation, which were not mapped beforehand and caused substantial problems, as the following quote suggests: "There should have been a transparent discussion before start, leading to a clear joint understanding. It reflects a bit how SDC as an organisation works and how much the hierarchy matters. It caused quite a lot of tensions. The local consultant mentioned at the end that he 'needs to heal after that experience'." (Int 2)

The scope of the evaluation is also not clear in many ToR. Only a good third delineates the period of the intervention's implementation (38.5%) and the geographic areas (34.6%) that should be covered by the evaluation. The relevant components of the intervention (i.e. activities implemented in a certain sector, for a specific target group or with a particular instrument) for the evaluation are mentioned in about half of the ToR (53.8%).

Evaluation criteria and questions

As indicated above, evaluation criteria and questions are always mentioned in the 26 ToR; they also seem adequate and specific to the evaluation object. When analysing further the evaluation questions, they are always referred to the OECD/DAC criteria and in more than half of the cases to cross-cutting issues (57,7%). They are well structured (e.g. in a list, bullets points under each evaluation criteria). It can, however, also be noticed that in none of the ToR any other than OECD/DAC criteria are used, such as for instance ALNAP criteria, gender or equality criteria, or criteria related to the SDGs. Two interviewees (Int 6,10) argue that an evaluation should be implemented to know something specific, which also implies that not all DAC criteria are required. Another interviewee claims the importance of the evaluation questions: *“Evaluations turned out to be most useful when the questions have been precise. But often they are too general because we might not know what we want. We should be as precise as possible so that the evaluator knows what he needs to do.”* (Int 3)

Methodology

The envisaged methodology is mostly specified in the ToR (80.8%) but could be more exhaustive. Indeed, expected data collection instruments (e.g. survey, interviews, Focus Groups Discussions (FGDs), observations) and analysis methods (e.g. qualitative content analysis, multivariate descriptive or inferential statistics) could be more detailed. Moreover, mix of qualitative and quantitative methods is not explicitly requested in most cases (74.1%). In fact, most of the evaluations follow qualitative methodologies based on desk reviews and empirical data collected with FGDs and interviews. Whereas obligatory data sources (e.g. project documents, partner staff, target groups) are specified in more than two thirds of the ToR (69.2%), triangulation is solicited considerably less often (38.5%). If at all, data collection tools are only quoted, mentioned as being indicative and rarely described (e.g. in bullet points in Rep 2.04).

A few ToR include specific 'requests', such as for applying a participatory approach (e.g. Rep 1.01,09,29), conducting a desk review, field visits or specific stakeholder workshops. For instance, some stipulate field visit to the intervention sites, also mentioning stakeholders to be interviewed (e.g. Rep 1.05,10,15,17,28,29). While many ToR are not precise on the methodology (e.g. bullet points only), lack guidance (Rep 1.09) or do not have a methodology section at all (Rep 1.35), some exceptions exist (e.g. Rep 1.20,30,39). Report 1.25 does not only suggest the methodology but also describes its justification. Betimes, the evaluation team is expected to suggest a methodology (e.g. Rep 1.01,04,06). The responsibilities and roles of the evaluators are specified in about two thirds of the ToR (65.4%). Unfortunately, thereby the task division between the regional and international consultant are not always differentiated. Finally, whereas most ToR outline the desired deliverables of the evaluation (96,2%), its overall deadline is specified in only about three quarter of the cases (73,1%) with the envisaged milestones and corresponding timeline even somewhat less often (58%).

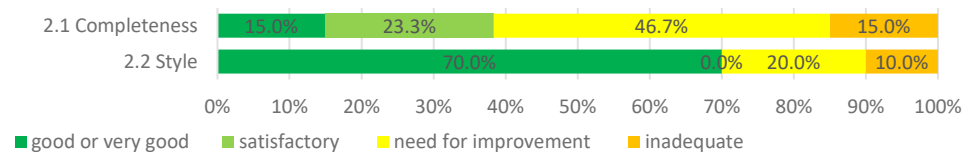
Feasibility

Four out of five ToR specify the number of working days allocated to the evaluation. An expert review of the expected evaluation methodologies reveals that it is mostly not possible (63.1%) to implement the evaluation within this time budget. In fact, the ToR underestimate the scope of work for the demanded tasks and deliverables. Plus, the evaluation process and how its quality will be assured is only explained in 42 percent of the ToR. Interesting to note that many interviewees agree that the effort for evaluations is systematically underestimated and that the deadlines are too ambitious. They confirm that the time it takes

to write the report including feedback loops and reviews, as well as to write the Management Response is often not sufficiently considered (Int 3,5,6,8-10), as the following quote illustrates: “You have to be very clear what exactly you expect from the evaluators in which form. This includes the formulation of recommendations which should be precise, target-group specific, time-bound, prioritised and have a max. number. Otherwise, it’s a back-and-forth which needs too much time.” (Int 8) It is also mentioned that the amount of time required highly depends on the evaluators competencies: “We invested too much time with the consultant. We had to explain over and over again.” (Int 8)

7.7.2 Executive summary

Figure 5: Overall assessment of Executive Summaries (n=60)



Completeness

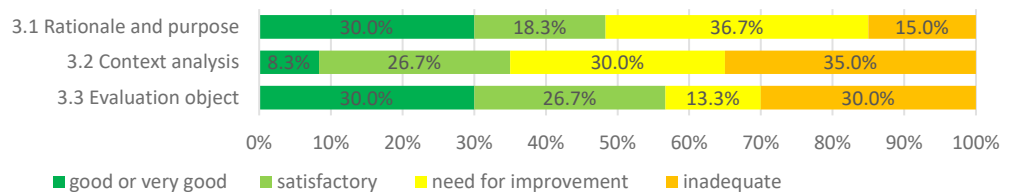
At large, the executive summaries are a mediocre preview of what will be addressed in the report. Most of them present what has been evaluated with a description of the intervention (66.7%), the main findings (80.0%) and recommendations (83.3%). However, few of them tackle the rationale or purpose of the evaluation (35.0%) and the evaluation design and methods (8.3% and 46.7% respectively); essential to understand the reasons why and how the evaluation is implemented. Further findings corroborate the lack of evaluation designs as also outlined in Chapter 3.4. Plus, approximately only half of the executive summaries describe the scope of the evaluation (48.3%), its goals (50.0%) and conclusions (48.3%). Eventually, very few include lessons learned (16.7%) if they are specified in the evaluation report (i.e. in 50.0%). Hence, the executive summaries feature considerable room for improvement as regards their completeness.

Style

In terms of their style, it can be concluded that the executive summaries are well written. It is easy to understand what has been evaluated and what the main findings and recommendations are. The language of the majority of the reports is good (83%), with a few exceptions. The executive summaries are also consistent with the report in most cases (75%). That means that the executive summaries do not contain any new information that is not covered in the report elsewhere. It is important to notice that some reports do not contain at all an executive summary such as reports number 1.06,07,14,16,33,46.

7.7.3 Introduction and context

Figure 6: Overall assessment of introductions (n=60)



Rationale and purpose

Whereas nine out of ten introductions (88.3%) contain a description of the evaluation's objectives, its rationale is specified in less than two thirds of the cases (61.7%). Furthermore, the intended use of the evaluation appears to be missing in half of the introductions (53.3%) and the intended users are lacking even more often (73.3%).

