



2020:2

Sida Studies in Evaluation

NIRAS Sweden AB

Meta-study of Partner-led Evaluations 2019

Final report

The views and interpretations expressed in this report are the authors' and do not necessarily reflect those of the Swedish International Development Cooperation Agency (Sida).

Authors: Cecilia Ljungman, Jonas Norén, Matilda Svedberg, Christina Paabøl Thomsen

Sida Studies in Evaluation 2020:2

Copyright: Sida

Date of final report: 2020-08-14

Published by Nordic Morning 2020

Art. no. Sida62323en

urn:nbn:se:sida-62323en

This publication can be downloaded from: <http://www.sida.se/publications>

SWEDISH INTERNATIONAL DEVELOPMENT COOPERATION AGENCY

Address: SE-105 25 Stockholm, Sweden. Office: Valhallavägen 199, Stockholm
Telephone: +46 (0)8-698 50 00. Telefax: +46 (0)8-20 88 64
E-mail: info@sida.se. Homepage: <http://www.sida.se>

Table of contents

| | |
|---|-----------|
| Table of contents | 7 |
| Abbreviations and Acronyms | 9 |
| 1 Introduction..... | 1 |
| 1.1 Background | 1 |
| 1.2 The assignment..... | 1 |
| 1.3 Sampling | 2 |
| 1.4 Report structure..... | 2 |
| 2 Quantitative Assessment..... | 3 |
| 2.1 Introduction and methodology | 3 |
| 2.2 Swedish strategy coverage | 4 |
| 2.3 Sidas' financial contributions | 7 |
| 2.4 Types of Sida partners | 7 |
| 2.5 Types of interventions | 7 |
| 2.6 Sectors | 8 |
| 2.7 Geographic focus of interventions | 8 |
| 2.8 Type of evaluations | 9 |
| 2.9 Evaluation criteria | 9 |
| 2.10 Evaluation approaches | 10 |
| 2.11 Data collection methods | 10 |
| 2.12 Evaluation budgets and duration | 11 |
| 2.13 Summary | 12 |
| 3 Quality Assessment | 13 |
| 3.1 Introduction and methodology | 13 |
| 3.2 Overall assessment..... | 15 |
| 3.3 Language and form | 16 |
| 3.4 Methodology..... | 16 |
| 3.5 Context..... | 16 |
| 3.6 Analysis..... | 17 |
| 3.7 Terms of reference | 17 |

| | |
|--|-----------|
| 3.8 Summary and conclusions | 19 |
| 3.9 Reflections on the assessment framework | 19 |
| 4 Utility Assessment..... | 21 |
| 4.1 Introduction..... | 21 |
| 4.2 Dissemination | 22 |
| 4.3 Utility of the process and report..... | 23 |
| 4.4 The evaluation team | 29 |
| 4.5 Follow up | 33 |
| 4.6 Summary and conclusions | 35 |
| 4.7 Reflections on the utility assessment..... | 36 |
| 5 Potential correlations? | 37 |
| 5.1 Type of evaluation & evaluation quality | 37 |
| 5.2 Evaluation budget & evaluation quality | 38 |
| 5.3 Evaluation quality & evaluation utility..... | 38 |
| 5.4 Quality of ToR & evaluation quality..... | 39 |
| 5.5 Quality of ToR & evaluation utility | 40 |
| 5.6 Concluding reflections | 41 |
| Annex 1: Evaluation Metadata Parameters | 42 |
| Overview evaluation metadata..... | 42 |
| Detailed Evaluation Metadata | 43 |
| Annex 2: Quality Assessment Framework..... | 45 |
| Annex 3: Utility Questions | 52 |
| Annex 4: Overview of Sample Evaluations..... | 53 |
| Annex 5: Terms of Reference | 57 |

Abbreviations and Acronyms

| | |
|-------------|--|
| AMDT | Agricultural Markets Development Trust |
| AMQF | A. M. Qattan Foundation |
| CSO | Civil society organisation |
| DSA | Daily subsistence allowance |
| FAO | Food and Agriculture Organisation of the United Nations |
| FLEGT | Forest Law Enforcement, Governance and Trade programme |
| GBV | Gender-based violence |
| HRD | Human rights defenders |
| MDG | Millennium Development Fund |
| MJF | Manusher Jonno Foundation |
| MTE | Mid-term evaluation |
| MTR | Mid-term review |
| NMFA | Norwegian Ministry of Foreign Affairs |
| OECD/DAC | Organization for Economic Cooperation and Development/Development Assistance Committee |
| PAO | Party-affiliated organisation |
| PO | Programme Officer |
| PRCD | Peacebuilding and Reconciliation Through Community Dialogues |
| SCI | Save the Children International |
| Sida | Swedish International Development Cooperation Agency |
| Sida/CIVSAM | Sida's Department for Support to Civil Society |
| Sida/UTV | Sida Department for Evaluation and Internal Audit |
| SNV | Stichting Nederlandse Vrijwilligers (Netherlands Development Organisation) |
| SRHR | Sexual and Reproductive Health and Rights |
| ToC | Theory of change |
| ToR | Terms of reference |
| UN | United Nations |
| UNDP | United Nations Development Programme |
| UNEG | United Nations Evaluation Group |
| UNOPS | United Nations Office for Project Services |
| UN Women | United Nations Entity for Gender Equality and the Empowerment of Women |
| VAFF | Visual Arts – a Flouring Field project |
| ZOA | Name of Dutch CSO |

1 Introduction

1.1 BACKGROUND

Sida's partner organisations are encouraged to commission and manage their own evaluations of Sida-funded support if they are deemed by Sida to have adequate monitoring and evaluation (M&E) systems and sufficient capacity. Sida lacks systems to establish an overall view of evaluations undertaken by its partners. It therefore has insufficient understanding of the number, type, credibility, quality, and utility of partner-led evaluations.¹ This data is important to Sida to help strengthen the credibility, quality and use of evaluations. Sida's internal audit has recommended assessing the quality of partner-led evaluations, monitoring how the evaluations are used after completion, determining if a management response/action plan based on the recommendations has been elaborated and how the evaluation has been used to improve operations.

Sida, therefore, commissioned NIRAS to undertake a stocktaking and meta-assessment of partner-led evaluations² between April and June 2020 to serve as input into further development of Sida's systems for evaluation, evaluation quality assessment, and evaluation use.

1.2 THE ASSIGNMENT

The assignment consisted of three main components:

- **Component 1** consisted of a stocktaking of all partner-led evaluations and reviews.³ This involved identifying the potentially wide range of evaluations that were completed in 2019 or the beginning of 2020.⁴
- **Component 2** consisted of assessing the **quality** of a sample of 14 partner-led evaluations, using Sida's and OECD/DAC's quality standards as a frame of reference.

¹ Sida defines partner-led evaluations as evaluations commissioned and managed by a Sida cooperation partner that were directly or indirectly financed under a contribution agreement between Sida/Embassy and the partner. This includes specific budget lines for evaluation in the agreement or evaluations undertaken as part of core support agreements.

² The meta-study does not include evaluations of interventions by third parties that receive funds via Sida partners who redistribute Sida funds.

³ In this report, the term "evaluation" will be used most often, although it refers to both reviews and evaluations.

⁴ As outlined in the ToR, the study was initially expected to cover all partner-led evaluations conducted during 2018 and 2019. However, due to the consequences of the Corona pandemic (limited capacity and time of Sida staff due to i) staff relocations; and ii) efforts to realign Swedish development cooperation assistance according to the changing context), Sida decided to limit the scope to evaluations from 2019 and the first months of 2020.

- **Component 3** consisted of assessing the **utility** of the same sample of 14 partner-led evaluations. The utility assessment included determining use by whom and how; and the extent the evaluation has been disseminated, communicated, and followed up.

Each component required its own methodological approach. These are described in the beginning of the following three chapters, which are each dedicated to the three respective components.

1.3 SAMPLING

The quality and utility analyses (covered in Chapters 3 and 4) were based on a sample of evaluations that amount to 15 percent of the total dataset. The sample set was selected as follows: in consultation with Sida/UTV, the team selected 14 evaluations for deeper quality and utility analysis and assessment. The aim was to create a sample that was as mixed as possible. The following was taken into consideration when selecting the sample:

1. Sida requested that partner-led evaluations conducted by Swedish partners were not included in the sample, since Sida has a closer dialogue with these partners and knowledge of these evaluations.
2. The Sida staff (survey respondents) uploaded the:
 - a. Email address of the Sida staff member responsible for the partner agreement
 - b. Email address of a partner organisation representative
 - c. Terms of reference for the evaluation

The first two criteria were important so that the team could contact the stakeholders to discuss utility. The last criterion was needed for the quality analysis and assessment.

3. The evaluations submitted needed to have been completed by 30 September 2019 so that there was sufficient time for the evaluation to have been used.
4. The sample had to include a mix of Swedish strategies which they represented, with 60 percent being country strategies, and 40 percent global and regional strategies.
5. The sample had to include a mix of types of partner organisations.
6. The sample had to include a mix of sectors.

Of the 91 evaluations and reviews that make up the total dataset, less than 20 met criteria 1 to 3 above. The resulting sample consists of a satisfactory mix of evaluations in line with the criteria set above. However, had more evaluations met the criteria, an even more varied sample may have been possible. Annex 4 contains a short summary of the sampled evaluations.

1.4 REPORT STRUCTURE

This study consists of five chapters. The first parts of Chapters 2 through 4 provide a description of the methodology applied, including their respective limitations. Chapter 5 discusses possible correlations.

2 Quantitative Assessment

2.1 INTRODUCTION AND METHODOLOGY

This chapter maps out the partner-led evaluations that the evaluation team collected from 2019 and early 2020. Seventy-five evaluations were completed in 2019 and 16 were completed in 2020.

The task of quantitatively analysing partner-led evaluations involved collecting evaluation metadata in relation to 20 different parameters (such as type of evaluation, type of partner, type of sector – please see Annex 1). Devising a methodology to collect this data presented several challenges:

1. The dataset of partner-led evaluation was unknown.
2. The only way to compile and obtain an overview of all partner-led evaluations was to contact Sida heads of units/counsellors responsible for each of the 43 development cooperation strategies, together with relevant controllers.
3. Sida/UTV wanted an approach that minimised the time that Sida staff were required to devote to assist the data collection process.
4. Receiving evaluations by email would be a messy and unreliable approach that could cause confusion and lead to some evaluations being lost in the email chains.
5. Platforms, like SurveyMonkey, allow for very limited amounts of documentation to be uploaded.

The team, therefore, custom-built a data collection and analysis tool especially for this study using the open-source platform Docassemble⁵ and a range of other Python packages for data analysis and data storage. The approach provided a robust, effective, and user-friendly means of managing and organising large quantities of different types of data (survey data and documentation) for analysis.

The procedure consisted of the following:

1. Sida/UTV contacted relevant staff members by email requesting them to contribute to the study and forwarded a link to the study's data platform.
2. Relevant Sida/embassy staff members completed a short electronic survey for each partner-led evaluation that was completed in 2019 and early 2020 (see Annex 1).

⁵ Docassemble was originally designed for large class action suits in USA, which need a range of data in different forms at different times from different plaintiffs. It allows stakeholders to easily upload: survey responses, documents, multimedia files (videos, photographs with geo coordinates, recordings); while at the same time allowing for an interactive process between those supplying data and those demanding it.

3. At the end of the survey, the staff at Sida/embassies uploaded the evaluation and respective terms of reference (if the latter was not included in the evaluation document).
4. Once the documents had been uploaded, the NIRAS team reviewed the evaluations and entered metadata in relation to 16 additional parameters for every evaluation (see Annex 1). The data was entered in line with what was stated in the evaluation report.
5. NIRAS created a database in Excel that contains information about every evaluation received and includes hyperlinks to the reports.

Sida/UTV was consulted regularly during the development of the data collection tools to ensure that the parameters selected, and methods used met Sida's needs.

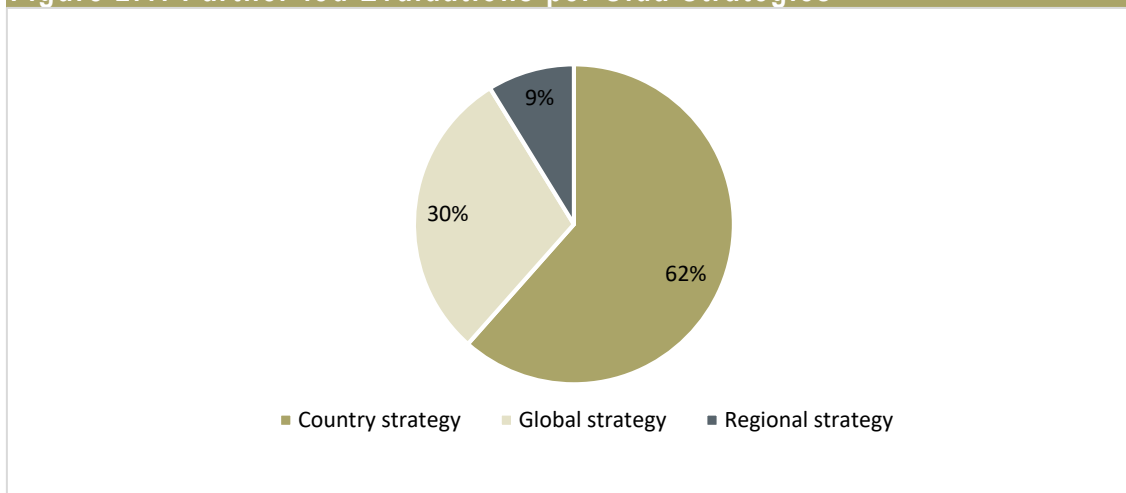
In all, 108 evaluations were uploaded. However, some were not partner-led (Sida-led), or were not actually evaluations (for example some were baseline studies or an audit reports). In some cases an inception report or PowerPoint were uploaded instead of an evaluation report. The final dataset includes 91 partner-led evaluations. Of these, 77 evaluations included terms of reference. For analysis and easy access to the data, a dashboard was subsequently created using various data science techniques.

The following sections provide an overview of the dataset in relation to 11 different parameters. These include Swedish strategies, financial contribution from Sida; types of partners; types of interventions; sectors; geographic focus; types of evaluations; evaluation criteria; evaluation approaches; data collection methods; and evaluation budget and duration.

2.2 SWEDISH STRATEGY COVERAGE

Sida's aim was that the study encompasses all partner-led evaluations undertaken under each of Sweden's 43 development cooperation strategies. Despite several reminders from Sida/UTV and the evaluation team, Sida staff were only able to upload evaluations conducted under 27 of the 43 current strategies.

Figure 2.1: Partner-led Evaluations per Sida Strategies



No partner-led evaluations were conducted under two strategies (Iraq and KAPAME). It is possible that some of the other strategies also had no evaluations to upload, but since there was no response, it is not possible to conclude. It was confirmed that partner-led evaluations had been undertaken by partners supported by the Strategy for Support through Swedish Civil Society Organisations. However, given the limited timeframe and the challenge of determining which of these evaluations met the Study's criteria, Sida's Unit for Support to Civil Society replied that it was not in a position to participate in the survey and upload the necessary documentation. Nevertheless, when counting the strategies known to not have partner-led evaluations in 2019, 67 percent of the total strategies are represented, 73 percent of the country strategies, 67 percent of the regional strategies, and 55 percent of the global strategies. The study team deems that the number and variety of evaluations received provides a decent dataset for analysis.

| Strategy | Number | Percent |
|---|--------|---------|
| 1. Afghanistan 2014-2019 | 4 | 4.4% |
| 2. Bangladesh 2014-2020 | 4 | 4.4% |
| 3. Bolivia 2016-2020 | 1 | 1.1% |
| 4. Burkina Faso 2018-2022 | 1 | 1.1% |
| 5. Cambodia 2014-2018 | 3 | 3.3% |
| 6. Capacity development, partnerships and methods that support Agenda 2030 for sustainable development (KAPAME) 2018-2022 | 0 | 0% |
| 7. Colombia 2016-2020 | 4 | 4.4% |
| 8. Democracy support through Swedish party affiliated organisations (PAO) 2016-2020 | 3 | 3.3% |
| 9. Environmental sustainability, sustainable climate and oceans, and sustainable use of natural resources 2018-2022 | 6 | 6.7% |
| 10. Ethiopia 2016-2020 | 4 | 4.4% |
| 11. Guatemala 2016-2020 | 2 | 2.2% |
| 12. Human rights, democracy and the rule of law 2018-2022 | 7 | 7.8% |
| 13. Iraq 2017-2021 | 0 | 0% |
| 14. Liberia 2016-2020 | 5 | 5.6% |
| 15. Mali 2016-2020 | 3 | 3.3% |
| 16. Mozambique 2015-2020 | 1 | 1.1% |
| 17. Myanmar 2018-2022 | 1 | 1.1% |
| 18. Palestine 2015-2019 | 7 | 7.8% |
| 19. Regional Asia and Oceania 2016-2021 | 1 | 1.1% |
| 20. Regional Eastern Europe, the Western Balkans and Turkey 2014-2020 | 5 | 5.6% |
| 21. Regional Sexual and Reproductive Health and Rights (SRHR) Strategy in Sub-Saharan Africa 2015-2019 | 1 | 1.1% |
| 22. Regional Sub-Saharan Africa 2016-2021 | 1 | 1.1% |
| 23. Somalia 2018-2022 | 3 | 3.3% |
| 24. Sustainable economic development 2018-2022 | 9 | 10.0% |
| 25. Sustainable peace strategy 2017-2022 | 2 | 2.2% |
| 26. Tanzania 2013-2019 | 6 | 6.7% |
| 27. Uganda 2018-2023 | 3 | 3.3% |

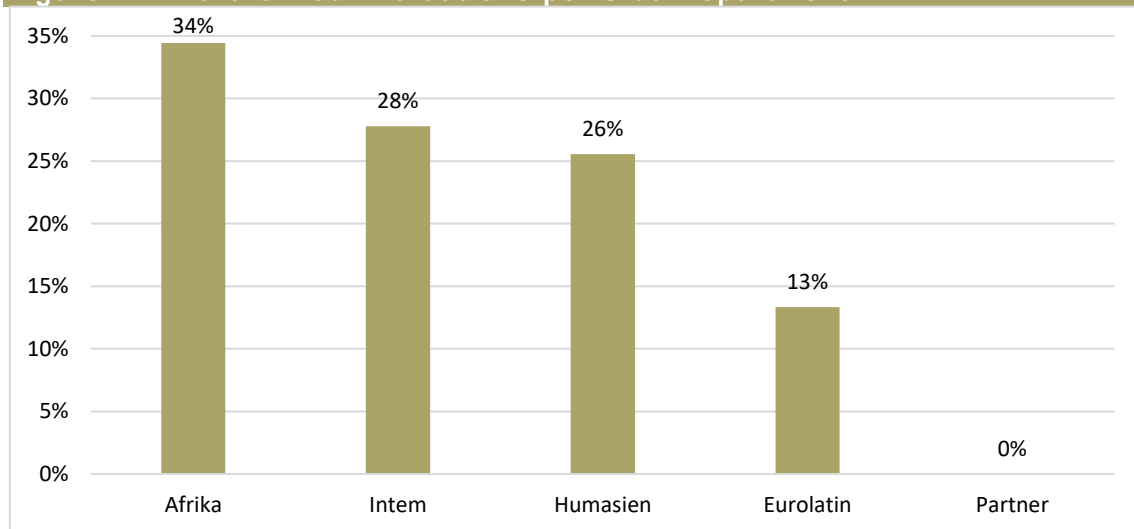
| | | |
|------------------------|---|------|
| 28. Zambia 2018-2022 | 2 | 2.2% |
| 29. Zimbabwe 2017-2021 | 1 | 1.1% |

No evaluations were received in relation to the following strategies:

1. Cuba 2016-2020
2. Democratic Republic of Congo
3. Kenya 2016-2020
4. Russia 2020-2024
5. Rwanda
6. South Sudan 2018-2022
7. Sudan 2018-2022
8. Regional Middle East and North Africa (MENA)
9. Regional Syria Crisis 2016 – 2020
10. Gender equality and women's and girls' rights 2018-2022
11. Humanitarian aid provided through Sida 2017-2020
12. Research cooperation and research in development cooperation 2015 - 2022
13. Sustainable social development 2018–2022
14. Swedish Civil Society Strategy (CIVSAM).

Just under two-thirds of the evaluations in the dataset have been undertaken through contributions via country strategies. The figure below shows the distribution of partner-led evaluation by Sida department.

Figure 2.2: Partner-led Evaluations per Sida Department



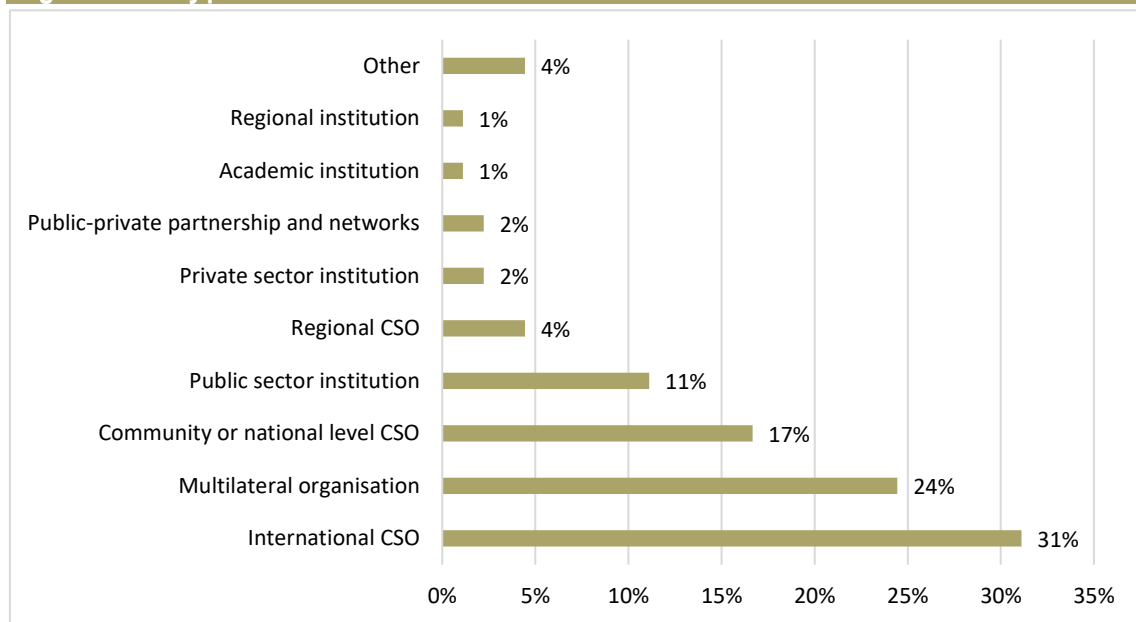
2.3 SIDAS' FINANCIAL CONTRIBUTIONS

The database includes information regarding the financial size of Sida's contribution for 90 percent of the evaluations in the dataset. The average financial size is SEK 58.5 million and the median size is SEK 43.2 million.⁶

2.4 TYPES OF SIDA PARTNERS

The most common type of partners represented in the evaluation dataset are international CSOs (31%), multilateral organisations (24%) and community or national level CSOs (17%).

Figure 2.3 Types of Sida Partners

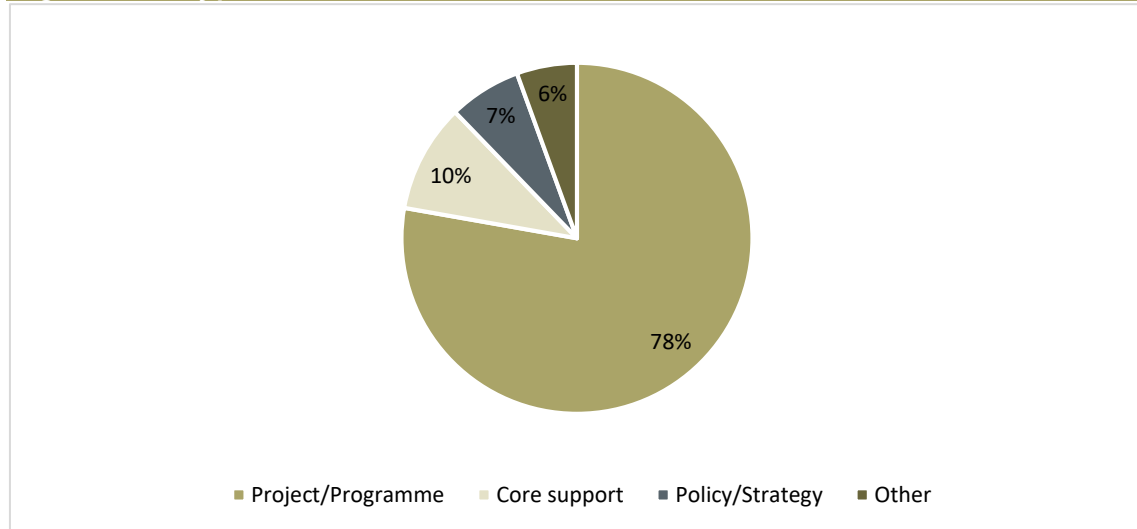


2.5 TYPES OF INTERVENTIONS

The largest type of intervention evaluated are projects and programmes. In total, 78 percent of the dataset represent project/programme evaluations. Evaluations undertaken by partners that receive core support amounts to 10 percent of the dataset, while policy/strategy evaluations amount to seven percent (had evaluations under the Strategy for support through Swedish Civil Society organisations been included in the dataset, the figure for evaluations by partners receiving core support would have been higher). The remaining six percent are comprised of other types of interventions.

⁶ The team has not prepared a diagram for this data. We are unsure of the reliability of the financial entered by Sida staff. Seven "0"s have been entered and nine contributions have been entered as being less than SEK 2 million. Four contributions have been entered as being over SEK 200 million.

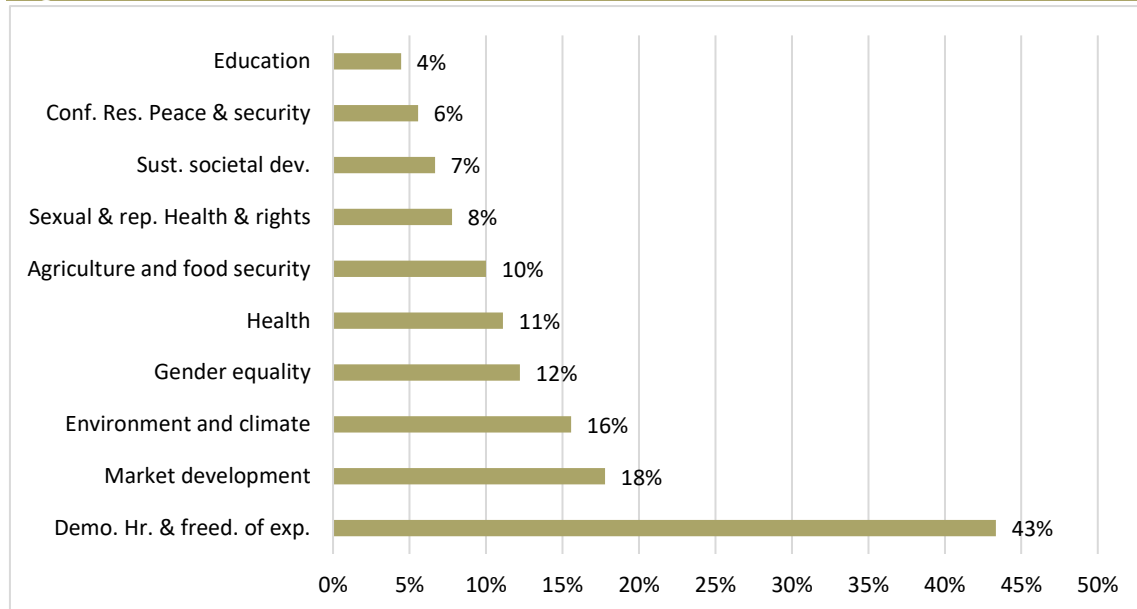
Figure 2.4: Types of interventions



2.6 SECTORS

The democracy, human rights and freedom of expression “sector” is the most common sector represented in the dataset (43%). Other relatively large sectors are market development (18%) and environment and climate (16%).

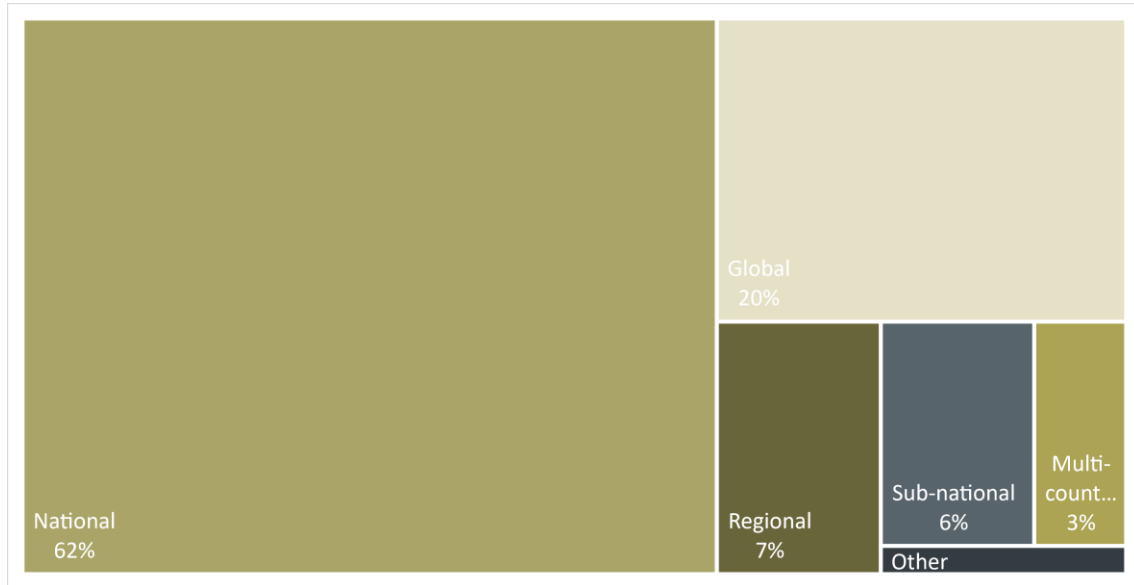
Figure 2.5: Sectors



2.7 GEOGRAPHIC FOCUS OF INTERVENTIONS

Sixty-two percent of the evaluated contributions had a national focus. One-fifth (20 %) had a global scope, seven percent had a regional focus, six percent had a sub-national focus, three percent are multi-county evaluations and the remaining one percent were labelled as “other.”

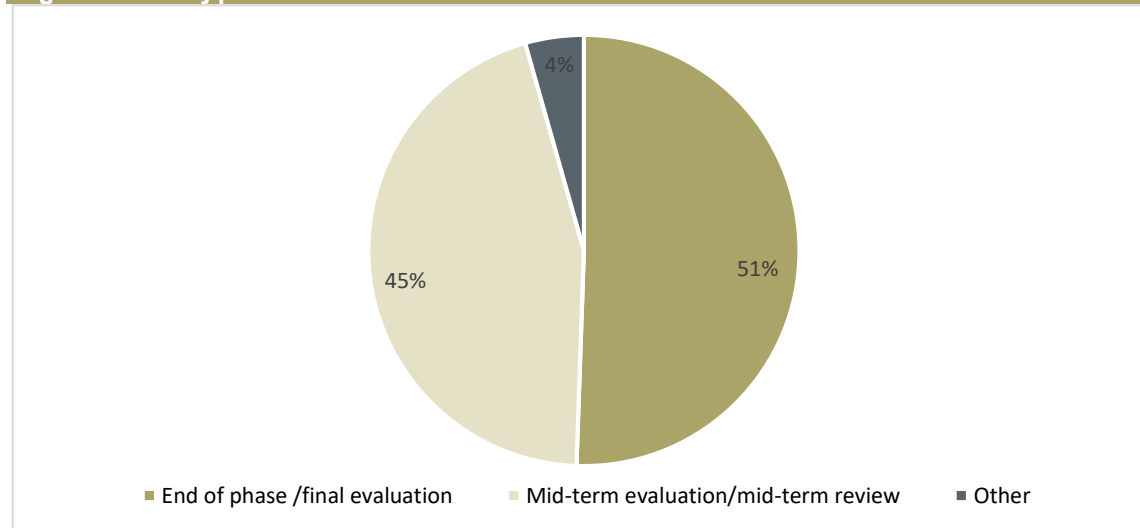
Figure 2.6: Geographic focus of interventions



2.8 TYPE OF EVALUATIONS

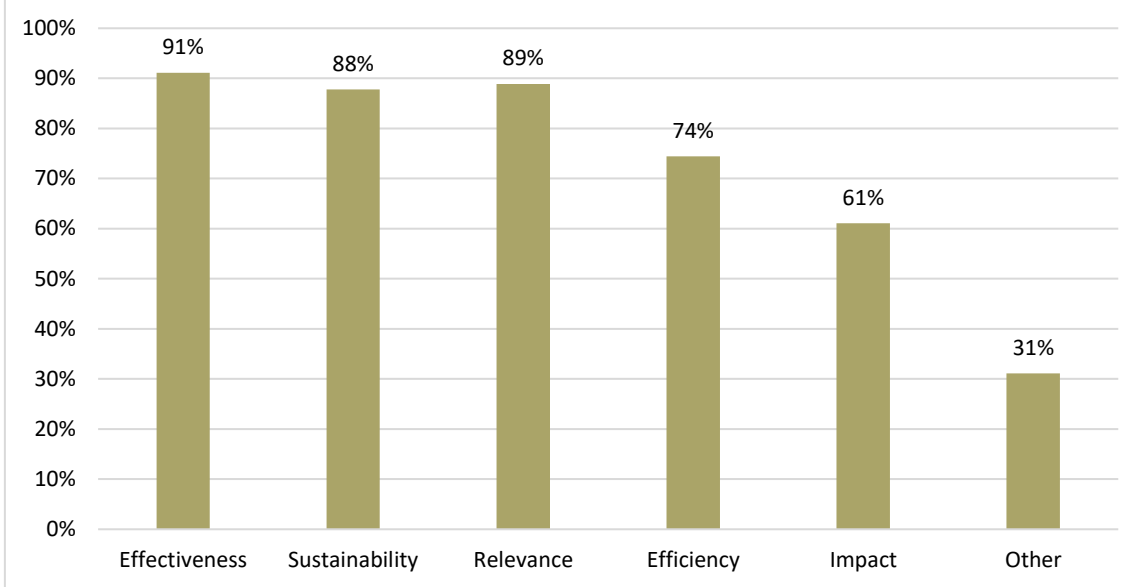
A little over half (51%) of the evaluations in the dataset are end of phase/final evaluations, and 45 percent are mid-term evaluations or reviews.

Figure 2.7: Type of evaluations



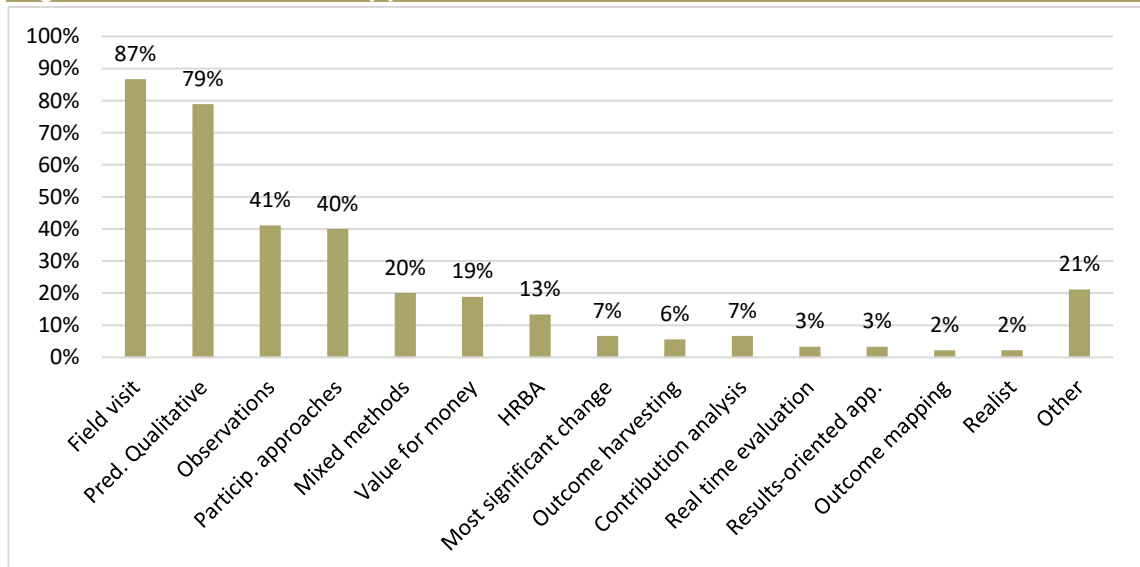
2.9 EVALUATION CRITERIA

Over 90 percent of the partner-led evaluations have applied OECD/DAC's effectiveness criteria. Relevance (89%) and sustainability (88%) were assessed by most evaluations. Efficiency was assessed in 74 percent of the evaluations and impact was assessed in 61 percent of the evaluations. Among the "other" evaluation criteria, were specific evaluation question areas relating to areas such as gender equality and organisational capacity. In most cases, these could have been organised under one of the five evaluation criteria.