Context analysis

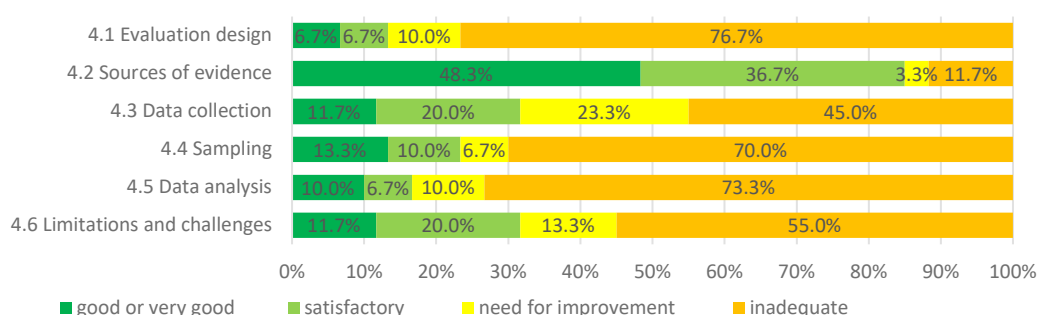
In most of the reports the context analyses are rather weak and appear insufficient, especially concerning their references to Swiss Development Policies (23.3%). On the one hand, only 58 percent of the reports contain a context analysis at all. Less than a half are referring to international, national or regional policies, strategies or context (43.3%). On the other hand, some contexts are quite extensively described, however, often then without sufficient connection to the intervention (58.3%). Some interviewees attribute this finding to insufficient sectorial and context experiences by many evaluators, so that they do not know what to write about the local environment (Int 3,8,10).

Evaluation object

As regards the evaluation object, almost half of the introductions fail to provide an adequate description (43.3%). Only the intervention's objectives are well outlined in most introductions (75.0%). All other criteria, such as the intervention's budget, area, target groups and further stakeholders or implementation arrangements are not consistently addressed. An intervention logic or results model is mostly missing (70.0%). Yet, some reports describing the evaluation object quite extensively (e.g. Rep 1.01) could be considered as good practices. However, some of these descriptions are of poor quality (e.g. Rep 02.09,12) making them difficult to read or understand the intervention subject comprehensively. In sum, it needs to be concluded that most reports look like 'insider' documents, not suitable for readers who were not involved in the intervention's implementation.

7.7.4 Methodology

Figure 7: Overall assessment of methodology sections (n=60)



Taking a closer look at the assessment criteria according to which the evaluations' methodologies were assessed, their rather negative rating reveal the main issues. While most evaluators manage to name their sources of evidence (85.0% rated very good/ good or satisfactory), barely a quarter is able to describe the sampling to at least a satisfactory extent. Data collection instruments are outlined adequately in only about a third of the cases, and so is the discussion of the limitations and challenges of the evaluation. By contrast, the vast majority fails at an adequate description the data analysis methods and the evaluation design, with the latter almost nine out of ten featuring substantial flaws or no description at all.

Evaluation design

Beside the fact that most reports do not contain any information on the evaluation design, those which do mostly lack key elements such as the presentation of an evaluation matrix (assessment grid) at least containing the evaluation criteria, data sources and methods applied, or information if an attribution or a contribution analysis has been undertaken. Eventually, in only ten cases (16.7%) the evaluation design can be deemed as appropriately adapted to the intervention. Another remarkable finding is that repeatedly evaluators were actually describing the evaluation 'approach' (e.g. Developmental Evaluation, participatory approach) or the practical implementation of the evaluation instead when referring to its design. Such mistakes indicate a misconceived terminology.

Sources of evidence

While being on average well rated, the description of the sources of evidence shows nevertheless some deficits in regard to what and how sources are referred to. While project documents and representatives of implementing partners are frequently cited (in 88.3% and 81.7% respectively), beneficiaries – which supposedly should be the most relevant data source – are comparatively less often mentioned (71.7%), as is project monitoring data (45,0%). The latter, however, may also be associated to lacking monitoring systems of the projects.

Another observation is a considerable discrepancy between what is stated in the method sections as data sources, what is listed in the annexes of the reports and what is actually being used as a source of evidence for a finding in the main body. It appears, that the general rule that an annex should only contain sources, which are actually used in the text does not apply for evaluations. Moreover, if at all, statements are rarely furnished with evidence or contrasted by referring to particular stakeholders.

Data collection instruments and sampling

As regards the data collection, astonishingly in almost a quarter of the cases the method section gets along without a decent description of the data collection instruments. Also, they are not always found in the annex. As a look into the instruments enables the reader to get an idea about the methodological quality of the findings, particularly the latter has to be considered as a major shortfall. This finding is backed up by the fact that in less than a fifth of the evaluations (18.3%) the validity and reliability of the data collection is discussed at all.

The available information about the sampling of empirical data shows likewise room for improvement. If such information is available at all, the sampling strategy is described and justified in only about two thirds (64.0%) and half of the cases (48.0%) respectively. Unfortunately, the document analysis yielded not sufficient evidence for assessing the adequacy of the sampling strategy since mostly information on the size and composition of the basic population was lacking. In one case (Rep 2.02) where a very thorough description could be found, the formula for calculating the sample size appeared to be unsuitable for the intended purpose of providing for external validity of the evaluation results.

Data analysis, limitations and challenges

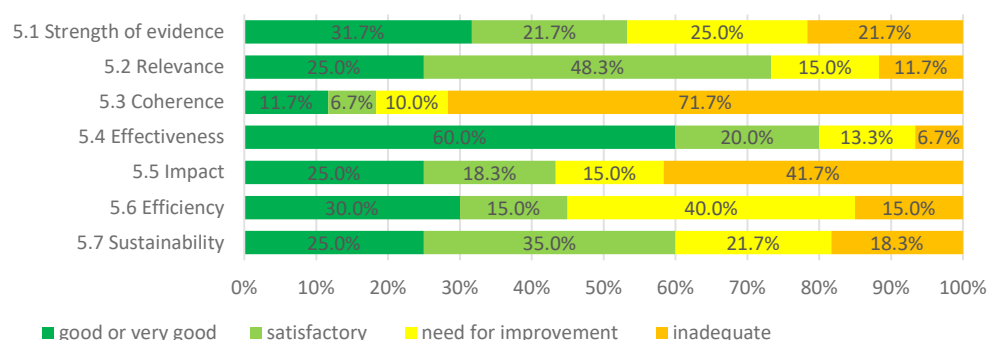
In contrast to the description of the data collection, the **data analysis** is considerably less frequently outlined in the evaluations' method sections, with only about a fifth of the reports (21.7%) containing information on the latter. The descriptions are, however, mostly not sufficient to assess the appropriateness of their application and mix in view of the evaluation subject. What strikes the eye though, is that apparently in only seven cases (11.7%) quantitative methods were applied, which does not speak for a sound method mix on the whole.

Consistent with the finding that method sections mainly focus on describing data collection is the fact that in those cases where **limitations and challenges** are discussed, they also rather refer to data collection (38.3%) than on data analysis (28.3%). After all, such discussions mostly also provide information on how the evaluation team dealt with these challenges in order to ensure data quality.

In light of this quite sobering picture about the methodological quality of the evaluations, respectively their descriptions in the evaluation reports, the incidental finding from an interview with an SDC representative (Int 1) who claims that the evaluators were exaggerating the methodological rigor of the evaluation is striking. This statement indicates a lacking understanding of and awareness for the necessity of a decent methodology for providing sufficiently valid and reliable results that are suitable for decision-making.

7.7.5 Evaluation findings

Figure 8: Overall assessment of findings sections (n=60)



Leaving aside the assessment of the reports' Coherence sections, the figure shows that the quality of the evaluations' findings about intervention impacts are assessed weakest with less than half (43.3%) being of at least satisfactory quality, while the findings about their effectiveness are assessed best (80.0% rated satisfactory or good/very good). Whereas about only a quarter (26.7%) of the relevance discussions feature considerable deficits, more than half (55.0%) of the efficiency and 40 percent of the sustainability sections do so. Eventually, in almost half of the cases (46.7%) the strength of the evidence provided for the evaluations' findings leaves doubts about their reliability.

Strength of evidence

A closer examination of the indicators of the evidence strength shows that in fact the data basis of the findings is outlined in the method sections and in most cases specified in the annexes (e.g. lists of documents, interview partners), however, not referred to appropriately in the findings chapters (cf. also Chapter 3.4). Instead, rather general references to empirical sources (e.g. formulations like "some/ many/ all interviewees confirmed ...") can be found and hardly any reference to documentary data. In only half of the evaluations, findings are put into perspective with reference to different data sources (e.g. comparing feedbacks from different stakeholders). Moreover, in almost three out of four cases (74.3%), the attribution of observed changes to the intervention or the intervention's contribution to them is not being adequately discussed. Another shortcoming is that a good third (35%) of the evaluations fails to separate findings from conclusions and betimes even recommendations. Thereby, both can be found, conclusions and recommendations included in the presentation of the findings as well as new findings introduced in the discussion of the conclusions and recommendations. What is even more concerning is the fact that the majority of the reports (62,7%) neither follow nor contain an analysis grid (evaluation matrix) in which the evaluation questions, indicators, data sources and methods are disclosed.

Discussion of OECD/DAC criteria

Apparently not all evaluators are familiar with the terminology of the OECD/DAC criteria or at least the criteria's current definitions. This leads to sometimes incoherent contents of the respective chapters with aspects being discussed under the wrong heading (e.g. Rep 2.04), project outputs in the impact chapter). What is striking is that **Relevance** is apparently the most extensively discussed criterion, at least in terms of the length of the respective chapters. Thereby, the intervention's consistency with the target groups'/ final beneficiaries' needs and with the policies of the partner government are likewise often captured in most cases (73.3%), as is the adequacy of the intervention design (71.7%). By contrast, international or regional policies and strategies are clearly less often addressed (56.7%) and so is the intervention's consistency with the SDC development cooperation policy (38.3%). Finally, the adaptation of the intervention to changing context conditions is frequently overlooked (considered in only 31.7%).

As mentioned, most evaluations under investigation deal well with the **Effectiveness** criterion. Yet, while output and outcome achievement are equally discussed adequately in most cases (81.7% and 86.7% respectively), the interventions' contributions to observed changes at outcome level are reflected somewhat less often (60.0%). It appears that often a lot of information is being provided about what has changed in the intervention area but hardly any evidence how the intervention contributed to those. Moreover, only half of the evaluations (50.0%) provide disaggregated figures for outcome achievement, e.g. at different target groups. Examining the narratives more in detail, it turns out that betimes Effectiveness assessments are presented on the basis of stakeholder feedbacks with only little to no analytical transfer. Also, outputs and outcomes are regularly confused and hence wrongly contextualised.

The discussion of the intervention's **Impacts** features comparably more shortcomings, starting with that less than two thirds of the evaluations (58.3%) address this criterion in the sense of its definition at all. Even less explicitly discuss impact achievement (53.3%) and the intervention's contribution to observed impacts (43.4%) in particular. Corresponding to these numbers, only 27 percent of the evaluations provide an impact analysis disaggregated by specific (target) groups. What is even more worrisome is the fact that only two out of ten evaluations (20.0%) make an effort to shed a light on unintended impacts, therewith impeding a comprehensive impact assessment in compliance with current evaluation standards.

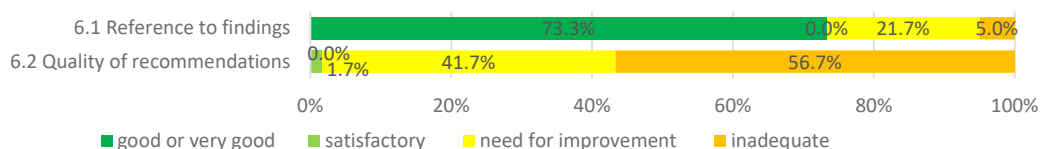
In contrast to Relevance, on average **Efficiency** turns out to be the shortest discussed criterion in the examined evaluation reports. Thereby, most put focus on the assessment of the intervention's management and cost efficiency (73.3% and 83.3% respectively), while production and allocation efficiency (43.3% and 28.3% respectively) are considerably less often considered. Also, the timeliness of the intervention's implementation is discussed in only 40 percent of the cases. Against the backdrop of these quantitative figures, the quality of the presented analyses leaves a lot to be desired. None of the reports features a decent cost-benefit analysis. Even less sophisticated methods like benchmarking or the so-called follow-the-money approach can hardly be found. Instead, most efficiency analyses appear rather particularistic, focusing only on a few key aspects such as the added value of the intervention's cooperation with other donors' activities (Rep 1.15).

The picture looks similar for the evaluations' **Sustainability** assessment. None of the criterion's indicators is considered in more than two thirds of the cases, with the likeliness of the continuation of the intervention's benefits and the risks to the sustainability of the intervention's results being discussed most often (both 63.3%). Whereas the capacities of the partners, target groups or beneficiaries to benefit sustainably from the intervention's results is also assessed relatively often (58.3%), the intervention's strategies to mitigate the risks to the sustainability of its results is only included in a third of the evaluations' sustainability assessment (33.3%). Eventually, only 17 evaluations (28.3%) discuss sustainability

as a multifaceted concept, taking at least two of its dimensions into account. Again, a closer look into the narratives reveals room for improvement. While the primarily anticipating assessments may be owed to the fact that the sample does not contain any ex-post evaluations, the analyses' lacking depth (e.g. not considering dissemination effects) cannot (only) be ascribed to the point in time when evaluations took place. It rather appears that in general the respective chapters are geared to provide recommendations only than actually assessing the intervention's sustainability.

7.7.6 Evaluation conclusions and recommendations

Figure 9: Overall assessment of conclusions and recommendations sections (n=60)



Reference to findings

When looking in detail at the different criteria, the conclusion chapters refer well to the findings (73.3% rated very good/good). Conclusions seem to be logically connected and derived from evaluation findings (80.0%); and recommendations appear to be well grounded and based on the conclusions (91.7%). However, their logical connection is not made explicit and betimes recommendations are just listed in bullet points.

Quality of recommendations

In contrast, more than half of the reports (56.7%) contain recommendations of inadequate quality. While the majority of recommendations are addressing target groups (63.3%), very few are prioritised (10.0%) or even time-bound (8.3%). At large, many appear rather general and not sufficiently specified to be actionable; a finding which is also confirmed by the empirical data from the survey and the interviews.

Only a half of the reports provide lessons learned. However, if they are presented, they are well structured (e.g. in sub-categories, themes) and relevant, and thus represent an added value for the intended users of the evaluation (e.g. Rep 1.19,34,36).

The finding that the evaluations' recommendations feature considerable shortcomings is insofar quite worrisome as they represent the most important part of the report to its readers, as is also confirmed by the interviewees. They also approve the experts' assessment in terms of considering recommendations often as not clear enough, too general, not prioritised and difficult to translate into concrete actions (Int 2,3,5,6,9). Some interviewees (Int 2,5,6) question the independency and impartiality of the evaluators if they have been commissioned by SDC many times, which has severe implications for the formulation of the recommendations. Evaluators would then tend not to be sufficiently critical (Int 2,5,6). Or as one interviewee explains: *"Often the national program officers put pressure on the formulation of the recommendations to go in their direction. The independence of the evaluation is then questionable. It's hard to draw the line between the commissioner and the evaluators and respect the boundary. More experienced evaluators are rather not obliged to the pressure."* (Int 5)