Figure 2.8: Evaluation criteria

2.10 EVALUATION APPROACHES

Various approaches were used in the evaluations. The most common was field visits (87%). Other commonly applied approaches include predominantly qualitative approaches (79%), observations (41%), and participatory approaches (40%).

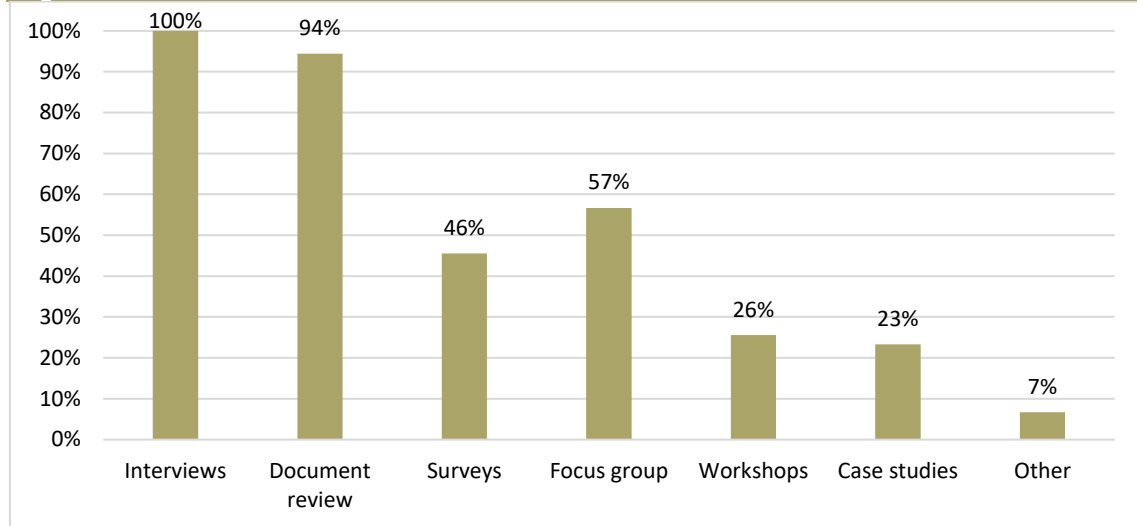
Figure 2.9: Evaluation approaches

2.11 DATA COLLECTION METHODS

All of the evaluations in the dataset relied on interviews as a data collection method. Document review (94%) was also common. Focus groups (57%) and surveys (46%) were

used in roughly half of the evaluations. Workshops (26%) and case studies (23%) were used in around one-fourth of the evaluations.

Figure 2.10: Data collection methods



2.12 EVALUATION BUDGETS AND DURATION

Data for budgets of the evaluations was difficult to obtain since this information was most often missing in the evaluations and terms of references. It was only possible to obtain the financial cost of the evaluation for 21 (23%) of the evaluations.

Table 2.1: Financial budgets of partner-led evaluations (23% of dataset)

| | SEK |
|----------------|-----------|
| Maximum | 1 120 000 |
| Minimum | 80 000 |
| Average | 357 238 |
| Median | 290 000 |

There is adequate coverage of data relating to the calendar time for the evaluations (70 evaluations or 77%). The average length of evaluations in calendar time was 17 weeks and median length was 13 weeks.

Table 2.2: Weeks of calendar time (77% of dataset)

| | Weeks |
|----------------|-------|
| Maximum | 70 |
| Minimum | 2 |
| Average | 17 |
| Median | 13 |

Data regarding the duration of the evaluation was available for less than one-third of the dataset (32%). It was most common that evaluations consisted of around 40 days (eight weeks) of input.

Table 2.3: Weeks of input (32% of dataset)

| | Weeks |
|----------------|--------------|
| Maximum | 28 |
| Minimum | 3 |
| Average | 10 |
| Median | 8 |

2.13 SUMMARY

While the dataset of partner-led evaluations is missing evaluations from one-third of Sweden's development strategies, the available evaluations and reviews from two-thirds of the Swedish development strategies comprise a great variety of interventions. They represent evaluations or reviews of varying financial size, undertaken by many types of partners, and in a range of sectors. The evaluations and reviews have been of modest financial size, and were typically conducted over a period of three-to-four months. Most evaluations have assessed all five OECD/DAC criteria, despite many being mid-term reviews of interventions that have only been implemented for a few years. Qualitative approaches have been used most often. Interviews, document reviews, focus groups discussions, and surveys have been the most common means of collecting data.

3 Quality Assessment

3.1 INTRODUCTION AND METHODOLOGY

The following section discusses the quality of 14 partner-led evaluations sampled from the dataset. The quality assessment aimed to provide an aggregate picture of evaluation quality among partner-led evaluations. The team created an assessment framework that was used for assessing the quality of each evaluation report in the sample. The framework is primarily based on OECD/DAC's Evaluation Quality Standards (included in Annex 2). It takes into consideration the evaluation's language and form; methodology; context; and analysis using 15 different parameters. These parameters include 52 detailed criteria (some of which are further divided into sub-components) which each evaluation was assessed against. A mark between 0 and 5 was awarded in relation to each parameter based on a pre-defined scale. This allowed for systematic assessment.

To minimise bias, one member of the team undertook the quality assessment of all 14 evaluations, which included a score and an explanation justifying each score. This was then quality-assured by another team member. The team also tested the robustness of the framework by analysing the effects of increasing or decreasing the score for a parameter by one full point. It found that the aggregate score is not sensitive to a changed score for a parameter because of their relatively large number.

To further ensure a fair and balanced assessment, the quality assessment also included analysing the terms of reference of each sample, using six criteria, also outlined in Annex 2. In general, ToR set important preconditions for the evaluation's quality, such as the resources available, the calendar time, the methodology to be applied, and the questions to be asked.

Once the utility assessment was undertaken, quality assessments for evaluations that had a utility score that was either very high or low in relation to the quality assessment were revisited to assure that all quality scores were merited and correct.

As discussed in section 1.2, the evaluations and reviews that were sampled were completed before September 30, 2019. The list below provides the full names of the sample evaluations and reviews that were quality assessed as well as the abbreviated names used in this report for easy reference:

1. Afghanistan, SCI (Save the Children International): *External Project Evaluation Report on Sida Child Protection Project in Afghanistan - 2016 to 2019*

2. Bangladesh, MJF (Manusher Jonno Foundation): *Strengthen Civil Society and Public Institutions to Combat Gender-based Violence and Build Community Resilience to Climate Change*
3. Bosnia, UN Women: *Standards for Engagement for Ending Violence against Women and Domestic Violence in Bosnia and Herzegovina*
4. Bosnia UNDP: *External Evaluation Report “Strengthening the Role of Local Communities/Mjesne Zajednice in Bosnia and Herzegovina 2015–2019*
5. FAO FLEGT: *Mid-term Evaluation of the FAO-EU Forest Law Enforcement, Governance and Trade Programme, Phase III*
6. Freemuse: *External Evaluation of Artistic Freedom Programme 2016-2018*
7. SNV Challenge Fund: *Innovations Against Poverty II: Mid-term Review*
8. Integrity Action: *Integrity Action’s 2016-2020 Sida Grant – Mid-Term Evaluation*
9. Liberia ZOA: *Mid-Term Review of the LBR180 Project: Peacebuilding and Reconciliation Through Community Dialogues (PRCD): Strengthening Social Cohesion and Civic Trust in Five Counties*
10. Myanmar, 3MDG Fund, UNOPS: *Final Evaluation of the Three Millennium Development Goal Fund*
11. Palestine Oxfam: *Mid-term Evaluation Report Developing Equitable Agricultural Production and Market Systems for Resilient Economic Development in the occupied Palestinian Territory*
12. Palestine VAFF: *Visual Arts: A Flourishing Field (VAFF) Mid-term Evaluation Report*
13. AMDT Tanzania: *Mid Term Review Agricultural Markets Development Trust*
14. Tanzania, HRD: *Review of the Tanzania Human Rights Defenders Coalition Protection Mechanism*

Annex 4 consists of an overview of the 14 evaluations sampled, and Annex 5 contains a summary of each evaluation. Below is an overall summary table:

| | Strategy | Type of Partner | Sector(s) |
|------------------|---|-------------------|---|
| Afghanistan, SCI | Afghanistan 2014-2019 | International CSO | Health/ Education/ Human rights |
| Bangladesh, MJF | Bangladesh 2014-2020 | National CSO | Gender equality/ Climate change |
| Bosnia, UN Women | Regional Eastern Europe, the Western Balkans and Turkey | UN Agency | Democracy/ Human rights and freedom of expression |
| Bosnia, UNDP | Regional Eastern Europe, the Western Balkans and Turkey | UN Agency | Democracy/ Human rights and freedom of expression |
| FAO- FLEGT | Sustainable economic development 2018-2022 | UN Agency | Agriculture/ Food security |
| Freemuse | Human rights, democracy and the rule of law 2018-2022 | International CSO | Democracy/ Human rights and freedom of expression |
| Integrity Action | Human rights, democracy and the rule of law 2018-2022 | International CSO | Democracy/ Human rights and freedom of expression |

| | | | |
|-----------------------------|---|--|---|
| Liberia, ZOA | Liberia 2016-2020 | International CSO | Conflict resolution/ Peace and security |
| Myanmar, 3MDG Fund UNOPS | Myanmar 2018-2022 | UN Agency | Health |
| Palestine, Oxfam | Palestine 2015-2019 | International CSO | Market development / Agriculture and food security |
| Palestine, VAFF | Palestine 2015-2019 | Regional CSO | Democracy/ Human rights and freedom of expression |
| SNV Challenge Fund | Sustainable economic development 2018-2022 | International CSO | Market development / Sustainable societal development |
| Tanzania, AMDT | Tanzania 2013-2019 | Donor established funding mechanism | Market development / Agriculture and food security |
| Tanzania, HDR | Tanzania 2013-2019 | National CSO | Democracy/ Human rights and freedom of expression |

3.2 OVERALL ASSESSMENT

The table below presents the assessed evaluations and reviews in order of highest average quality score to lowest, out of a score of 5.0. The top four evaluations are end-term evaluation and three of them were procured by UN partners.

| Evaluation | Language and form | Method- ology | Context | Analysis | Average | ToR |
|---|----------------------|------------------|------------|------------|------------|------------|
| 1. Myanmar, 3MDG Fund UNOPS, <i>end-term</i> | 5.0 | 4.5 | 5.0 | 4.5 | 4.8 | 4.6 |
| 2. Freemuse, <i>end-term</i> | 4.5 | 4.2 | 5.0 | 5.0 | 4.7 | 4.0 |
| 3. Bosnia, UN Women, <i>end-term</i> | 5.0 | 5.0 | 4.0 | 4.7 | 4.7 | 4.0 |
| 4. Bosnia, UNDP, <i>end-term</i> | 5.0 | 4.8 | 4.0 | 4.0 | 4.5 | 3.7 |
| 5. SNV Challenge Fund <i>mid-term</i> | 4.0 | 4.4 | 4.0 | 4.8 | 4.3 | 3.2 |
| 6. Tanzania, AMDT <i>mid-term</i> | 4.0 | 4.2 | 4.0 | 4.6 | 4.2 | 4.0 |
| 7. FAO-FLEGT <i>mid-term</i> | 3.0 | 4.2 | 4.0 | 5.0 | 4.1 | 2.7 |
| 8. Tanzania, HDR <i>mid-term</i> | 4.0 | 4.6 | 3.0 | 4.4 | 4.0 | 0.7 |
| 9. Afghanistan, SCI, <i>end-term</i> | 4.5 | 5.0 | 2.0 | 3.8 | 3.8 | 3.0 |
| 10. Integrity Action <i>mid-term</i> | 4.5 | 3.8 | 2.0 | 3.8 | 3.5 | 4.0 |
| 11. Liberia, ZOA <i>mid-term</i> | 3.5 | 3.0 | 3.5 | 3.8 | 3.5 | 4.7 |
| 12. Palestine, Oxfam <i>mid-term</i> | 3.5 | 3.2 | 4.5 | 2.4 | 3.4 | 3.8 |
| 13. Palestine, VAFF <i>mid-term</i> | 2.5 | 2.8 | 3.0 | 3.5 | 3.0 | 3.7 |
| 14. Bangladesh, MJF <i>mid-term</i> | 2.0 | 2.8 | 3.0 | 2.2 | 2.5 | 4.0 |
| AVERAGE SCORE | 3.9 | 4.0 | 3.6 | 4.0 | 3.9 | 3.6 |

The five lowest scoring evaluations are all mid-term evaluations or reviews.

Sections 3.3 to 3.7 that follow discuss the five areas of assessment.

3.3 LANGUAGE AND FORM

In assessing language and form, the structure and clarity of the report, as well as the existence of supporting documentation, was taken into consideration. With two exceptions, the reports were fairly well written. A main shortcoming was the lack annexes with information on interviewees, documents reviewed, evaluation tools used, and/or evaluation matrix. The latter was rarely included. The average score for the 14 reports is 4.0/5.0 and the median is 4.2/5.0.

3.4 METHODOLOGY

The methodology sections of the evaluation and review reports varied from a score of 2.0 to 5.0. The average score was 4.0/5.0, and the median score was 4.3. The assessment of methodology included the following criteria:

- How well the methodology was explained including limitations;
- How the sources of information were described;
- The representativity of the data and sampling strategy;
- The extent to which the report discussed validity and reliability of the data; and
- The transparency of the sources and the extent to which ethical principles were applied.

A majority of the evaluations did not include a discussion of relevant ethical evaluation standards applied. Generally, the UN evaluations were better at this, perhaps since the UNEG standards require this to be included.

In three cases, the evaluation team formulate or reformulated the evaluation questions. This was necessary since one ToR had no questions and the other two had several dozens of questions listed.

3.5 CONTEXT

The average quality score for context was 3.6/5.0, with the median at 4.0. This cluster of parameters received the lowest score. The context (country/regional/global context) and the scale and structure of the evaluation object was often insufficiently described to allow for a basic understanding. A typical shortcoming was the absence of financial information on the object being evaluated.

In assessing the context, the study team also assessed the extent that the intervention's theory of change was described and analysed. With a few exceptions, such a discussion was relatively rare.

3.6 ANALYSIS

Both the average and the median score for the quality of the analysis undertaken in each evaluation was 4.0/5.0. The assessment of the analysis included assessing the clarity and quality of:

- The analysis
- The findings
- The conclusions
- The recommendations
- The lessons learnt
- The extent that the evaluation questions were answered.

In most cases, the analysis presented in the evaluation report was fairly good – findings were usually backed by supporting evidence; the structure had a logical flow; and recommendations were separated from findings and conclusions. Around 70 percent of the evaluations answered all evaluation questions in the ToR, the rest either missed a question or had too many questions in the ToR to answer. Gender equality concerns were analysed and reflected in the findings, conclusions, and recommendations in all but two evaluations. These two evaluations did mention gender equality concerns but did not reflect on them very well.

The study team also found that there were some common shortcomings. First, most evaluations did not separate the evidence from conclusions. Instead conclusions were interspersed in the findings. Second, the way recommendations were presented did not always facilitate easy follow up. They lacked numbering. Some were not formulated in full sentences. They were most often not directed at specific stakeholders, even though it can be assumed that they were addressed to the partner organisations. Third, several evaluations took a mechanical approach to the intervention’s logical framework rather than analysed and discussed the theories of change and underlying assumptions. Fourth, sometimes evaluation teams had insufficient understanding of evaluation terminology – mixing up outcome and impact; effectiveness and efficiency; and effectiveness and impact. A few evaluations meticulously monitored the extent quantitative targets had been met, rather than analyse the extent change at different levels had occurred.

Very few ToR requested lessons learnt in the OECD/DAC sense of the term. Some of the evaluations labelled sections “lessons learnt” or discussed “lessons learnt,” but these are either lessons learnt by the intervention or conclusions drawn by the evaluation. They are not lessons that represent contributions to general knowledge or refine or add to commonly accepted understanding.

3.7 TERMS OF REFERENCE

The score for the terms of references for the 14 evaluations/reviews ranged from a score of 0.7 to 4.7 (out of 5.0). Six criteria were used to assess the ToRs:

1. *How clearly the rationale and purpose of the evaluation was conveyed:* the extent to which the ToR contains information on why the evaluation is being undertaken at this point in time, for whom it is undertaken, and how the evaluation is to be used for learning and/or accountability functions. Around two-thirds of the ToRs scored full points in this area.
2. *How well the specific objectives of the evaluation were presented:* to what extent the ToR clarifies what the evaluation aims to find out e.g., ascertain results (output, outcome, impact) and assess the effectiveness, efficiency, relevance and sustainability of a specific development intervention; provide recommendations; and/or draw lessons for future design and implementation. Around two-thirds of the ToRs scored full points in this area.
3. *The description of the evaluation object:* the extent to which the evaluation object is introduced and clearly described. Just over half of the evaluations scored full points in this area. Some ToRs that were otherwise clear, failed to provide an adequate description.
4. *The extent the scope of the evaluation was realistic:* The extent to which the time period to be covered, geographical area, target groups, evaluation criteria, and the depth and breadth of the analysis expected is clarified. Only three evaluation scored full points in this area. The average score was 3.57.
5. *The quality of the evaluation criteria and questions:* How well the evaluation questions or areas of inquiry apply DAC criteria correctly; if there is a reasonable number of evaluation questions (12 or less); and whether the evaluation questions are answerable. No evaluation scored full points in this area and the average score was 2.9.
6. *The allocated timeframe:* The extent to which days and calendar time are in relation to the scope. One ToR scored full points and the average score was 3.2.

The average and median scores for the ToRs were 3.6 and 3.8 respectively. There was no discernible relationship between type of commissioning partner and quality of the ToR: included among the top three and bottom three scorers are a UN agency, an international CSOs and a national CSO.

Common shortcomings of the ToR included the following:

- There were too many evaluations questions, often over 20 and sometimes over 40.
- Some ToRs clearly stated in the beginning the objectives that specified two or three evaluation criteria to be assessed but, later, still included all evaluation criteria in the evaluation questions section.
- The timeframes for the evaluation were often unrealistic and the number of days allocated did not match the broad scope in the ToRs.