7.8 Detailed analysis of the survey and interview results

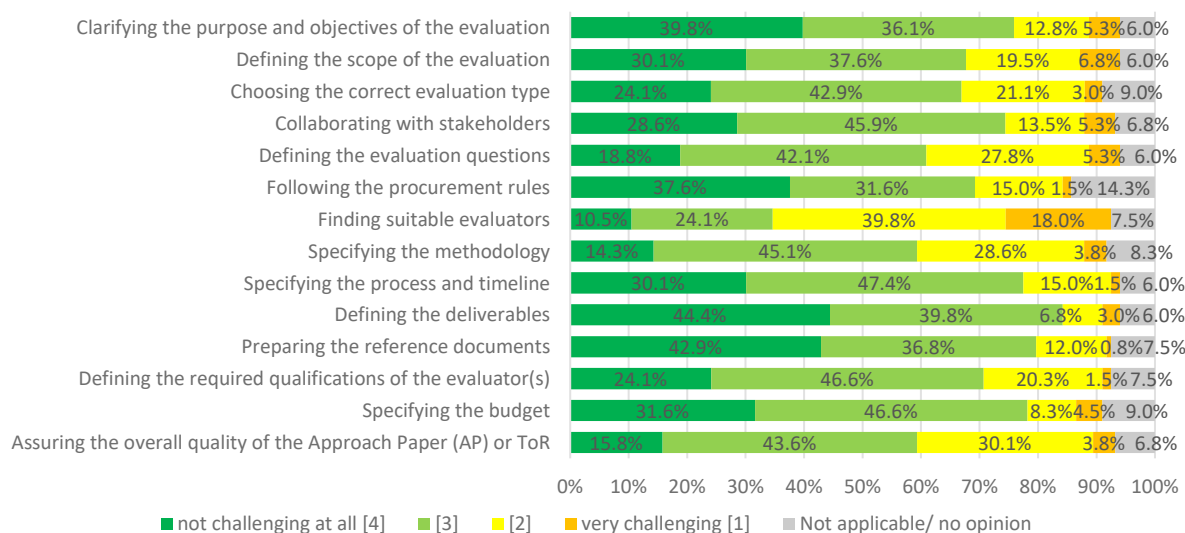
7.8.1 Implementation and Process Use

For assessing the implementation of the evaluations and their process use, SDC staff was asked in the online survey and the interviews how challenging they perceive the **preparation of the evaluation**, how they assess their **collaboration with the evaluators** and the **evaluators' competences** as well as how they rate the **adequacy of their own effort** for the evaluation.

Preparation of the evaluation

The majority of the survey participants feels comfortable with most of the tasks associated with the preparation of an evaluation: define the deliverables of an evaluation (84.2%), prepare reference documents (79.7%), specify the budget (78.2%) or the process and timeline (77.5%). Most challenging for the survey respondents is to assure the quality of the Approach Paper (AP) or Terms of Reference (ToR) (33.9%), followed by specifying the methodology (32.4%) and defining the evaluation questions (33.1%).

Figure 10: Assessment of the evaluation's preparation



The findings from the interviews confirm the survey results to the widest extent and furthermore provide explanations for some observations. Most interviewees agree that finding suitable evaluators (Int 2-9) and writing the ToR (Int 2,3,6,8,10) are the biggest challenges when preparing an evaluation. In particular finding local experts and mid-level evaluators is deemed very difficult. Interviewees (Int 4,5,7,8,9) explain that they often address this challenge by picking consultants they already know from the roster, if there is one, or ask colleagues from other projects in the region to recommend one. From their perspective, this makes it on the one hand easier to find suitable evaluators who already know the organisation, but on the other hand enforces to first think about the person and then about the evaluation, whereas it should be vice versa. 2 respondents (Int 5,9) claim that complicated and time intensive tender processes support this practice, so that tender processes are often avoided. They suggest improving the tender process and receive better guidance from HQ in that regard.

Apparently, the issue of the evaluators' independence is reflected controversial among the interviewed SDC staff. While some (INT 3,9) emphasise the experience with and good knowledge of SDC being a necessary prerequisite for good quality and usable evaluation results, others (INT 2, 6) are concerned about the danger of lacking independence introduced thereby. Interesting to note that even those who express their concerns, nevertheless, admit having recruited the same consultants repeatedly, simply because it was easier to organise.

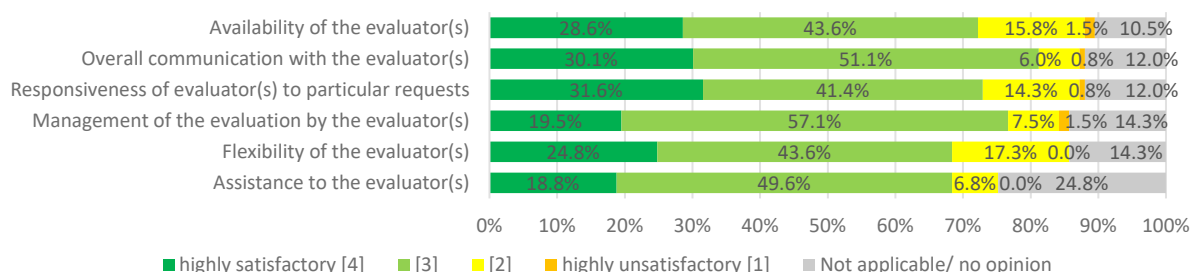
As regards the challenge of drafting ToR the interviewees provide a number of reasons with lack of professional evaluation knowledge and experience being the most important one (Int 2,3,8). Although the project managers usually are aware of the SDC PCM and RBM system and the field handbook which contains a ToR template, or at least know previously used templates of ToRs, results from the interviews suggest that often unexperienced staff without a clear understanding of how to write adequate ToR for their planned evaluation is assigned with this important task. The downside of this 'learning-by-doing' practice is recognised as the following quote indicates. *"If you do not have a clue about evaluations, this is hard to write ToR and has implications on the whole evaluation process and the evaluation as a whole could be a waste of time"* (Int 8).

This finding is insofar worrisome as several interviewees (Int 2,4,9) also confirm that betimes conflicts with evaluators are rooted in unclear terms. Moreover, they clearly emphasise that ToR are an instrument for reaching consensus about the purpose, objectives and tasks of an evaluation not only between the contractor and the evaluator but also among all other stakeholders involved, so that everyone has a clear vision of what can be expected from this exercise and what not. Without clear ToR apparently conflicts are inevitable, as the following quote suggests. *"Overall, it was a very negative experience and a huge disappointment for everybody. Next time I would define a clear process and how to communicate with the different units, find an agreement and then develop the ToR. It would have been good to reflect on this experience and draw lesson learnt, but they did not do that as they agreed not to talk about it anymore. So, lesson learnt got lost as this happens often, also due to the rotation realities."* (Int 2)

Collaboration with the evaluators and evaluators' competencies

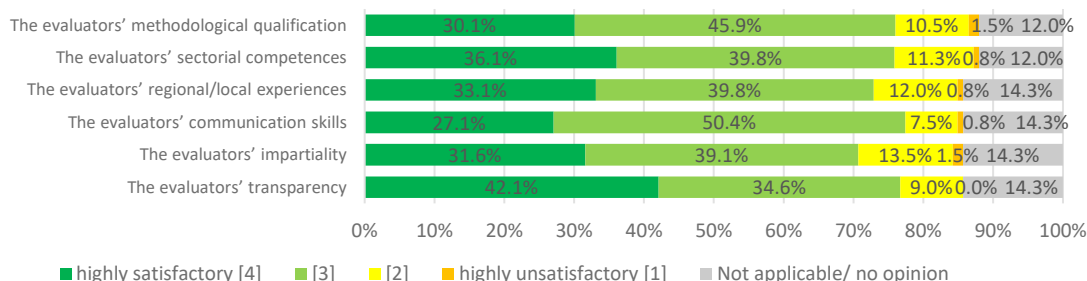
Moving on in the timeline of an evaluation, the next aspect relates to SDC staff's satisfaction of their collaboration with evaluators. The online survey results provide an overall positive picture. Except for the evaluator's flexibility and the assistance provided to them, roughly three quarter of the respondents are rather or highly satisfied with all related aspects such as the evaluators' availability, responsiveness, as well as their communication and management of the evaluation.

Figure 11: Satisfaction with the collaboration with the evaluators



Likewise, the evaluators' competencies are rated positively across the board. With approximately 70 to 80 percent rather or highly satisfactory ratings, all aspects are scored remarkably well by the survey respondents. Particularly, the evaluators' communication skills and their transparency are appreciated by the majority.

Figure 12: Satisfaction with the evaluators' competencies



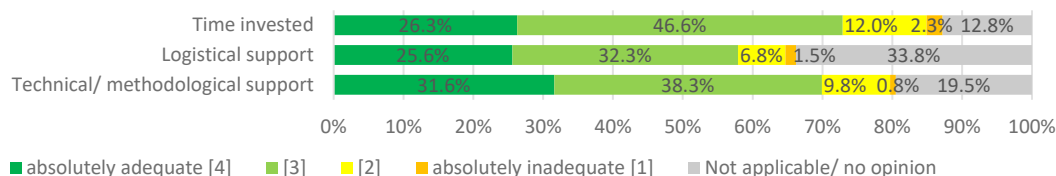
Interesting to note is that the majority of the respondents of the interviews (Int 2-5,8-10) emphasise the lack of specific competencies of the evaluators they have worked with especially regarding their analytical skills, sectorial expertise and regional experience. As a result, the respondents had to invest more time than planned in communication and revising reports, as the following quote illustrates. *“I had to jump in many times, and sometimes I was almost writing the report.”* (Int 2) Thereby, it is also recognised that intervening too much may compromise the evaluators' independence (Int 2,5,8).

The comparatively positive assessment of the evaluators' methodological competencies by SDC staff in the survey also somewhat contradicts the experts' assessment as outlined in Chapter 3.4. In sum, the results on this aspect are not fully conclusive, which (at least partially) may also be ascribed to the poor representativeness of the samples. It might also be that in some cases survey respondents are not fully capable of assessing the evaluation methodology as this requires respective in-depth knowledge. In any case they indicate the need to elaborate this issue further.

SDC's staff own efforts

The last aspect discussed under the heading of process use is the effort made by the SDC staff during the implementation of the evaluation. The survey participants were asked to assess the adequacy of their own support for the evaluation in terms of time invested, logistical support as well as technical and methodological support provided. At least 70 percent of the respondents considered their efforts as rather or absolutely adequate, yet, apparently a third of the respondents did not provide any logistical support.

Figure 13: Adequacy of own support for the evaluation



In order to quantify one's own support to the evaluation, the respondents were further asked to estimate the number of working days (WD) they have invested. It turned out that on average SDC staff invests 12,4 WD for managing an evaluation. This compares well with experiences from other donor organisations, which lie in the same range.

Taking a closer look on the duties the time was spent for, the interviews indicate that most effort was put in preparatory tasks (e.g. writing/ reviewing ToR, selecting evaluators, stakeholder meetings), as well as providing backstopping and quality assurance (e.g. revising reports, participating in briefing and debriefing workshops). Thereby, it appears that the time invested varies quite a lot among the interviewees and after all depends on the evaluators'

competencies. The majority of the interviewees agrees (Int 3,5,6,8-10) that the required time is systematically underestimated, deadlines are often too tight with feedback loops, reviews or drafting management responses not being sufficiently considered. Eventually, these shortcomings lead to a poor exploitation of the potential of evaluations as an instrument for evidence-based decision-making.

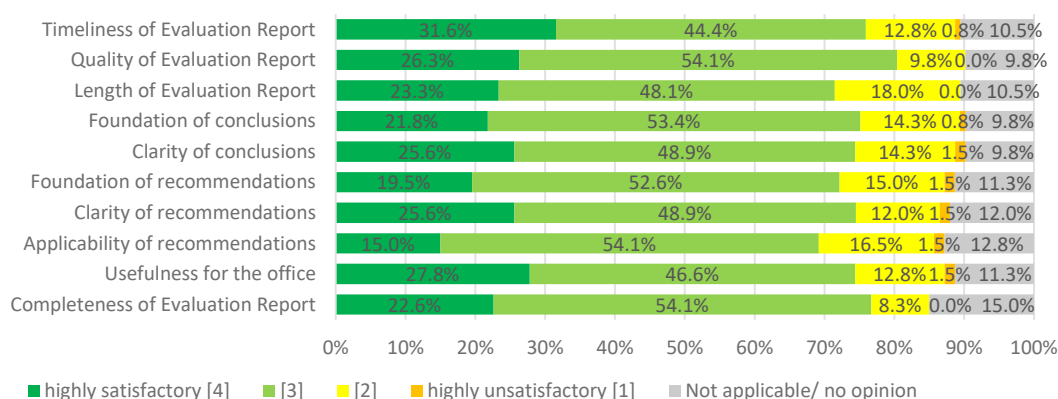
7.8.2 Usability of Decentralised Evaluations

For assessing the usability of evaluations, SDC staff was asked about its **satisfaction with the evaluation report**, the **actionability of recommendations**, their **usefulness for particular purposes** and the overall **added value of the evaluation**.

Satisfaction with the evaluation report

The online survey results suggest that most respondents are quite satisfied with the evaluation reports they received. Only the length of the evaluation report, the applicability and the foundation of recommendations were rated somewhat less positive. These findings contradict the rather mediocre expert rating of the decentralised evaluation reports, as well as the assessment of most interview respondents. What is worth mentioning is the fact that although the analysis shows severe shortcomings of the reports, those commissioning the evaluation consider the evaluation report as useful for their office. Nevertheless, the findings once more indicate the urgent need of Evaluation Capacity Development among SDC staff.

Figure 14: Satisfaction with evaluation report



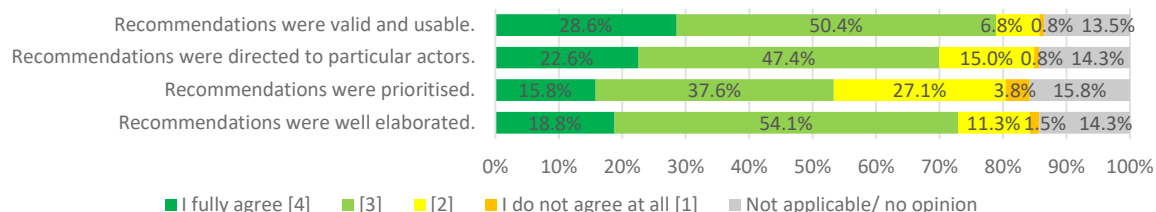
The interviews reveal some explanations for the contradictory assessments: Some (Int 1,4,6,9) interviewees deplore that evaluators tend to go the easy way and collect data selectively (e.g. only positive assessments, only feedback from particular stakeholders such as government institutions). Accordingly, evaluation reports are deemed to contain predominantly information that is already known by the project staff. Another danger is seen in biased recommendations caused by project managements and other stakeholders compromising the evaluations independence by forcing the evaluators to consider recommendations they 'want' to receive. Finally, the lacking tangibility of recommendations is criticised by a number of interviewees (Int 2,3,5,6,9), which leads to project staff having a hard time to translate what has been recommended in the evaluation report into concrete actions.

2 interviewees stated the following main factors highly influencing the quality of the report (INT 7, 8): A short presentation to give feedback to the government partners; a graphic on one page to summarise the main findings; standardised rules and requirements: max. 20 pages plus annexes; Executive Summary, max. number of recommendations, target-group specific, prioritised, focused and time-bound. Since they obligatory demand those factors, the reports became more useful for the office as well as for the partners.

Actionability of recommendations

In contrast to the interviewees' perceptions, the vast majority of the survey respondents assess the recommendations as valid and usable. Also, the elaboration and alignment of recommendations to particular actors is rated mostly positively. Merely the prioritisation of recommendations leaves room for improvement with only slightly more than half of the respondents rather or fully agreeing with the respective statement.