3.8 SUMMARY AND CONCLUSIONS

While there is scope for improvement, the quality of the reports is decent, with an average score of 3.9/5.0. Especially noteworthy is the solid effort of evaluation teams to analyse gender equality concerns and reflect this in the conclusions and recommendations.

One area of weakness is the description of the context of the intervention, including the intervention logic/theory of change. Several evaluations failed to describe the structure and scale of the evaluation object, often missing to provide financial information. Only four evaluations adequately described and analysed the intervention logic or theory of change.

The terms of reference were another weak aspect of the evaluations. The average score for these were 3.6/5.0. The partners need to better consider what questions and evaluation criteria are most important and useful to apply. Having too many questions can mean that nothing gets analysed and answered to any useful depth. Asking evaluation teams to assess impact after two to three years of implementation is also a waste of resources.⁷

Some evaluators reformulated the evaluation questions presented in the ToR. These evaluations have had better scores for both quality and utility. The dialogue process in the inception phase to achieve a set of mutually agreeable questions probably improved the understanding among the parties which translated into a better evaluation process and report.

There appears to be some correlation between type of evaluation and level of quality: all end-term evaluations except for one scored at the top. Three of the top-scoring evaluation were procured by UN agencies. Perhaps the higher scores are a result of UNEG's work to improve the quality of evaluations undertaken by UN agencies.

3.9 REFLECTIONS ON THE ASSESSMENT FRAMEWORK

The OECD/DAC's quality standards provided a very helpful basis on which to develop the framework. It guided the development of the framework's five components and the formulation of the criteria. The team found the assessment framework relatively straightforward to use. Many of the scoring criteria were formulated to minimise the risk of subjective judgement – either something is present, or it is absent, and can thus be scored accordingly. Having 15 parameters with a total of 52 criteria contributed to a relatively demanding assessment process. However, because of the large number of

⁷ However, partners and Sida could consider conducting evaluations that examine longer-term results for interventions that have received repeated funding from Sida over several cycles.

criteria, the aggregate score was not sensitive to a changed score on any single parameter. Thus, the team deems that the framework has the correct amount of detail.

Mid-term reviews and evaluations generally scored in the bottom half, while end-term evaluations were the top scoring evaluations. This is discussed further in chapter 5.

4 Utility Assessment

4.1 INTRODUCTION

An evaluation's *utility* depends on how useful the evaluation has been or will be for its stakeholders for internal learning, accountability and/or management purposes. *Credibility* is the extent to which stakeholders perceives the evaluation process and results to be reliable. Credibility is determined to the extent that: i) the evaluation is perceived as being independent, impartial and/or lacking conflict of interest; and ii) the evaluator is perceived as being familiar with the context and has expertise in evaluation and the subject matter being assessed.

This chapter discusses the utility of the 14 sampled evaluations from the dataset (see Annexes 4 and 5) as experienced by Sida and its partner organisations. The utility assessment consisted of collecting the views of two stakeholders involved in each of the 14 evaluation processes – one from the partner organisation, and one from Sida. Since assessing utility involves determining use by whom and how; and how well the evaluation has been disseminated, communicated, and followed up, the team devised questions relating to:

- Evaluation report dissemination
- Utility of report and process
- Credibility of the evaluation team
- Follow up process

(See Annex 3 for the list of the questions asked).

These questions were posed to one Sida stakeholder and one partner organisation stakeholder. To ensure as much feedback as possible, the evaluation team offered stakeholders different options for how to convey their information:

1. Complete a specially devised short electronic survey with closed questions
2. Complete the survey and add comments as relevant
3. Complete the survey and answer a set of open-ended questions via email
4. Complete the survey and convey information orally to the team
5. Complete the survey and convey information orally to the team and answer a set of open-ended questions via email.

This strategy proved effective as all 28 stakeholders provided the necessary data via one of the options listed above. Most stakeholders provided feedback via options 3 or 4 above.⁸

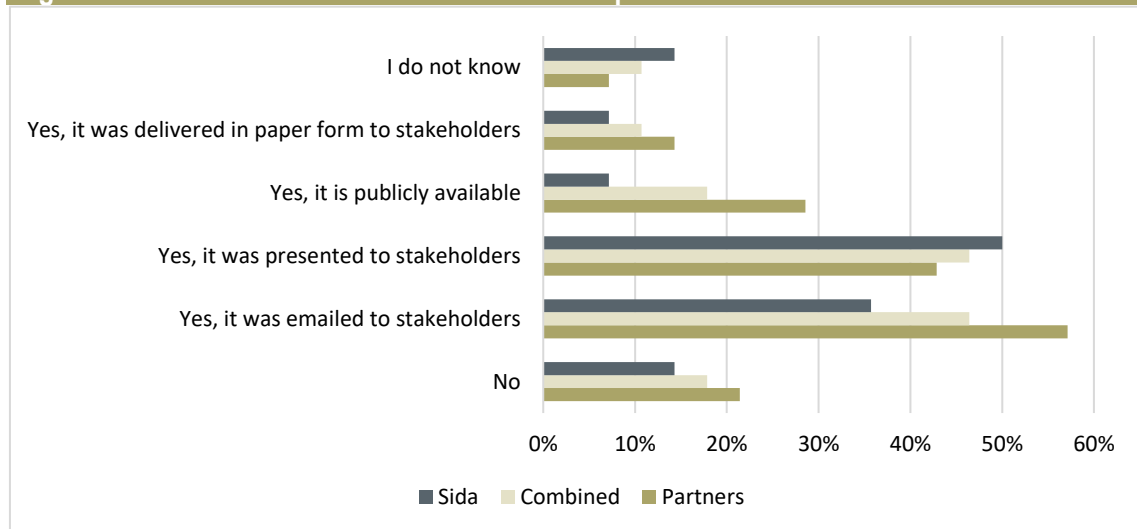
There are a few limitations to the assessment. First, some Sida staff have rotated since the evaluation process began and current officials did not have the knowledge necessary to respond to some of the questions, particularly with regard to the evaluation process. However, rotation did not affect the ability of the partner organisations to answer the team's questions relating to utility. Second, for some of the questions (e.g., relating to dissemination of the evaluation report, level of participation in the evaluation process, use afterwards) Sida staff did not have complete information to be able to answer, but partner organisations were able to provide credible information related to these questions.

The discussion below is divided into sections on dissemination; the utility of the evaluation process and report; the perceived credibility of the evaluation team; and the extent to which the evaluation has been followed up. The final section provides a summary and concluding remarks.

4.2 DISSEMINATION

In most cases, the evaluation reports were shared using more than one channel. Dissemination by email was most common, as seen in the diagram below.

Figure 4.1: Was the evaluation/review report shared with stakeholders?



⁸ As seen in Figure 4.1, there are some differences between the answers from partner representatives and the responsible Sida desk officer. There are different potential reasons for this. First, several of the Sida staff members have moved on and current staff may not have participated in the evaluation process. For questions relating to utility, each party responded from their perspective. For questions regarding dissemination and whether there is prepared management response, partner organisations have more insight than the Sida counterpart.

Another common dissemination approach has been to conduct seminars at which the evaluation report was presented, often by the consultants themselves, to a select group of stakeholders. At least a quarter of the evaluations produced evaluation briefs. In two cases partners mentioned that evaluation briefs were prepared and widely shared. In two other cases, evaluation briefs had been or were being translated from English into national and local languages to enable an even wider audience to access the evaluations results.

It seems that three reports have not been shared. Distribution of one report will take place when an executive summary in Arabic is completed (delayed because of Covid 19). Another report is being prepared for the public and will be placed on the website once finalised. In the third case, no explanation was provided.

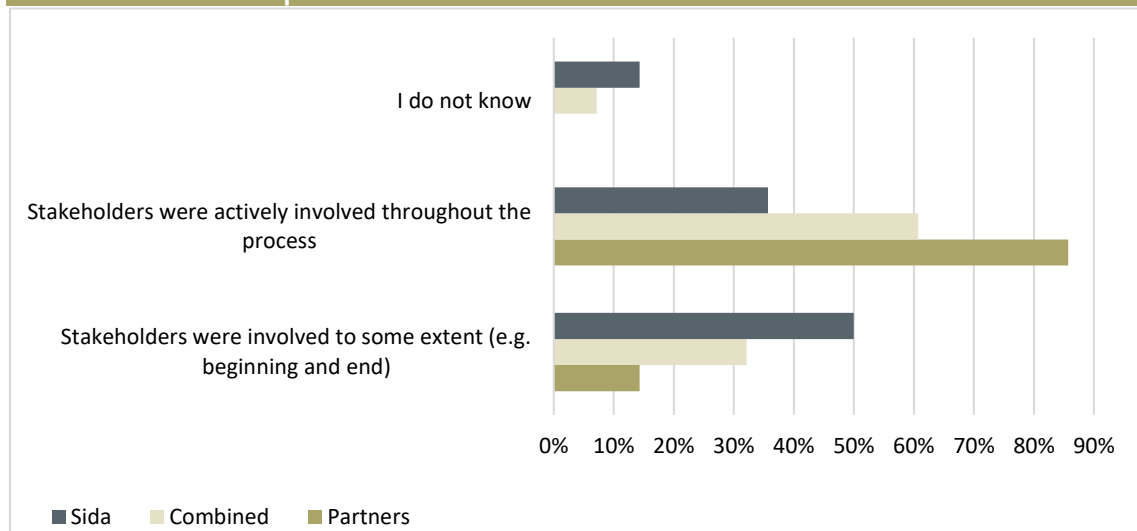
4.3 UTILITY OF THE PROCESS AND REPORT

This section looks at the utility of the evaluation process and the evaluation report. The final sections focus on feedback on what could have made the processes and/or report more useful.

4.3.1 Stakeholder participation in the process

Stakeholder participation in an evaluation or review process can be a good indicator of evaluation utility. Over 80 percent of the partner organisations responded that stakeholders were actively involved throughout the process (Figure 4.2).

Figure 4.2: To what extent were stakeholders involved in the evaluation/review process?



Stakeholders were involved as informants during field work, in interviews, field visits to sites, workshops, and focus group discussions.

The evaluation process was incredibly participatory, with multiple approaches – from online questionnaires, direct interviews, focus groups, etc. – the process also helped better understand the intricacies of the contextual nature and ambitions of the project. – Partner respondent

The discussion with the evaluation team, especially on specific aspects around M&E was a very good process. These sessions almost became training processes. The MTR process and interactions with the evaluation team made it easier to accept recommendations from the MTR. It was a very inclusive process that also included the two consortium partners. – Partner respondent

In two cases, partners felt that participation could have been better:

There could have been more focus on including beneficiaries. There was dialogue and discussion with main stakeholders, but beneficiaries [could have been included] as well. – Partner respondent

Sida staff were more inclined to respond that stakeholders participated in the evaluation “to some extent,” since this typically reflected how they themselves had been involved. On the other hand, partners understood “stakeholder” in a broad sense, meaning actors involved in the interventions – target groups, partner organisations, government counterparts, staff, donors, etc. Two Sida respondents did not know enough about stakeholder participation to be able to respond.

Sida staff typically reported that they had been involved in the preparation phase, especially in relation to developing the terms of reference, as well as in the beginning of the evaluation. Most had also been involved during the final stages – providing comments on reports and attending the evaluation presentation. Sida staff have also generally followed the management responses developed by the partner organisations. However, for over 40 percent of the sample evaluations, Sida staff rotations during 2018 and 2019 hindered the Sida representative’s participation in the full evaluation process:

It was conducted in the turn-over period between the outgoing and incoming project officer at the embassy. It started as the former PO was leaving and was submitted the week after the current PO arrived. This meant that the embassy was not as actively involved as it normally would have [been], and thus the donor perspective and experiences were not reflected in the report. ... When the embassy can provide comments on draft report, the final version is [more useful]. Had the PO been able to be more involved throughout the process, it would have also been a better learning opportunity/experience for the embassy/PO. – Sida respondent

4.3.2 Usefulness of the evaluation process

Combined, 80 percent of the stakeholders surveyed found the evaluation process moderately (28%) or greatly (52%) useful. Partners generally found the process more useful than Sida staff, which is understandable since they were more involved in it. Several partner organisations mentioned that the process of reflection, discussions with the evaluators, getting the perspective of an external party, and having the programme’s stakeholders involved were particularly useful aspects of the evaluation.

A good external eye definitely helped me/my colleagues verify [our achievements] in a very challenging environment that often needs adjustments. – Partner respondent

The process was very helpful as it provided opportunity for programme partners to give input to an independent external party regarding the objectives and practical realities (including challenges) of our work. – Partner respondent

The evaluation process was extremely useful for the many stakeholders involved (over 200) as it reflected on the relevance, project, approach, interventions and achievements. – Partner respondent

The evaluation provided a voice for all stakeholders and everyone was honest in providing feedback. – Partner respondent

It facilitated a discussion around the key assumptions that had been part of the programme from the beginning and challenged these. – Partner respondent

It resulted in internal reflection and created a good opportunity for discussion between the Programme and its Steering Committee and donors which was very helpful. – Partner respondent

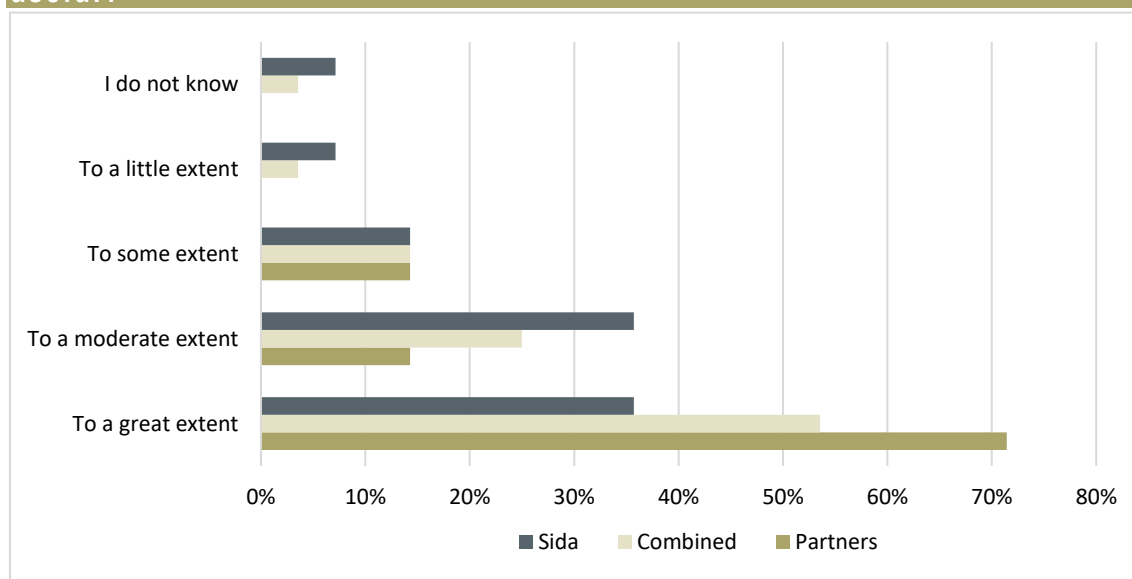
One of the Sida respondents mentioned the exercise of planning for the evaluation as particularly useful:

The most useful in the process was the dialogue on the ToR with the partner. It was a great learning experience, for a small partner to conduct an evaluation in accordance with the criteria of OECD DAC. – Sida respondent

One partner commented on the timeliness of the evaluation (for comparison, see section 4.3.5 below).

The MTR came at a perfect timing, as it was at the end of the first cycle. The MTR fed into the process, before the second cycle started, which allowed the findings and recommendations to be used. Perfectly placed in the project time frame. – Partner respondent

Figure 4.3: To what extent do you think the evaluation/review process was useful?



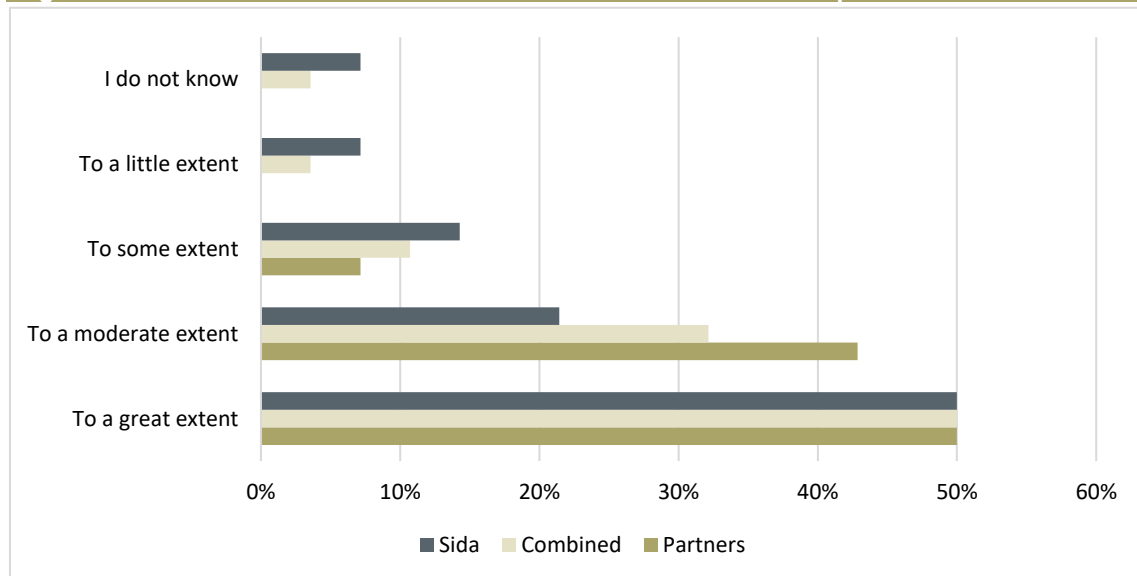
Two partners felt that the evaluation process was useful only to some extent, but none rated the usefulness of the process as a little useful or not useful. One commented on the process as being too pressed for time. The partner held that if more time had been dedicated to the process to allow for more data collection and discussion, it would have taken less time to adopt some of the recommendations. One Sida representative found the process to be of little use. He/she held that the evaluators had insufficient knowledge of the context and were poor at dialogue. One of the two Sida staff members who found the process to be useful to only some extent clarified that while the process was not useful for Sida, it was for the partner:

I know the MTE was useful for the partner, and I assume that stakeholders involved also benefited from the process. – Sida respondent

4.3.3 Utility of the evaluation report

Over 80 percent of the respondents found that the evaluation reports were useful to a moderate (32%) or great extent (50%). The difference between the number of Sida and partner respondents using the report to a great extent was minimal.