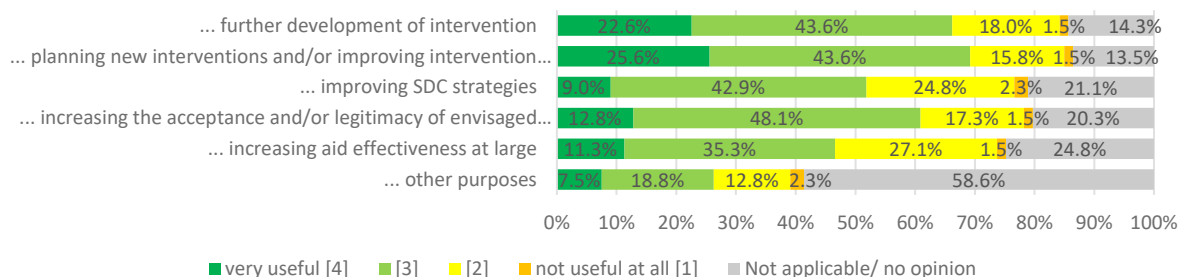
Figure 15: Actionability of recommendations



Usefulness of recommendations

Eventually, coming to the perceived usefulness of recommendations for particular purposes, the survey respondents' feedback is once again very positive but somewhat more differentiated. Recommendations are regarded most useful for planning new interventions and improving intervention approaches (69.2%) as well as for developing an already running intervention further (66.2%) and for increasing the acceptance or legitimacy of foreseen changes (61%). Merely about half of the respondents see recommendations most useful for improving SDC strategies or increasing aid effectiveness at large. Other purposes are mentioned only by very few respondents: informing or getting into discussion with project stakeholders, promoting or advocating for beneficiaries' needs, increasing comprehension at HQ level for the perspective of field office or attracting additional funding.

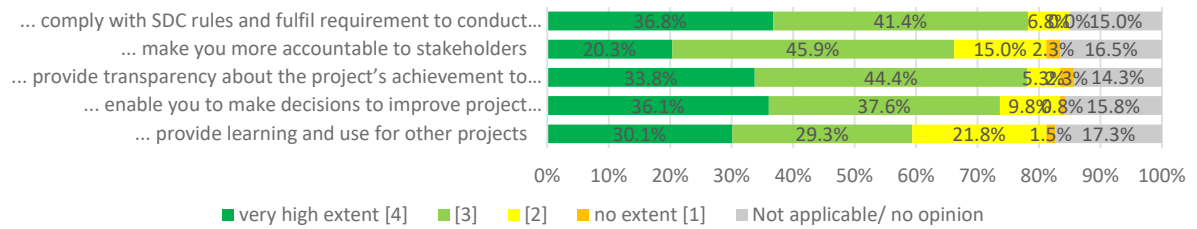
Figure 16: Usefulness of recommendations for particular purposes



Added value of the evaluation

A likewise heterogeneous picture yields the question about the added value of the evaluation at large. Whereas almost eight out of ten online survey respondents award the evaluation to provide transparency about the project's achievements to a high or very high extent, considerably less but still more than half state that it actually provided learning and use for other projects. The survey respondents recognise the usability of evaluations, although to different extent regarding generating knowledge, enabling organisational learning, providing for accountability, as well as providing the basis for decision-making.

Figure 17: Added value of the evaluation



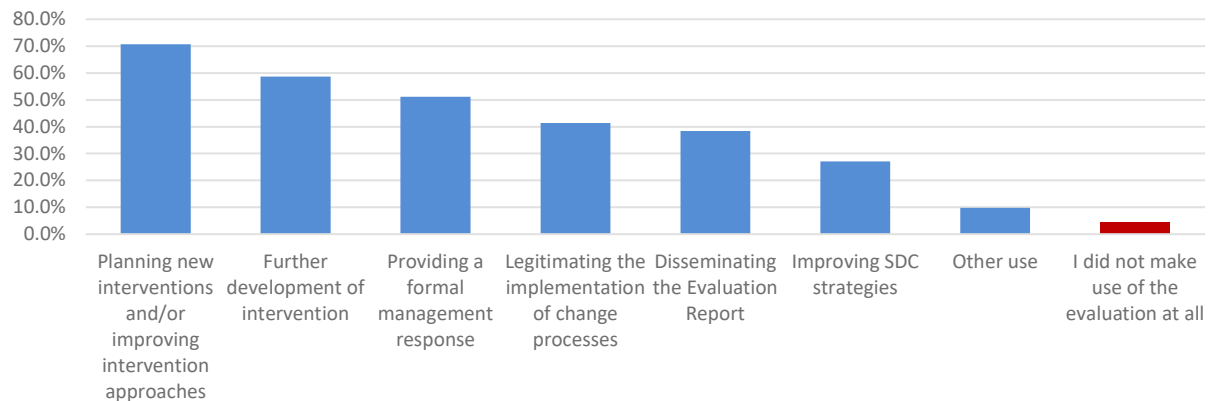
7.8.3 Use of Evaluation Results

This section focuses on the use of evaluation results, starting with SDC staff's feedback about **for what they use recommendations** and with **further evaluation outcomes** are relevant to them. Furthermore, the **limiting factors** are discussed as well as the **options available to facilitate the use of evaluations**.

Use of recommendations

Looking first at the use of the evaluations' recommendations, the survey provides a clear picture, which is also in line with the perceived usability as discussed in the previous section. Most respondents use them for planning new interventions and improving intervention approaches, for the further development of a current intervention as well as for preparing a formal management response. In contrast, for legitimising change processes, disseminating the evaluation report and improving SDC strategies recommendations appear to be somewhat less important. It is interesting to note that anyhow almost five percent state to not having made use of recommendations at all.

Figure 18: Actual use of recommendations



Once more the interviews provide justifications for these quantitative figures. Most interviewees (Int 4-7,9,10) confirm to use evaluations for deciding on how to continue with a project, particularly when entering a new phase. Therefore, new topics and ideas (e.g. for activities, products), good practices and lessons learnt from the evaluated project but also from others, as well as suggestions on project management and steering are most appreciated. As regards the dissemination of the evaluation report, apparently the opinions differ among the interviewed SDC staff. While two interviewees highlight its persuasive potential (Int 4,6), another one challenges the meaningfulness of publishing evaluation reports as it would take very much effort to put it into an appropriate, in terms of diplomatic, language for the different stakeholders, e.g. the partner government (Int 8).

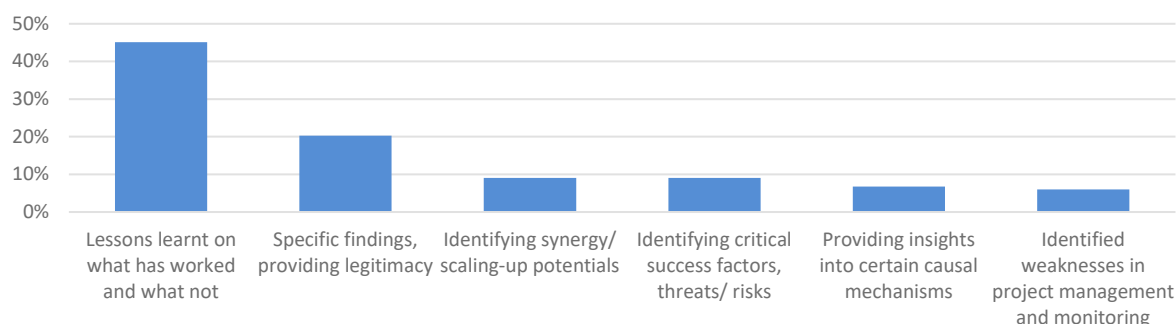
Relevant evaluation outcomes

Adding to the above-mentioned uses of recommendations deriving from evaluations, SDC staff was further asked what (further) outcomes are relevant for them. Apparently, lessons learnt are by far the most important aspect in this regard. Thereby, the respondents indicate to particularly appreciate necessary changes being highlighted or current approaches being confirmed. Providing legitimacy by outlining specific findings, for instance in which areas the project was particularly successful, showing beneficiaries' perspective, confirming of goal achievement, or providing information about expectable sustainability, is received well by a fifth of the respondents. Further outcomes which are mentioned to be relevant comprise (a) the identification of synergies and scaling-up potentials, (b) of critical success factors for impact achievement and sustainability, (c) providing insights into and suggesting improvements for the intervention logic and the adequacy of its operationalisation (e.g. developing better results indicators) and (d) identifying weaknesses in project management and monitoring (e.g. 'forgotten' stakeholders, lacking baseline data).

The following changes due to evaluations, clustered around two main areas, are reported by the survey respondents:

- (a) Adaptation of the intervention: adapting the project's exit strategy, adding, changing or stopping activities, revising budget allocations, goal concepts and implementation design, changing staffing, priorities, institutional setups, establishing linkages between components and a monitoring system.
- (b) Secondary effects of the evaluation's implementation: changes in attitudes, awareness and ownership of intervention staff, partners and further stakeholders to the evaluation, increased interest by other donors to collaborate and replicate activities, increased efforts in revising the intervention's terminology, improving its accountability, conducting research studies on specific, intervention-related topics.

Figure 19: Relevant evaluation outcomes



Limiting factors for the use of evaluation results

In order to find out what hampers the use of evaluation results, including the implementation of their recommendations, SDC staff was asked about impeding factors. Comparatively few responses make suggestions on this aspect, resulting in only a few factors to be identified:

- ✓ Insufficient comprehensible results, lacking clarity, persuasiveness and actionability of recommendations (mentioned by 13.5% of the survey respondents)
- ✓ Superficial, inadequate and/ or incomplete analysis and lacking understanding of the intervention by the evaluators as well as lacking objectivity (12.8%)
- ✓ Resistance of stakeholders to adopt changes and unsupportive regulations (e.g. at strategic level, superiors, partners, shifting political priorities, bureaucratic hurdles) (12.0%)
- ✓ (Unforeseen) Changing or deteriorating of intervention framework making recommendations obsolete (5.3%)

Further factors mentioned by only a few (≤ 3 respondents) comprise an inapt timing of the evaluation (e.g. conducted too late, when no further activities are planned), an inadequate evaluation framework (i.e. unsuitable evaluation questions, criteria), lacking involvement of relevant stakeholders in the evaluation and finally – for the later evaluations in the sample – the restrictions caused by the CoViD19 pandemic.

Interestingly, the interviews yield yet another perspective on impeding factors, which are not so much related to the evaluations themselves but rather conceived as systemic issues rooted in SDCs organisation culture. The factors can be summarised under three head-words: (a) Fear to failure, (b) lacking learning interest and (c) missing commitment. As regards (a) some interview respondents (Int 6,7,10) believe that within SDC dominates an atmosphere of fear to fail resulting amongst others in the avoidance of external evaluations. Concerning (b) one interviewee (Int 1) supposes that many perceive evaluations rather as a tool for providing accountability than for institutional learning and further development. As long this is the case the interviewee doubts that any progress will be made in making better use of evaluations, no matter what technical measures are taken. In this regard, it is also mentioned that such a change in culture would need to be introduced top-down, as only then NPOs and further staff would adopt necessary changes, which finally refers to (c). Some (Int 6,8) see the main challenge in little commitment, particularly at HQ level, or as one interviewee brings it to the point: *“RBM is not in the focus of SDC anymore. This can be seen since a couple of years. Before, RBM was clearly wanted by the director. Now it is rather a question of surviving politically. There is no vision. As long as it is not clear where SDC should aim for, in which direction to go, it is difficult to think about how and for what evaluations could be used for. There is pressure on the four topics of the strategy and the Erfolgsquoten have superimposed it.”* (Int 6)

Facilitation of use

Finally, in the online survey SDC staff was asked what they think could be done in order to foster the use of evaluation results. The suggestions mirror to a certain extent the required support.

- ✓ **Learning and knowledge sharing platform:** Around a fifth of the survey respondents considers a platform for learning exchange, aggregating findings from evaluations and knowledge management as beneficial. Such a platform would enable sharing results of, and experiences made during evaluations with each other and further stakeholders. It is further supposed that therewith, evaluation results could be fed into the design of new interventions and the further development of SDC policy more systematically. Eventually, some suggest exploiting evaluations for developing learning material for SDC staff trainings.
- ✓ **Changes in attitude:** About an eighth regards a fundamental change of SDCs attitude towards evaluation as essential, as already laid out above by interview statements on limiting factors of evaluation results. Instead of understanding evaluation as a “punishment” and “additional effort” it should be taken as a learning exercise providing its stakeholders to opportunity to learn and improve their work. This would also require more intensive stakeholder involvement (already during the planning of an evaluation) and openness to criticism by all involved in such an exercise.
- ✓ **Usability as focus in all evaluation stages:** Provision of stakeholder specific reports (instead of one large evaluation report), applying better adapted evaluation criteria, engaging more suitable evaluators, clarify expectations about the evaluation with all involved stakeholders, reducing the scope of evaluations to fewer key questions, and only “really go where it might hurt”.

Most of these suggestions are also reflected in the interview findings. Moreover, the interviewees raise further aspects like the necessity of a better structured management response (Int 4-7,9) with a clear set out follow-up plan for implementing recommendations including timelines and responsibilities. Another suggestion refers to a more systematic

capitalisation of evaluation results (Int 1,2,4,7,10) by planning learning and dissemination strategies beforehand, as well as aggregating findings from several evaluations in a sector or in particular contexts. Eventually, exchanging experiences within SDC and with other Swiss Development actors such as SECO or the embassies and conducting joint evaluation are regarded as further learning opportunities.

7.8.4 Support Needs and Quality of Support

As the focus of the Quality Assessment is also to assess the adequacy of the current support provided to SDC staff for managing evaluations and to provide recommendations for improving it, the received **support** and **its perceived quality** as well as further needed support measures are investigated. Once again, perceptions and information by the survey and the interview respondents are provided in the following.

Received support

The vast majority of the survey respondents (86.5%) searched for support along the way of planning and managing an evaluation:

- written guidance (e.g. handbooks, checklists) (75.2%)
- personal counselling (32.2%)
- external support (18.8%)
- only very few favor an in-house training (7%)

Approximately nine out of ten (88.7%) who seek for support, actually find it as follows:

As regards **written guidance**, for most survey respondents this is the SDC field handbook (40.3%), followed by other publicly available guidebooks and papers (32.7%). Other SDC strategic and policy documents and the How-to Note, appear to play a comparably lesser role (each only about 18%). Interviewees confirm that guidelines and tools are not known well, and they suspect that many are not aware of them, including the How-to Note.

When looking for **personal support**, most respondents state to approach superiors (44.9%) or other knowledgeable colleagues (34.7%). Only few (10.2%) request support directly from the E+C unit or external experts. Further support is found in local networks, in other evaluations or consultants.

Again, the survey results match quite well the findings from the interviews during which knowledgeable colleagues were mentioned as the primary source of support (Int 1-5, 7-9). It is, however, also mentioned that it can be quite difficult to identify the 'right' expert, as the following quote indicates. *"SDC is a jungle, if you know the people, then you find the right person who can give support for your evaluation. If you are young and new, it is difficult. Knowing the people is very helpful."* (Int 4)

The survey results indicate a **high satisfaction of SDC staff with the support they receive** (36,3% rates it highly helpful and 46% as rather helpful). Considerably less assess the support as rather not helpful. All in all, the respondents from the survey as well as from the interviews are more satisfied with personal guidance than with written material.

How-to Note Evaluation

The majority of the survey respondents perceive the How-to Note as a very helpful tool, with the vast majority rating it very (35.2%) or at least rather (41.8%) helpful. However, while no one questions its helpfulness in general apparently 15 percent does not have an opinion about it. Nevertheless, given the fact that, as mentioned above, it is only mentioned at all by 18 percent as a relevant source of information, these values are actually better than expected.