Figure 4.4: To what extent has the evaluation/review report been used?



The evaluation reports have been used to for planning, adjusting methods and operations, and re-orienting the programmes:

The report has been used: (i) as a public record for reference by all interested parties; (ii) to stimulate internal reflection and planning; and (iii) to share back with Programme Steering Committee and receive orientation on the recommendations. – Partner respondent

MTR informed [us of] efficiency of operations, how to manage expansion. – Partner respondent

One mid-term review was seen as a benchmark for the upcoming final evaluation of the programme:

It provides a significant and reliable benchmark for the project's final evaluation. – Partner respondent

Over half of the respondents found that the evaluation report helped provide input for a second phase of the programme:

The evaluation's findings provided a road map for the development of the second phase. Learning from the evaluation... has to a great extent developed the implementation of the second phase. – Partner respondent

The Evaluation revealed some main weaknesses in indicators and reporting which was very useful in particular as lessons learned to factor into the set-up of the successor Fund. So in some regard the evaluation has had a great influence of continued programming. – Partner respondent

Likewise, for the Sida respondents, the evaluations and reviews have served to provide important information for the remainder of Sida's support and/or for continued cooperation:

The report is of good quality and gives a number of insights to Sida for the remaining programme period. – Sida respondent

The evaluation itself has proven to be extremely useful in the appraisal of a continued cooperation. The MTE assessed relevance and effectiveness of the partner and its 'fit for purpose' which gave a solid foundation for further assessments. – Sida respondent

It is one of the central documents in the appraisal of a proposal for renewed support. – Sida respondent

Only one Sida stakeholder responded that Sida had used the report to a little extent so far, but expected to use it in the coming year:

The report can be used in preparation for a potential phase III of the project on issues related to scaling up or narrowing down in terms of focus, result areas, number of targeted provinces. However, this process has not yet started, since the project has been extended for another year. Hopefully, the report will be used to a larger extent in the coming year. – Sida respondent

Two Sida staff members reported that the report had been used to some extent. One found the partner's response to the evaluation to be more useful than the report itself:

Most useful was [the partner's] management response. – Sida respondent

The other Sida representative explained that staff turnover due to maternity leave affected how Sida had used the report.

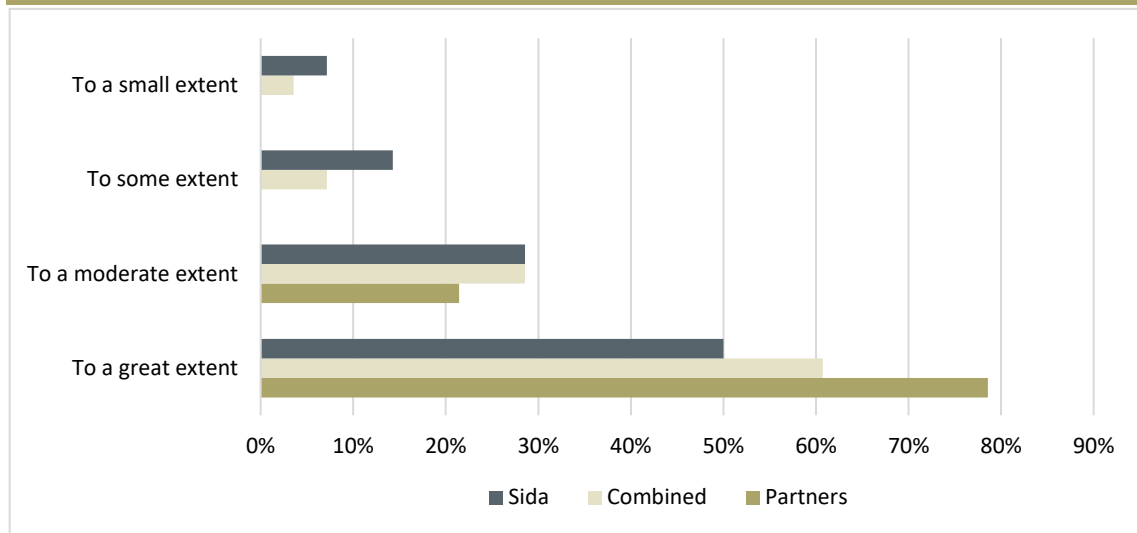
The report could have been better utilised at Sida internally... for knowledge sharing and follow up with the partner on several issues. – Sida respondent

One partner found the evaluation report to be useful to only some extent and developed action plans in response to the report but felt the evaluators' grasp of the organisational context was insufficient.

4.3.4 Meeting stakeholder needs

Ninety percent of the respondents found that the evaluations have met the needs of its intended users to a great (62%) or moderate (28%) extent.

Figure 4.5: To what extent do you feel that the evaluation/review addressed the needs of its intended users?



Feedback from partners and Sida was mostly positive:

It was extensive in its areas of coverage. We do not feel any areas were left out. – Partner respondent

The evaluation pointed out critical points for improvement of the organisation and that the partner organisation do deliver on important results. The evaluation was very valuable both for the partner organisation and for us as a donor. – Sida respondent

One Sida stakeholder who found that the evaluation did not sufficiently meet stakeholder needs, provided the following explanation:

The recommendations issued are very generic and the evaluation did not bring any new angle or insights for future programming. The evaluation does not assess whether the interventions have achieved outcomes or not. However, the partner partly agrees with the recommendations. – Sida respondent

4.3.5 Utility shortcomings

While the majority of respondents (over 80%) found the evaluation utility (process, stakeholder participation, evaluation report and meeting stakeholder needs) to be of satisfactory or high standard, some stakeholders saw room for improvement. They identified several ways that would have made the evaluation or evaluation process more

useful. Four partner organisations had expected more detailed recommendations or specific support with the programme's logical framework:

If possible, it would have been great to have in addition to general a more specific recommendations for the way forward. However, we do understand that this is a challenge for external people unaware of the specificities that usually take longer for a proper familiarization. – Partner respondent

Least useful was the rather generic nature of the recommendations and the team did not fully accept a number of them. – Partner respondent

Not all of the recommendations were specific enough to guide concrete action by the Programme as a result. – Partner respondent

Development of the logical framework was not included as part of the review. However, we have gone through the process after finalising the evaluation. – Partner respondent

A few respondents felt that the evaluation process would have been better if evaluators had access to the right data from the start; encompassed more time for data collection from target groups; and included more time and a better process for triangulation:

It would have been good with more time. The interactions in the field were very limited. This was [however] compensated by the fact that the consultants made use of the time between end of field work and delivery of report to get more information. – Partner respondent

It would have been much better if we could have had a better process on the fact checking and joint discussion on the report together with evaluators and the partner. – Sida respondent

Make sure that the documentation provided is used correctly and that evaluators get access to the right information. – Partner respondent

The timing of the evaluation was raised by four stakeholders. Two partner organisations felt the evaluation was conducted too late, leaving too little time for adjustments to become meaningful before the programme formally ended. Meanwhile, a third partner felt the evaluation was conducted too early:

We should have pushed the evaluation forward into the lifetime of the project to better capture and measure a more substantial impact. – Partner respondent

A fourth partner reflected on timing being both too late and too early at the same time:

The timing of the evaluation was both the most useful and the most challenging. It was good to be able to plan ahead and use the evaluation. But it also proved difficult to see the results, as the programme was not completely finalised. – Partner respondent

4.4 THE EVALUATION TEAM

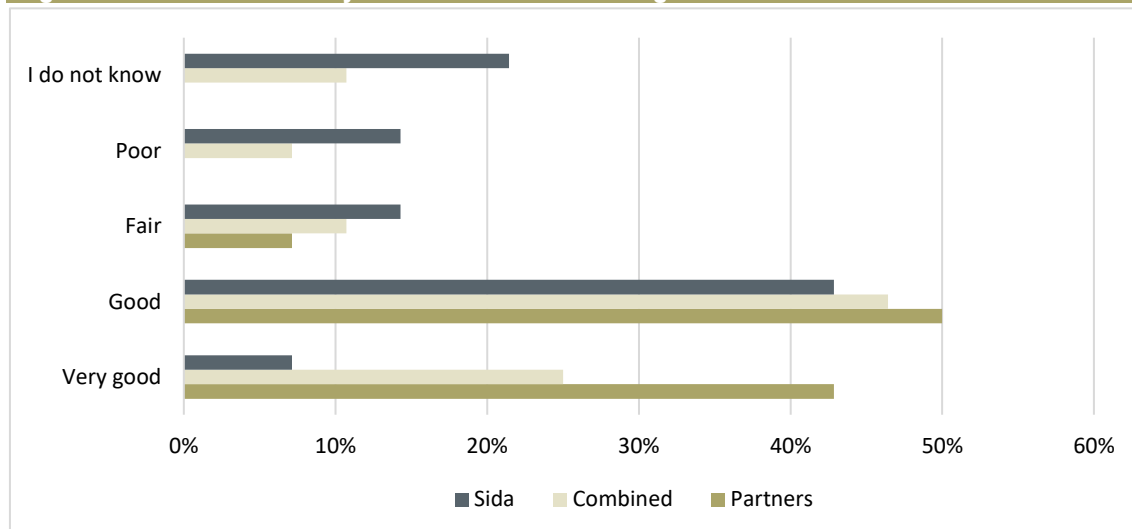
To obtain feedback from the stakeholders on the perceived credibility of the teams that undertook the partner-led evaluations, respondents were asked to comment or provide

feedback on the dialogue with the evaluation team, their view of the team's qualifications, and the extent to which they perceived them as independent and impartial.

4.4.1 Dialogue with the evaluators

Over 90 percent of the partners felt that the dialogue with the evaluation team was good or very good.

Figure 4.6 How would you assess the dialogue with the evaluators?



Since the evaluations were partner-led, some Sida staff had limited dialogue with evaluators.

I cannot recall any contact at all with evaluators, since the evaluation report was prepared for the partner. An email or two at most on more practical issues. – Sida staff member

In some cases, the limited interaction with the evaluation team was due to turnover of Sida staff.

The former program officer left when the Mid-Term Evaluation had just started and the current program officer took up the position the week before the report was submitted. – Sida staff member

In one case Sida staff had to actively reach out to obtain more information about the ongoing evaluation.

Partners, on the other hand, had closer interaction with evaluators throughout the evaluation process. In cases of very close interaction/communication, the evaluators' role was much appreciated as facilitators of processes that were relevant not only for the evaluation, but for the organisation as a whole (see section 4.2).

Communication with the evaluation team was very strong and consistent throughout the process. – Sida partner

In two cases, the need for closer communication only became evident at the reporting stage, when misconceptions and misunderstandings were reflected in draft reports or final reports. Regarding one evaluation both Sida and the partner felt the dialogue was suboptimal:

There was a lot of tension originally, as the evaluators used their main finding to dismiss almost every achievement of the Fund. It took several iterations between Donors and the evaluators for the report to look at things in perspective, balancing the importance of [an issue raised by the team] with the acknowledgment of achievements. – Sida respondent

The evaluators were not easy to dialogue with and did not respond well to comments, questions etc. This view was shared among all four contributing donors who all struggled in communication with the evaluators. – Partner respondent

The partner acknowledged that in part the poor dialogue could have been prevented by the stakeholders:

I think a mistake made by the donors and the Fund was to not invest enough time in the interviews [with the evaluation team], timely follow up around relevant documentation, and relevant stakeholders etc. – Partner respondent

The report of this evaluation is of high quality. The interviews suggest that this could be due to revisions and discussions that led the partner to find the final product to be useful.

4.4.2 Qualifications of evaluators

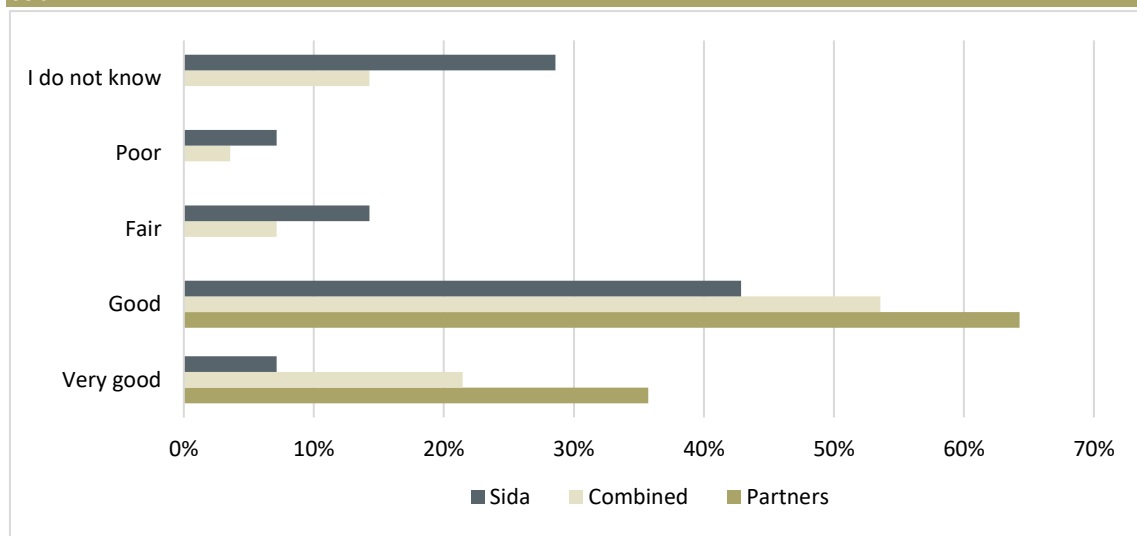
Over three-quarters of the respondents assessed the qualifications of the evaluation team to be good (54%) or very good (21%). The partners generally rated the evaluators' qualifications higher than the Sida respondents.

For one-third of the sample evaluations, it was felt that the evaluation would have benefited from evaluators with more knowledge of the country context, better technical knowledge, or sector knowledge. In two cases the evaluation skills of the evaluators were raised as a problem:

The team of consultants did not crosscheck findings from data collection or triangulated. – Partner respondent

It was noted that the presentation of findings of the evaluation team was not of top quality. It lacked proper structure and coherence. The lead evaluator was stressed and had English language difficulties. – Sida respondent

The evaluation report itself could have been of a better quality. It didn't reflect in detail several findings that enables an external person to capture important issues in detail related to implementation or to Sida's partner processes. – Partner respondent

Figure 4.7: How would you assess the qualifications of the evaluation team?

In one case, when the skills of the evaluator were insufficient, the partner hired a second consultant:

We had two evaluators working on the evaluation. The first one was not strong enough given the complexity of the topic and partners' work, therefore we decided to contract another one to carry on the work and finalise the final report that was eventually stronger and satisfactory.
– Partner respondent

4.4.3 Independence and impartiality

Overall, the evaluators were considered as independent and impartial by both Sida and partners, as shown in Figure 4.8. Partner organisations assessed the evaluators' independence higher compared to Sida staff.

There is nothing in the report that shows any signs of compromised independence. –Partner respondent

Nothing has come to the Embassy's attention which would indicate that the evaluators have not been impartial and independent in their review. – Sida respondent

In terms of ensuring independence, both Sida and partners highlighted that procurement process with clear criteria was important to ensure the selection of an independent evaluator. However, when the pool of qualified evaluators is small, the choice may be limited. Two stakeholders – one from Sida and one from a partner organisation – responded that they deemed the evaluation team to be impartial and independent only to some or to a moderate extent. Both commented that the evaluator was already known to the partner organisation and Sida. They suspected that evaluator produced a more measured report to protect his relationships.

A Sida representative for another evaluation deemed the evaluators to be impartial to some extent:

The dialogue [concerning] the first draft [revealed that] the evaluators' judgement was a bit skewed towards AMDT staff preferences. – Sida respondent

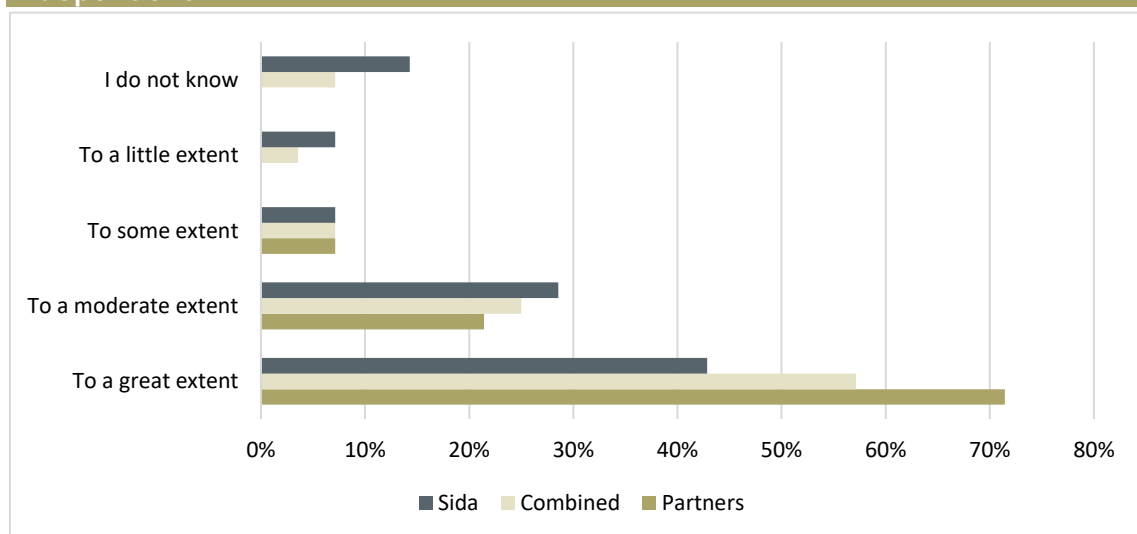
In one evaluation process both Sida and the partner felt that the dialogue was suboptimal. Sida judged the evaluation team to be impartial and independent to a little extent, but the partner, who judged the evaluator to be moderately impartial and independent commented as follows:

The evaluators were independent and impartial. However, once they made their main finding, they seemed to have lost the capacity to judge impartially and what the finding explains or does not explain. – Partner respondent

In two cases, the evaluations/reviews were carried out by or with the partner organisation's own independent evaluation unit, but the stakeholders still found the evaluations to be independent and impartial to a moderate extent.

The reader gets a sense of impartiality and independence. With that said, it was managed by] an internal evaluation unit. – Sida respondent

Figure 4.8: To what extent did you view the evaluator(s) as impartial and independent?



4.5 FOLLOW UP

Among the sample, all partners had prepared a management response for the evaluation. Partners explained that the management response outlined how the recommendations would be followed up, who would be responsible for each component and by when. While some organisations had standardised procedures for evaluation follow-up, other organisations had more flexible procedures, depending on the nature of the recommendations. Some recommendations have been easier to be addressed quickly by organisations, such as those focused on adjusting administrative procedures. Other recommendations have been more complicated to address, such as those focused on more

strategic and programmatic issues. For many of the evaluations, the follow-up process is still ongoing.

The recommendations gave guidance on what should be done. The recommendations were detailed and were presented in a prioritised order by the evaluation team. This was very helpful when following up on the MTR. – Partner respondent

All recommendations from the mid-term evaluation have been discussed with the project team. The team has assessed them and decided on which recommendations are relevant to implement. Based on this, some changes have been made in project methodologies and activities with a focus on sustainability and impact of the project. Furthermore, several learning sessions have been organized to draft lessons learned and use these in current and similar and/or new projects. – Partner respondent

The management team proposed how to deal with the recommendations. Some of the recommendations were easy and very quick to deal with, e.g., operational matters, administrative gaps in procedures that should be added to internal guidelines and policies. Some took a bit longer, and are only now being approved, e.g., deeper understanding of results and changes in programme components. – Partner respondent

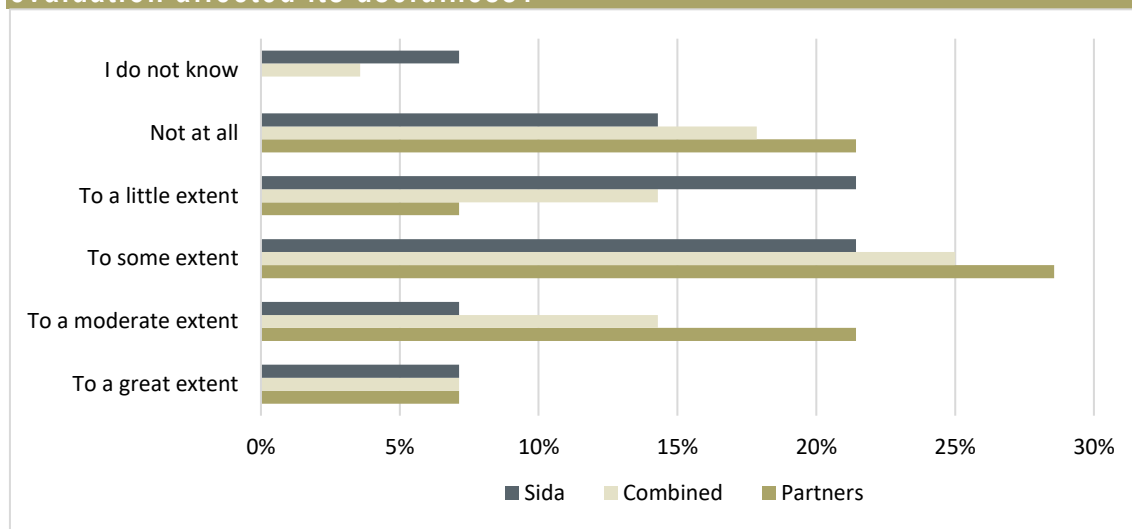
The immediate follow-up was related to training activities and documentation. The more long-term follow-up related to work with line ministries and partnerships. This is still ongoing. – Partner respondent

Follow-up and use of an evaluation not only depend on the evaluation itself, but can also be influenced by the external environment and contextual changes. More than half of the partner organisations report that the context has changed to an extent that it affects the usefulness of the evaluation to some extent. The pandemic was mentioned as a main cause. Partner organisations were more inclined to respond that contextual changes were affecting the usefulness of the evaluation: almost 60 percent expressed that changes in the context were affecting the usefulness of the evaluations to at least some extent, compared to 34 percent of the Sida respondents.

To some extent, especially COVID, which implied implementation of some recommendations have to be looked at again. – Partner organisation respondent

COVID-19 has happened... at this point most project activities are on hold or have been adapted to fit the new context. – Partner respondent

Figure 4.9: To what extent have any contextual changes since the evaluation affected its usefulness?



4.6 SUMMARY AND CONCLUSIONS

The average score for utility of the evaluations is at 4.2/5.0, which is higher than the aggregate quality score (3.9/5.0). The use of the evaluations among partner organisations has been higher than the use by Sida staff. This would seem logical. There may of course be a proclivity among partners to be more positive in their assessment of utility since they themselves had procured the evaluation. However, the fact that management responses have been prepared by the respective partners for the entire sample of 14 evaluations is fairly robust evidence that the evaluations are filling a function and are not being disregarded.

| | Meeting needs out of 4 | Evaluation process out of 4 | Evaluation report out of 4 | Credibility of team out of 4 | Combined score out of 5 ⁹ |
|---------------------------|------------------------|-----------------------------|----------------------------|------------------------------|--------------------------------------|
| Bosnia, UNDP | 4 | 4 | 4 | 3.67 | 4.9 |
| Palestine, VAFF | 4 | 4 | 4 | 3.67 | 4.9 |
| SNV Challenge Fund | 4 | 4 | 4 | 3.33 | 4.8 |
| FAO FLEGT | 4 | 4 | 3.5 | 3.67 | 4.7 |
| Integrity Action | 4 | 3 | 4 | 3.67 | 4.6 |
| Bosnia, UN Women | 3.5 | 3.5 | 4 | 3.33 | 4.5 |
| Tanzania, HRD | 4 | 4 | 3 | 3.50 | 4.5 |

⁹ This score was calculated by taking the average score of the responses to the survey questions relating to: 1) to what extent the evaluation met needs of its users; 2) the extent the process was useful; 3) the extent the evaluation report is/was used; and 4) the credibility of the evaluation team. The latter was calculated taking the average score for a) the quality of the dialogue with the team; b) the perceived qualifications of the team; and c) the team's perceived independence and impartiality.

| | | | | | |
|-------------------------------------|------|------|------|------|-----|
| Liberia, ZOA | 3.5 | 3.5 | 3 | 3.67 | 4.3 |
| Tanzania, AMDT | 3.5 | 3.5 | 3 | 3.00 | 4.1 |
| Freemuse | 3.5 | 3.5 | 3 | 2.67 | 4.0 |
| Bangladesh GBV | 2.5 | 3 | 3 | 3.00 | 3.6 |
| Palestine, Oxfam | 3.0 | 2.5 | 2.5 | 2.83 | 3.4 |
| Afghanistan, SCI | 3.0 | 2.5 | 2 | 2.80 | 3.2 |
| Myanmar, 3MDG Fund UNOPS | 3.0 | 1.5 | 3.5 | 1.83 | 3.1 |
| Average | 3.57 | 3.37 | 3.39 | 3.22 | 4.2 |

4.7 REFLECTIONS ON THE UTILITY ASSESSMENT

The utility assessment is relatively “light-touch” in that only two people from each of the 14 sample evaluations have been consulted – one from Sida and one from the partner organisation. In some cases, the representative from Sida and the partner organisation consulted with colleagues in responding to the questions on utility, but in other cases the representative based their response on their own experience. It would have been useful to also interview each evaluator to have a balanced view. However, this was not included in the ToR.

Giving stakeholders the options to reply to questions in written form or via interview or both allowed the team to obtain more data than if only interviews had been used. However, the most important data was gathered via the electronic survey. It allowed for comparable quantitative data as well as additional qualitative data in the form of comments. The questions developed for the survey proved to be clearly understood. The relatively short number of questions and the ease of replying resulted in a response rate of 100 percent.

5 Potential correlations?

Although it was not included in the terms of reference, the team has attempted to determine the correlations between utility and quality. How can we ensure quality evaluations? Having clear, realistic and well-formulated terms of reference and adequate evaluation budgets seem important. What factors contribute to making an evaluation useful? Intuitively, an evaluation of high quality would be expected to be more useful and credible than one of poor quality. However, high quality and credibility may not guarantee utility.

The team plotted the graphs, included in this chapter, to identify potential correlations between: i) quality of the report and the utility of the evaluation; ii) the quality of the ToR and utility of the evaluation; iii) the quality of the ToR and the quality of the evaluation report; and iv) the budget of the evaluation and the quality of the evaluation. The team also reviewed other variables, such as the type of evaluation (end-term or mid-term). Findings show that the **correlations reviewed are generally weak if existent**. Because many factors contribute to quality and utility, outliers are likely and the team suspects that **the sample is too small to show a clear pattern**. The relational review of variables is discussed below.

5.1 TYPE OF EVALUATION & EVALUATION QUALITY

As noted in Chapter 3, final or end-term evaluations are more represented among the evaluations scoring high on quality. Indeed, the four top scorers are all end-term evaluations, while the bottom five are mid-term evaluations. This could be because more thorough processes were dedicated to these end-term evaluations and they, thus, better comply with the assessment framework criteria. The timing of the evaluation is the strongest indicator of quality among the sampled evaluations.

| Evaluation | Average |
|--|---------|
| Top scorers | |
| 1. Myanmar, 3MDG Fund UNOPS, <i>end-term</i> | 4.8 |
| 2. Freemuse, Global, <i>end-term</i> | 4.7 |
| 3. Bosnia, UN Women, <i>end-term</i> | 4.7 |
| 4. Bosnia, UNDP, <i>end-term</i> | 4.5 |
| Bottom scorer | |
| 10. Integrity Action <i>mid-term</i> | 3.5 |
| 11. Liberia, ZOA <i>mid-term</i> | 3.5 |
| 12. Palestine, Oxfam <i>mid-term</i> | 3.4 |
| 13. Palestine, Qattan Foundation <i>mid-term</i> | 3.0 |
| 14. Bangladesh, MJF <i>mid-term</i> | 2.5 |

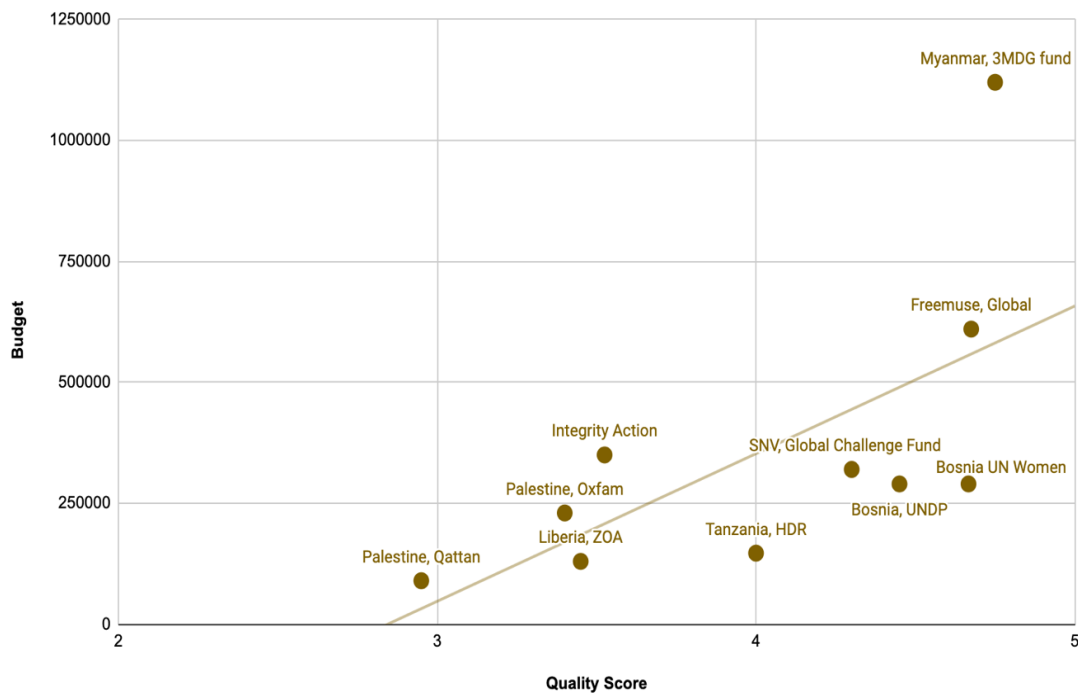
Three of the top scoring evaluations were procured by UN entities. The higher scores could be a result of UNEG's effort to strengthen evaluation quality within the UN system.

5.2 EVALUATION BUDGET & EVALUATION QUALITY

Evaluation budget figures were only available for nine of the 14 evaluations in the sample set. Nevertheless, the pattern below suggests a correlation – the smaller the budget, the lower the quality of the evaluation report, and vice versa.

The four evaluations with the smallest budgets are all mid-term evaluations. The two top scorers are end-term evaluations. The two UN evaluations in Bosnia, both end-term evaluations, have achieved relatively higher quality at lower cost, but it is unclear if the budgets for these evaluations include the daily subsistence allowance (DSA) for the evaluators. UN agencies often draw from other budget lines if the evaluators are engaged as individuals, as opposed to consulting companies, as in the case of these two evaluations.

Figure 5.1: Relationship between quality and evaluation budget of 9 sample evaluations

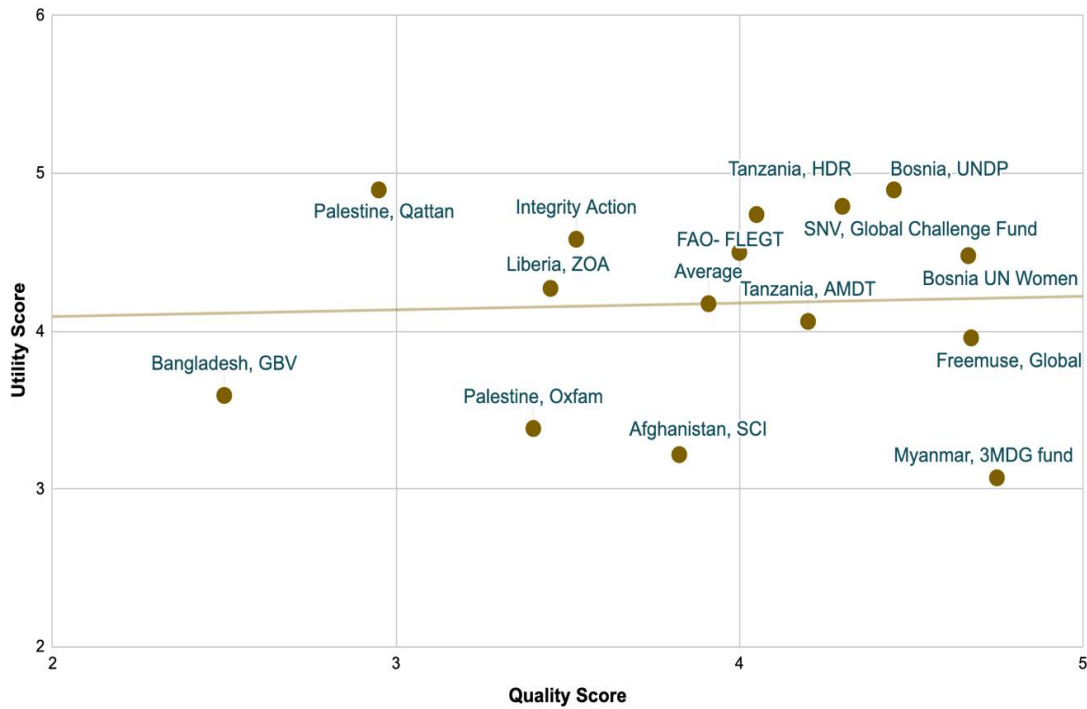


5.3 EVALUATION QUALITY & EVALUATION UTILITY

When plotting the quality of the evaluation report against the perceived utility of the evaluation process and report, no clear correlation is evident. While around half of the evaluations of higher quality have also scored high in usefulness (see upper right hand quadrant), one of the most high scoring evaluations was not seen as particularly useful

(the evaluation process for Myanmar 3MDG Fund was particularly unsatisfactory); and one of the evaluations with high utility did not score high on quality (Palestine Qattan).

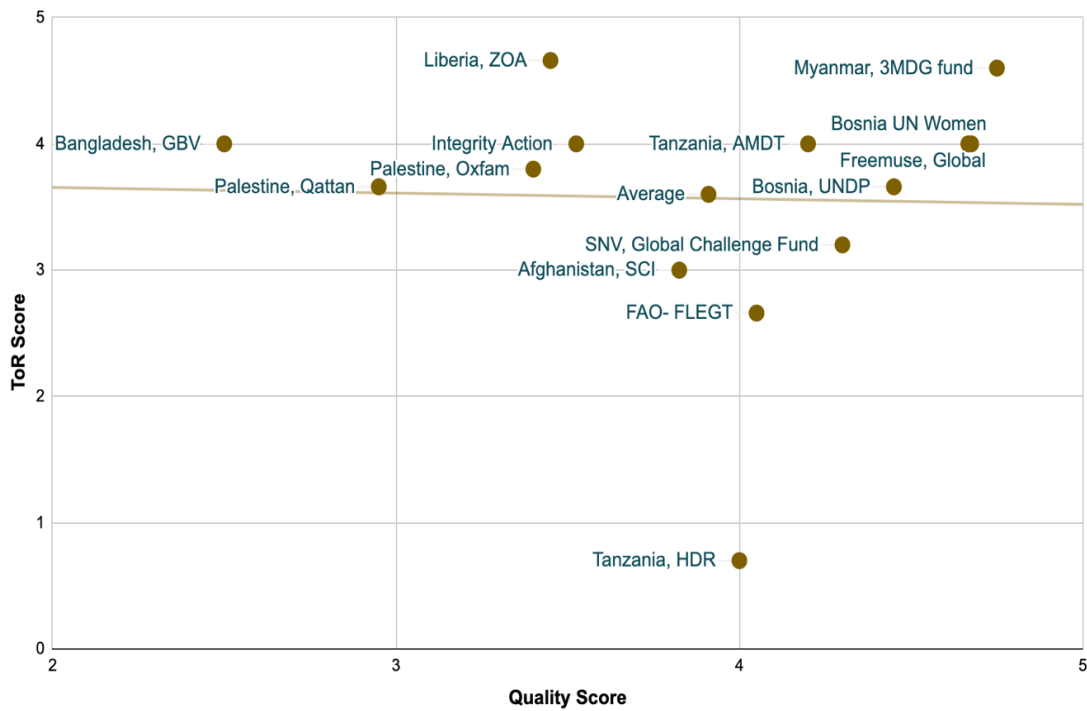
Figure 5.2: Relationship between quality of report and evaluation utility



5.4 QUALITY OF TOR & EVALUATION QUALITY

When plotting the quality of the ToR against the quality of the evaluation report, it is again difficult to see any correlation or pattern. If almost half of the reports are discounted for being outliers (while extreme, this might be reasonable considering some background information for some of these evaluations), there is a cluster in the upper righthand corner that suggests a positive relationship between the quality of the ToR and the quality of the evaluation report.