Asked about which contents of the How-to Note are considered most helpful, the majority of the respondents (40.9%) refers to the practical guidance on the individual steps of an evaluation. Practical examples/ templates, tools and links to further resources (13.6%), information on how to write a ToR (12.1%) and the provision of evaluation questions (10.6%) are regarded as particularly helpful too. Nevertheless, a number of respondents (16.7%) also can think of improvements of the document. As regards its content some wish a clearer terminology (e.g. review vs. evaluation), adding suggestions on how to find suitable evaluators, an improved part on impact evaluation, the provision of more practical examples and links to further resources, more guidance on contracting and scoping an evaluation as well as on different roles and responsibilities of evaluation stakeholders. Concerning its format a few (6.1%) find it too long and would prefer an interactive format instead of one large document.

Again, the interview findings complement the survey results by on the one hand confirming the overall positive assessment of the document but admitting that they used it only once, and on the other hand by providing further suggestions on its improvement. In particular, including further guidance on stakeholder involvement, budget and time planning, structuring an evaluation report in a way that it fulfills the information needs of its addressees, as well as on phrasing and prioritising recommendations is mentioned in this regard. Furthermore, the interviewees would appreciate references to best practice examples and recommendations on who should be involved in the management response and follow-up process in the How-to Note. Eventually, some assess it as most useful for new and unexperienced staff (Int 1,2,4,7,9).

Useful support measures

Returning to the survey, the respondents were asked what else could be done in order to support evaluation planning and management. Unfortunately, a close-ended question with the categories “online helpdesk”, “online video tutorials”, “virtual trainings”, “face-to-face trainings” and “written manuals and guidelines” did not shed much light on most urgent needs, as the entries feature insufficient variance, as the following figure shows.

Figure 20: Preferred type of support



Additionally, the respondents suggest the following support:

- ✓ more involvement and supervision by superiors (28.1%)
- ✓ online toolbox with exemplary ToR, evaluation questions, report templates and criteria for assessing the quality of an evaluation (28.1%)
- ✓ establishing an expert database with profiles and experiences by sector, region, topic etc. (15,6%)

- ✓ a repository of good and bad evaluations (12.5%)
- ✓ an exchange platform for field offices to provide mutual support (9.4%)
- ✓ general guidance on implementing evaluations and making use of them (21.6%)
- ✓ personal exchange with colleagues about scoping an evaluation or formulating evaluation questions (16%)
- ✓ provision of practical tools, templates or examples (e.g. for ToR) (16%)
- ✓ sharing experiences from other evaluations (11%)
- ✓ only very few wish for specific methodological trainings (5%).

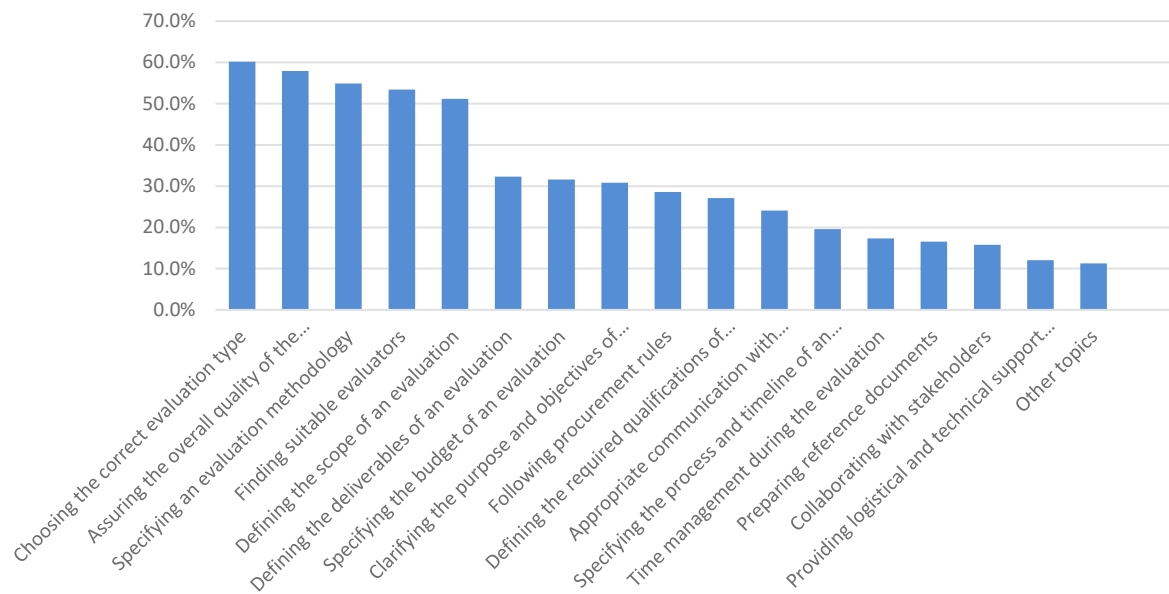
Once more the findings from the interviews coincide with the survey results and yet specify concrete support measures. The following topics are mentioned to be of particular use by many interview respondents (Int 1-6,8,10):

- ✓ Templates for ToR in specific fields of intervention such as for instance Humanitarian Aid, including a standard glossary of general technical as well as SDC specific terms, and a guideline for drafting management responses with timelines and to-do lists
- ✓ An obligatory on-the-job training program for all being assigned to commission an evaluation with concrete recommendations for instance for stakeholder involvement, communicating evaluation results and conducting evaluations in fragile contexts, including a refresher course (e.g. every two years)
- ✓ An online accessible repository for evaluators and evaluations by region and sector, including information on experiences made with these (e.g. what worked and what not, success and failure stories, value added, overall quality of the evaluation, respectively qualification of the evaluator)
- ✓ A mentoring network or focal points with more experienced SDC staff that provide case-based support to evaluations of younger colleagues
- ✓ A regulatory framework for deciding on the implementation of an evaluation, chose its type, who to involve, what to do, who is responsible for what

Specifically asked about the perceived added value of an online helpdesk, only two interviewees (Int 2,4) consider this as a meaningful option, at least not if it is being managed by a person or a small team. Doubts prevail that a small group will be able to keep the overview and provide competent support on the multitude of topics related to evaluation. Accordingly, such a facility is rather deemed as an entry point for accessing further support by other experts or data sources.

Finally, looking at specific topics of support provided through trainings, mentoring or any kind of technical measures, the online survey points to six priorities of the SDC staff, that were also mentioned before either under suggestions for improving the How-to Note or when referring to further requested support measures. As the following figure shows, clearly more than half of the respondents wish to learn how to define evaluation questions, chose the right type of evaluation, assure its overall quality, specify its methodology and scope, and find suitable evaluators.

Figure 21: Support topics



All other topics are mentioned considerably less often. For proving the validity of this result, it is worth to mention that the ranking of the topics does not correlate with the order of items in the questionnaire. So, a primacy or recency effect⁶ can be ruled out.

7.9 References to appendices

- Appendix 1: Online survey questionnaire
- Appendix 2 & 3: Survey raw data and further tables and figures

⁶ Primacy and recency effects describe an effect caused by the fact that humans better recall information presented in the beginning or at the end of a longer list (approx. > 5) e.g. of items, objects or properties (cf. e.g. <https://www.sciencedirect.com/topics/psychology/recency-effect>). This can lead to biased results during an interview or in a questionnaire. For identifying this bias, the data of such type of questions needs to be checked for partial correlation with the item position.

Imprint

Publisher:
Federal Department of Foreign Affairs FDFA
Swiss Agency for Development and Cooperation SDC
3003 Bern
www.sdc.admin.ch

Pictures Cover page: © SDC

Orders:
E-mail: info.deza@eda.admin.ch

Specialist contact:
Swiss Agency for Development and Cooperation SDC
Evaluation and Controlling
Freiburgstrasse 130, 3003 Bern
deza.evaluation-controlling@eda.admin.ch

This publication can be downloaded from the website
<https://www.eda.admin.ch/deza/en/home/results-impact/berichte/evaluationsberichte.html>

Bern, June 2023