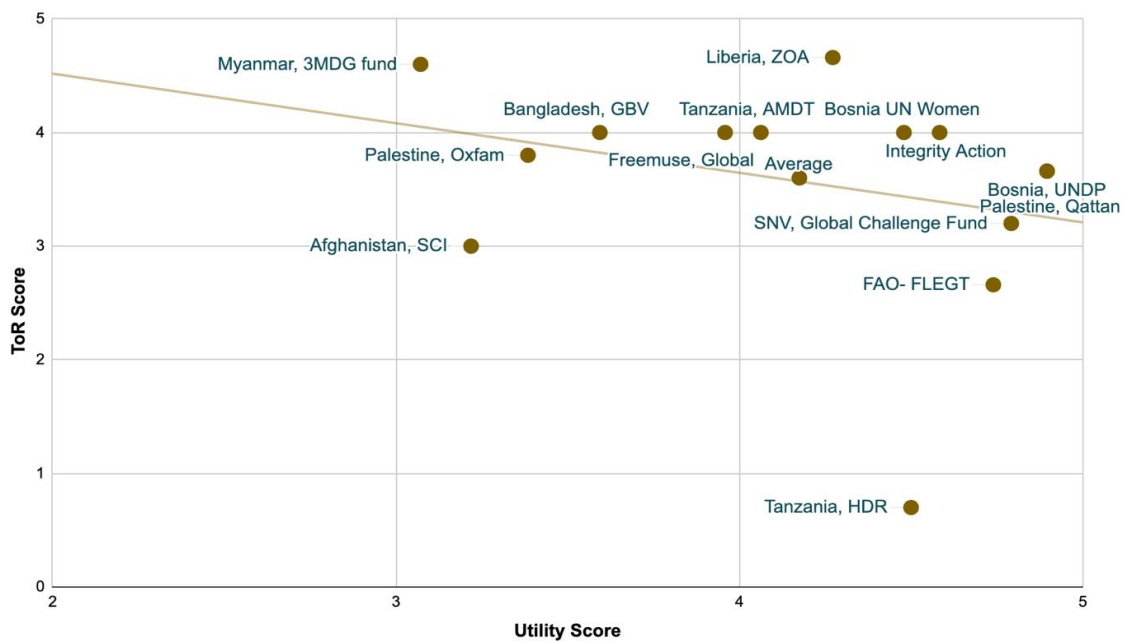
Figure 5.3: Relationship between quality of ToR and report quality



5.5 QUALITY OF TOR & EVALUATION UTILITY

Plotting quality of the ToR against the perceived utility of the evaluation also shows little correlation. If there is one, it seems to lean towards being a negative correlation.

Figure 5.4: Relationship between quality of ToR and evaluation utility



5.6 CONCLUDING REFLECTIONS

The terms of reference for this assignment states:

“One core hypothesis behind conducting evaluations is that IF there is an intended user who has an intended use of an evaluation and the evaluation is based on these identified needs, is conducted independently and with high quality, THEN the evaluation process and product will generate credible findings, conclusions and lessons learnt on what works and not, AND HENCE provide an evidence-base for how to improve operations and be better equipped to achieve the intended objectives.”

The hypothesis implicitly assume that if there is a defined intended use for the evaluation, the ToR would be drafted accordingly and would be of quality.

The data collected by this study and the analysis undertaken of the sample set of evaluations, do not support the hypothesis:

- For all sampled evaluations, the evaluation report was used to improve operations – even when the utility of the process and the quality of the evaluation were not assessed as high;
- The relationship between quality and utility is weak;
- There is no clear relationship between the quality of the ToRs and the quality or utility of the evaluation.

However, the data analysed by this study suggests that quality of a partner-led evaluation depends to some extent on whether it is commissioned as an end-term or mid-term evaluation, and the size of its budget – the latter of which is often dependent on the first.

The size of the sample used in this study is likely to be too small to reveal clear correlations. There is, thus, an opportunity for Sida to further explore these correlations by regularly assessing the quality and utility of partner-led evaluations. Sida could consider applying the tools developed by this study to a larger sample and/or aggregate the results for samples taken over consecutive years. This would provide a larger dataset which may more clearly reveal existing correlations. To ensure a more systematic approach, the Docassemble survey and the utility questionnaire could become part of the contribution management system at Sida. Staff could enter the basic data for each evaluation and then upload the final evaluation report when the process is completed, as well as fill out the utility survey and administer it to the partner after six months have passed.

As part of its support to partners with regard to monitoring, evaluation and learning, Sida could potentially enhance the utility and quality of partner-led evaluations by providing more guidance and/or engaging in dialogue relating to: i) drafting of ToRs; ii) meeting evaluation quality standards; and iii) ensuring an appropriate budget size for evaluations.

Annex 1: Evaluation Metadata Parameters

OVERVIEW EVALUATION METADATA

The data collected from Sida staff via survey in relation to 91 partner-led evaluations was structured along the following parameters:

1. Swedish development cooperation strategy

- | | |
|---|--|
| 1) Afghanistan 2014-2019 | 30) Regional Sexual and Reproductive Health and Rights (SRHR) Strategy in Sub-Saharan Africa |
| 2) Bangladesh 2014-2020 | 31) Regional Sub-Saharan Africa 2016-2021 |
| 3) Bolivia 2016-2020 | 32) Regional Syria Crisis 2016-2020 |
| 4) Burkina Faso 2018-2022 | 33) Capacity development, partnerships and methods that support Agenda 2030 for sustainable development (KAPAME) 2018-2022 |
| 5) Cambodia 2014-2018 | 34) Democracy support through Swedish party affiliated organisations (PAO) |
| 6) Colombia 2016-2020 | 35) Environmental sustainability, sustainable climate and oceans, and sustainable use of natural resources 2018-2022 |
| 7) Cuba 2016-2020 | 36) Gender equality and women's and girls' rights 2018-2022 |
| 8) Democratic Republic of Congo | 37) Humanitarian aid provided through Sida 2017-2020 |
| 9) Ethiopia 2016-2020 | 38) Human rights, democracy and the rule of law 2018-2022 |
| 10) Guatemala 2016-2020 | 39) Research cooperation and research in development cooperation 2015 - 2022 |
| 11) Iraq 2017-2021 | 40) Sustainable economic development 2018-2022 |
| 12) Kenya 2016-2020 | 41) Sustainable peace strategy 2017-2022 |
| 13) Liberia 2016-2020 | 42) Sustainable social development 2018-2022 |
| 14) Mali 2016-2020 | 43) Swedish civil society organisations 2016-2022 |
| 15) Mozambique 2015-2020 | 44) Other |
| 16) Myanmar 2018-2022 | |
| 17) Palestine 2015-2019 | |
| 18) Russia 2020-2024 | |
| 19) Rwanda 2015-2020 | |
| 20) Somalia 2018-2022 | |
| 21) South Sudan 2018-2022 | |
| 22) Sudan 2018-2022 | |
| 23) Tanzania 2013 - 2019 | |
| 24) Uganda 2018-2023 | |
| 25) Zambia 2018-2022 | |
| 26) Zimbabwe 2017-2021 | |
| 27) Regional Asia and Oceania 2016-2021 | |
| 28) Regional Eastern Europe, the Western Balkans and Turkey | |
| 29) Regional Middle East and North Africa (MENA) | |

2. What is your email address?

3. What is the full name of the evaluation?

4. **What year was the evaluation concluded?**
 - a. 2019 → will be asked to upload evaluation
 - b. 2020 → will be asked to upload evaluation
5. **Were Sida/the Embassy and/or you personally engaged in any way in the evaluation process?**
 - a. Yes, Sida/the Embassy
 - b. Yes, me personally
 - c. No
 - d. Don't know
6. **Has Sida/the Embassy located the terms of reference (ToR) for the evaluation? (they should normally be annexed to the final evaluation report or shared at the design stage of the evaluation)**
 - a. Yes
 - b. No
 - c. Don't know
7. **What is the *financial size* of the Sida contribution that has been evaluated? (i.e., what is the agreed amount of the contribution as per the agreement)**
8. **Do you have the email address of the:**
 - a) Sida/embassy person responsible for the partner organisation?
 - b) contact person at the partner organisation?
9. **Do you have any concluding comments?**

DETAILED EVALUATION METADATA

NIRAS entered the following metadata for each of the partner-led evaluation collected:

- | | |
|--|---|
| <ol style="list-style-type: none"> 10. Type of evaluation <ol style="list-style-type: none"> a. Mid-term evaluation/mid-term review b. End of phase /final evaluation c. Ex post evaluation d. Other 11. Evaluation criteria <ol style="list-style-type: none"> a. Relevance b. Effectiveness c. Efficiency d. Sustainability e. Impact f. Other 12. Year of evaluation completion | <ol style="list-style-type: none"> 13. Swedish partner organisation 14. Type of partner <ol style="list-style-type: none"> a. Public sector institution b. Private sector institution c. Community or national level CSO d. Subregional CSO e. Regional CSO f. International CSO g. Academic institution h. Multilateral organisation i. Regional institution j. Public-private partnership and networks k. Other |
|--|---|

15. Type of intervention of intervention evaluated

- a. Project/Programme
- b. Policy/Strategy
- c. Core support
- d. Other

16. Geographical coverage

- a. National
- b. Subnational
- c. Multi-country
- d. Global
- e. Other

17. Sector

- a. Agriculture and food security
- b. Conflict resolution, peace and security
- c. Democracy, human rights and freedom of expression
- d. Education
- e. Environment and climate
- f. Gender equality
- g. Health
- h. Humanitarian aid
- i. Market development
- j. Migration
- k. Sexual and reproductive health and rights
- l. Sustainable societal development

18. Financial size of the evaluation

19. Duration of evaluation (weeks of man hours)

20. Duration of evaluation (weeks in consecutive time)

21. Programme implementation *start date*

22. Programme implementation *end date*

23. Evaluation approaches used

- a. Appreciative inquiry
- b. Beneficiary assessment
- c. Contribution analysis
- d. Developmental evaluation
- e. Field visit

- f. Human rights-based approach
- g. Mixed methods
- h. Most significant change
- i. Observations
- j. Outcome harvesting
- k. Outcome mapping
- l. Participatory approaches
- m. Predominantly qualitative
- n. Qualitative impact assessment protocol
- o. Randomised control trials
- p. Real-time evaluation
- q. Realist evaluation
- r. Social return on investment
- s. Value for money
- t. Other

24. Data collection methods

- a. Document review
- b. Interview
- c. Survey
- d. Workshop
- e. Focus group
- f. Other

Annex 2: Quality Assessment Framework

| I. Language and Form | Average of I.1 to I.2 to 6 is score for Language and Form |
|--|--|
| <p>I.1 Structure and clarity of the evaluation report</p> <p>The structure/form of the evaluation report can readily be understood by the intended audience(s) and is appropriate given the purpose(s) of the evaluation.</p> <ol style="list-style-type: none"> 1. The report is easy to read and understand (i.e., written in an accessible language appropriate for the intended audience) with minimal grammatical, spelling or punctuation errors. 2. The report is structured in a logical way. 3. Graphs and tables are clear, labelled and useful. 4. Key basic information is provided on title page and in opening pages. <ul style="list-style-type: none"> • Name of the evaluation object • Time frame of the evaluation and date of the report • Locations (country, region, etc.) of the evaluation object • Names and/or organizations of evaluators • Name of the organization commissioning the evaluation • Table of contents • List of acronyms | <p>5 = 90%-100% of bullets covered 4 = 80%-90% of bullets covered 3 = 70%-80% of bullets covered 2 = less than 60%-70% of bullets covered 1 = less than 60% of bullets covered</p> |
| <p>I.2 Supporting documentation</p> <p>The following annexes are included:</p> <ul style="list-style-type: none"> • Terms of reference • Methodological tools used | <p>5 = 5 annexes 4 = 4 annexes 3 = 3 annexes</p> |

| | |
|---|--|
| <ul style="list-style-type: none"> • The evaluation matrix • List of persons interviewed, and sites visited • List of documents consulted | <p>2 = annexes</p> <p>1 = 1 annex</p> <p>Zero = No annexes</p> |
| II. Methodology | Average of II.1 to II.5 is score for Methodology |
| II.1 Explanation of methodology used <ul style="list-style-type: none"> • The evaluation methodology is transparently described and explained. • The evaluation matrix includes the evaluation questions, indicators, data sources and methods for data collection. • The report describes the data collection methods/tools and analysis, the rationale for selecting them, and their limitations. • The design, methods and tools are gender-responsive and appropriate for assessing relevant cross-cutting issues. • The report acknowledges any constraints encountered and how these have affected the evaluation, including the independence and impartiality of the evaluation. | <p>5 = 85%-100% of bullets covered</p> <p>4 = 70%-85% of bullets covered</p> <p>3 = 55%-70% of bullets covered</p> <p>2 = less than 40%-55% of bullets covered</p> <p>1 = less than 40% of bullets covered</p> |
| II.2 Sources of information <ul style="list-style-type: none"> • The evaluation report describes the sources of information used (documents, respondents, administrative data, literature, etc.) in sufficient detail so that the adequacy of the information can be assessed. • Stakeholder consultation processes are clearly described, including the rationale for selecting the particular level and activities for consultation. • The report includes discussion of how the mix of data sources was used to obtain a diversity of perspectives, ensure data accuracy and overcome data limits. | <p>5 = 80%-100% of bullets covered</p> <p>4 = 60%-80% of bullets covered</p> <p>3 = 40%-60% of bullets covered</p> <p>2 = 20%-40% of bullets covered</p> <p>1 = 20% or under of bullets covered</p> |
| II.3 Representativeness <ol style="list-style-type: none"> 1. The evaluation report explains the selection of case studies or any samples. 2. The sampling strategy is described. 3. The sampling strategy's rationale is explained. 4. Limitations of the representativeness of the samples are identified. | <p>5= 100% of bullets covered</p> <p>4 = 3 bullets covered</p> <p>3 = 2 bullets covered</p> <p>2 = 1 bullet covered</p> <p>0 = no bullets covered</p> |

| | |
|---|--|
| | Not applicable |
| II.4 Validity and reliability of data The evaluation: <ol style="list-style-type: none"> 1. Makes use of reliable qualitative and quantitative data sources. 2. Triangulates data from different sources. 3. Makes explicit and critically assesses the validity, reliability and limitations (bias, data gaps, etc.) in primary and secondary data sources. | 5 = 80%-100% of bullets covered 4 = 60%-80% of bullets covered 3 = 40%-60% of bullets covered 2 = 20%-40% of bullets covered 1 = 20% or under of bullets covered |
| II.5 Transparency and ethical considerations <ul style="list-style-type: none"> • Complete lists of interviewees and other information sources consulted are included in the report, to the extent that this does not conflict with the privacy and confidentiality of participants. • The evaluation report discusses ethical safeguards and measures that were taken to ensure that the evaluation process conformed with relevant ethical standards, including but not limited to, informed consent of participants, privacy and confidentiality considerations. | 5 = both bullets covered 3 = one bullet covered 0 = no bullet covered |
| III. Context | Average of III.1 to III.2 is score for context |
| III.1 The context of the development intervention The evaluation report describes the context of the evaluation object. <ol style="list-style-type: none"> 1. The contextual information includes descriptions of the: <ul style="list-style-type: none"> • Policy context (development agency and partner policies, objectives and strategies) • Development context – including socio-economic, political and cultural factors, such as gender equality and human rights issues; roles, attitudes and relations 2. The evaluation analyses the context in relation to the development intervention. 3. Scale and structure of the object of the evaluation are clearly described. <ul style="list-style-type: none"> • Number of components • Geographic context and boundaries • Purpose, goal and organisation/management of the object | 5 = 85%-100% of bullet content covered 4 = 70%-85% of bullet content covered 3 = 55%-70% of bullet content covered 2 = less than 40%-55% of bullet content covered 1 = less than 40% of bullet content covered |

| | |
|---|---|
| <ul style="list-style-type: none"> • Total resources from all sources including humans and budgets • Key stakeholders involved in the object implementation, including the implementing agency(ies) and partners, other stakeholders and their roles. | |
| III.2 Intervention logic <ul style="list-style-type: none"> • The evaluation report describes the intervention logic or theory. • The evaluation analyses intervention logic or theory. • The evaluation assesses the intervention logic or theory or provides an alternative ToC. • The evaluation discusses and analyses underlying assumptions. | 5= 100% of bullets covered 4 = 3 bullets covered 3 = 2 bullets covered 2 = 1 bullet covered 0 = no bullets covered Not applicable |
| IV. Analysis | Average of IV.1 to IV.6 is score for Analysis |
| IV.1 Clarity of analysis <ul style="list-style-type: none"> • The evaluation report presents findings, conclusions, recommendations and lessons separately and with a clear logical distinction between them. • The analysis takes into consideration contextual factors. • The evaluation analyses gender equality, cross-cutting issues and/or rights-based issues. • Assumptions underlying the analysis are made explicit. | 5 = 80%-100% of bullets covered 4 = 60%-80% of bullets covered 3 = 40%-60% of bullets covered 2 = 20%-40% of bullets covered 1 = 20% or under of bullets covered |
| IV.2 Findings Findings are objectively reported and substantiated by evidence. | 5 = all findings are substantiated 4 = most findings are substantiated 3 = some findings are substantiated 2 = a few findings are substantiated 0 = most findings are not substantiated |
| IV.3 Conclusions The evaluation report presents conclusions that: | 5 = 80%-100% of bullets covered 4 = 60%-80% of bullets covered |

| | |
|--|---|
| <ul style="list-style-type: none"> • Are well substantiated by the evidence presented. • Logically flow from the analysis. • Reflect reasonable evaluative judgments relating to key evaluation questions. • Provide insights into the identification and/or solutions of important problems or issues pertinent to the prospective decisions and actions of evaluation users. | <p>3 = 40%-60% of bullets covered 2 = 20%-40% of bullets covered 1 = 20% or under of bullets covered</p> |
| <p>IV.4 Recommendations Recommendations are clearly written and:</p> <ul style="list-style-type: none"> • Flow logically from conclusions. • Are relevant to the object and purposes of the evaluation. • Clearly identify the target group for each recommendation. • Are actionable and reflect an understanding of the commissioning organization and potential constraints to follow-up. | <p>5 = 80%-100% of bullets covered 4 = 60%-80% of bullets covered 3 = 40%-60% of bullets covered 2 = 20%-40% of bullets covered 1 = 20% or under of bullets covered</p> |
| <p>IV.5 Lessons learnt</p> <ul style="list-style-type: none"> • When presented, lessons drawn represent contributions to general knowledge. • Lessons learnt may refine or add to commonly accepted understanding but should not be merely a repetition of common knowledge. | <p>5 = lessons learnt are well presented and contribute to learning 3 = lessons learnt are included but not well presented 0 = lessons learnt is mentioned in ToR but not covered - = Not applicable</p> |
| <p>IV.6 Evaluation questions answered</p> <ul style="list-style-type: none"> • The evaluation report answers all the evaluation questions. • Where answers are not possible, explanations are provided. • The original questions, as well as any revisions to these questions, are documented in the report for readers to be able to assess whether the evaluation team has sufficiently addressed the questions. | <p>5 = all questions are answered well or mostly well 4 = most questions are answered 3 = some questions are answered 2 = some questions are answered, not all questions are answered well 1 = questions are not addressed or answered poorly</p> |

| TOR | Average of 1 to 6 is score for ToR |
|--|--|
| <p>1. The rationale and purpose of the evaluation</p> <p>The rationale, purpose and intended use of the evaluation is stated clearly, addressing:</p> <ul style="list-style-type: none"> • why the evaluation is being undertaken at this point in time • why and for whom it is undertaken • how the evaluation is to be used for learning and/or accountability functions. | <p>5 = the rationale and purpose are well presented</p> <p>4 = the rationale and purpose are somewhat well presented</p> <p>3 = the rationale given is “to undertake an evaluation.”</p> <p>0 – no rationale or purpose is given</p> |
| <p>2. Specific objectives of the evaluation</p> <p>The specific objectives of the evaluation clarify what the evaluation aims to find out and are clearly stated in the ToR: For example to:</p> <ul style="list-style-type: none"> • ascertain results (output, outcome, impact) and assess the effectiveness, efficiency, relevance and sustainability of a specific development intervention; and • provide findings, conclusions and recommendations with respect to a specific development intervention, strategy or policy in order to draw lessons for future design and implementation. | <p>5 = the objectives are well presented</p> <p>4 = the objectives are somewhat well presented</p> <p>3 = the objectives are poorly presented</p> <p>0 – not rationale or purpose is given</p> |
| <p>3. The evaluation object</p> <p>The evaluation object is introduced and clearly described.</p> | <p>5 = the evaluand is sufficiently well described to provide an overview</p> <p>4 = the evaluand is somewhat well described</p> <p>3 = the evaluand is poorly described</p> <p>0 – the evaluand is not described</p> |
| <p>4. The scope of the evaluation</p> <p>The time period to be covered, geographical area, target groups, evaluation criteria, the depth and breadth of the analysis expected.</p> | <p>5 = the scope is sufficiently well described to provide an overview</p> |

| | |
|---|--|
| <p>The scope is realistic.</p> | <p>4 = the scope is somewhat well described 3 = the scope is poorly described 0 = the scope is not described</p> |
| <p>5. Evaluation criteria and questions</p> <ul style="list-style-type: none"> • The evaluation questions / or areas of inquiry apply DAC criteria correctly. • There is a reasonable number of evaluation questions (12 or less). • The evaluation questions are answerable. | <p>5 = 80%-100% of bullets covered 4 = 60%-80% of bullets covered 3 = 40%-60% of bullets covered 2 = 20%-40% of bullets covered 1 = 20% or under 0 = there are no evaluation questions</p> |
| <p>6. Time frame</p> <p>The days and calendar time are in relation to the scope.</p> | <p>5 = the time frame and/or input is realistic 4 = the time frame and/or input appear mostly realistic 3 = the time frame and/or input appear somewhat realistic 2 = the time frame and/or input are not realistic 0 = no time frame is given</p> |

Annex 3: Utility Questions

The following questions were used to gather data about the utility of the 14 sample partner-led evaluations. The first four set of questions were closed questions (multiple choice) with the option of adding comments.

Dissemination

1. Was the evaluation/review report shared with stakeholders? How?

Utility of report and process

1. Was the evaluation/review report shared with stakeholders?
2. To what extent were stakeholders involved in the evaluation/review process?
3. To what extent do you feel that the evaluation/review addressed the needs of its intended users?
4. To what extent do you think the evaluation/review process was useful?

The evaluation team

1. How would you assess the dialogue with the evaluators?
2. How would you assess the qualifications of the evaluation team?
3. To what extent did you view the evaluator(s) as impartial and independent?

Follow up

1. To what extent has the evaluation/review report been used?
2. Was a management response or action plan prepared based on the recommendations in the evaluation/review?
3. To what extent have any contextual changes since the evaluation affected its usefulness?

Open-ended questions

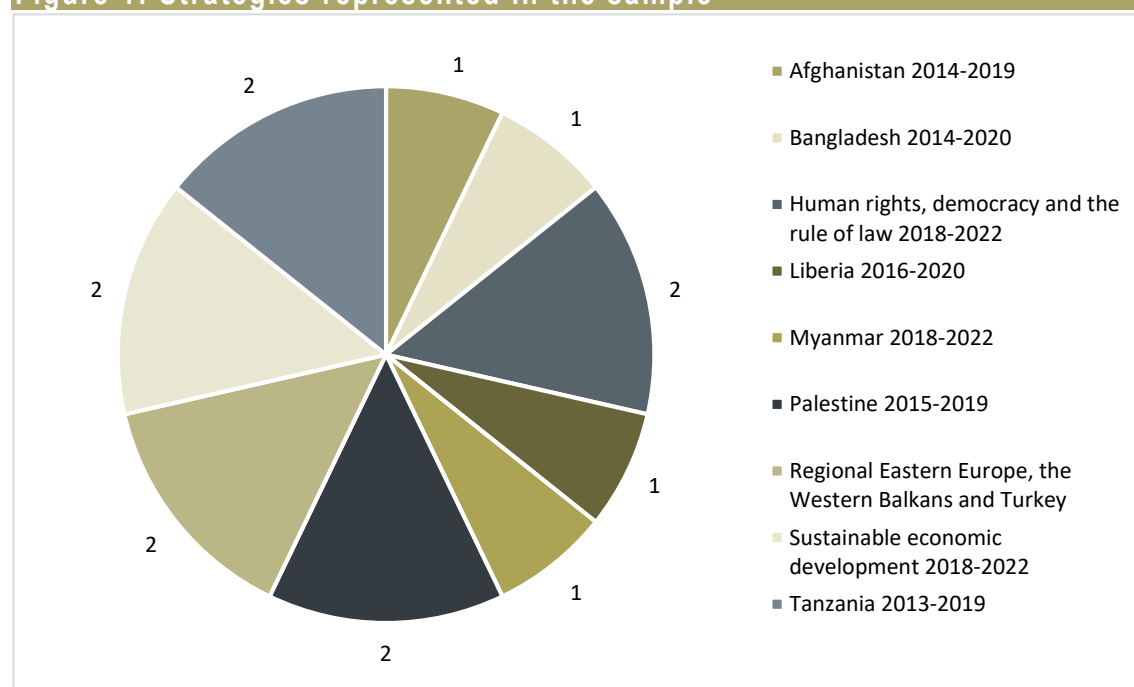
1. What was most useful about the evaluation and/or evaluation process?
2. What was least useful about the evaluation and/or evaluation process?
3. What would have made the process and/or report more useful?
4. How (if at all) has the evaluation been used?
5. Do you know the approximate cost of the evaluation?

Annex 4: Overview of Sample Evaluations

The following annex provides an overview of the sample used for the quality and utility assessments.

Eight Swedish strategies are represented by the 14 sample evaluations, amounting to 30 percent of the dataset and 19 percent of the full population of strategies.

Figure 1: Strategies represented in the sample



In line with the sampling strategy, around 60 percent (57%) of the sample evaluations represent country strategies and around 40 percent (43%) represent global or regional strategies.

Figure 2: Type of strategies represented in the sample

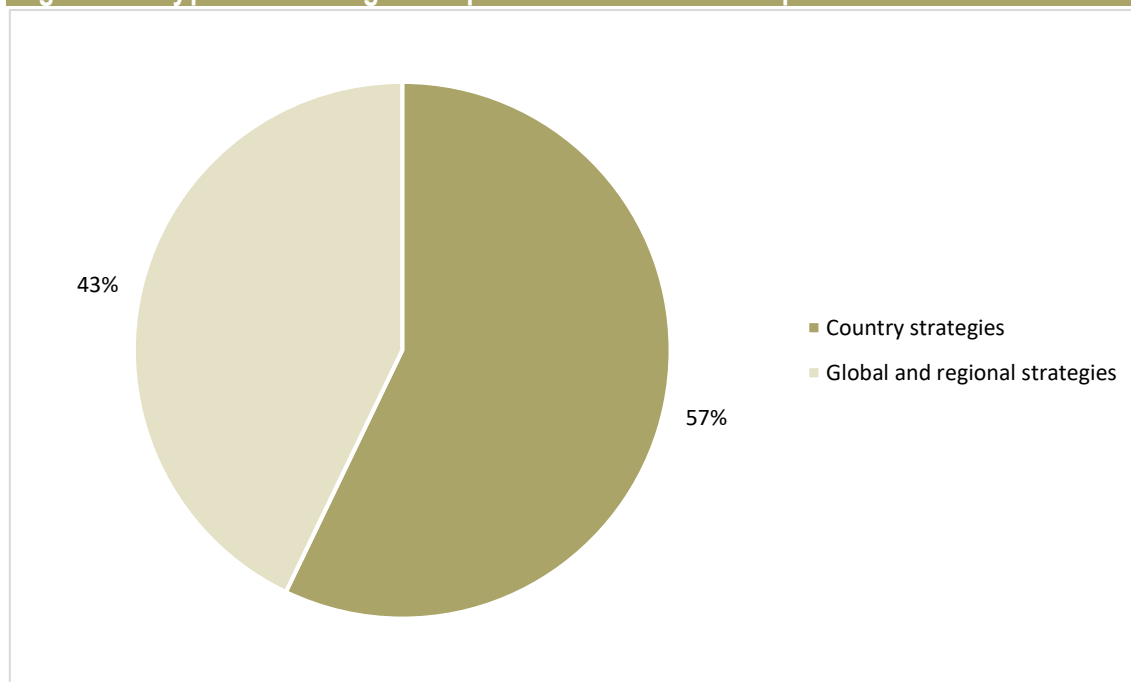
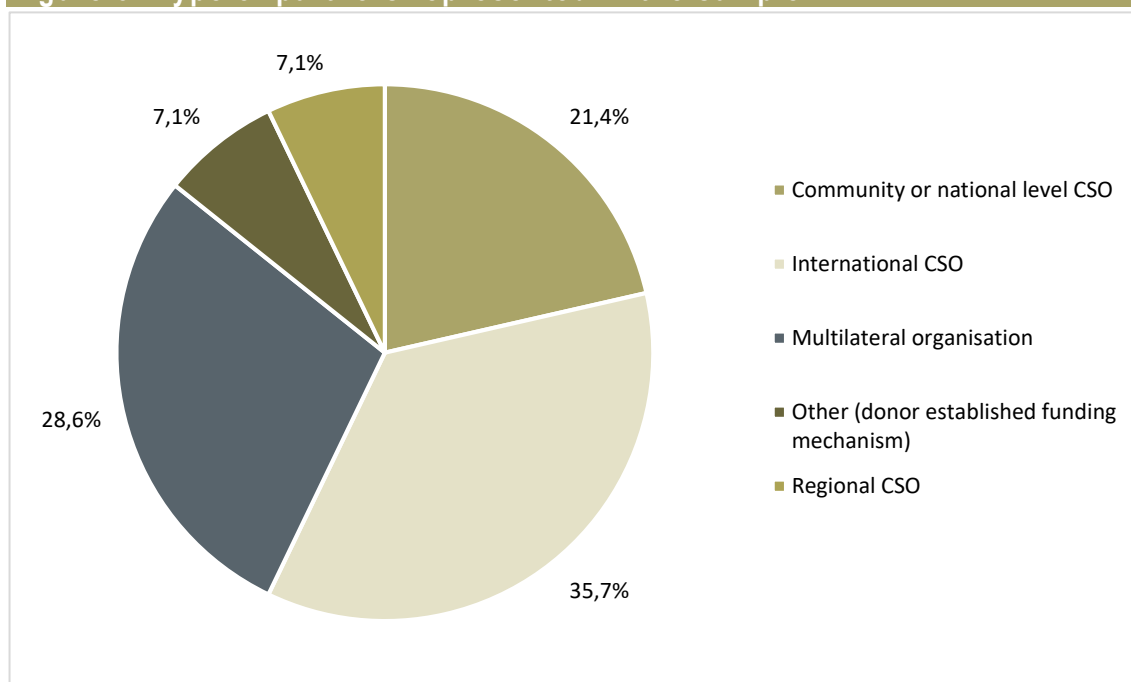
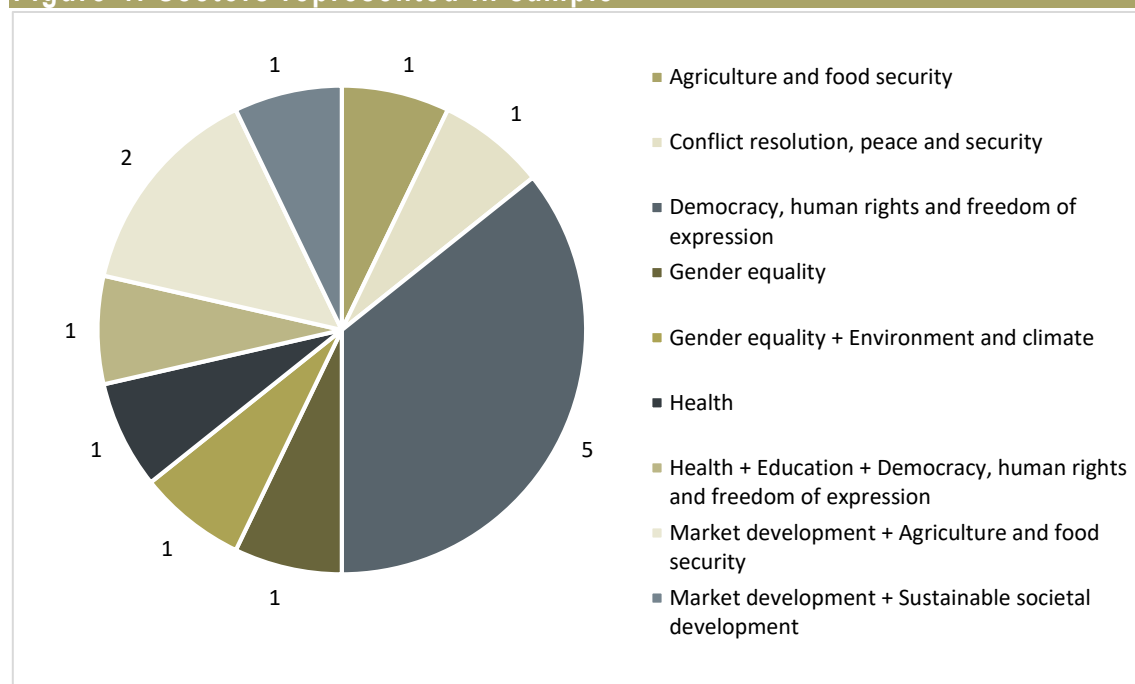


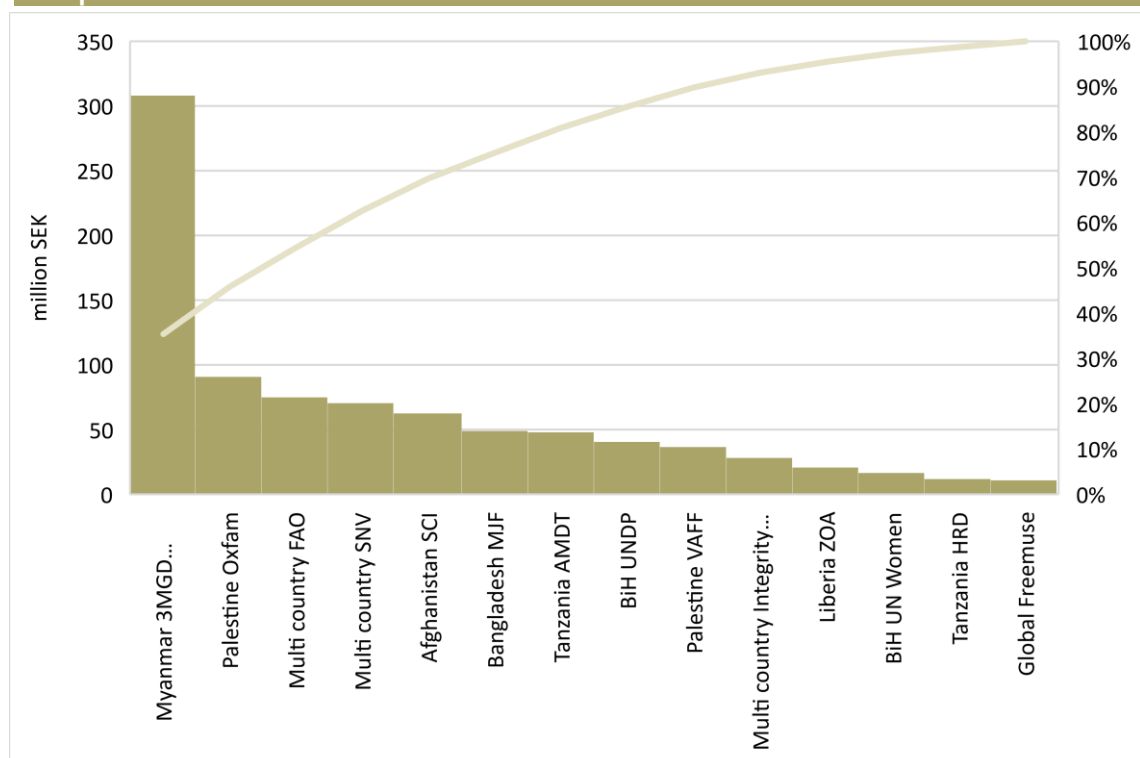
Figure 3: Type of partners represented in the sample



Six of the evaluations in the sample were commissioned by an international CSO, four by a multilateral organisation, and three by a community or national level CSO. This reflects the overall dataset well, in which international CSO is the most common type of partner and these three categories together comprise a large share of the total number of evaluations. In all, 6 of 10 types of partners are represented.

Figure 4: Sectors represented in sample

A large share of the sample evaluations (40%) are within the sector of “Democracy, Human Rights and Freedom of Expression.” Nine of the evaluations/reviews were undertaken at mid-term, five were final or end-term evaluations.

Figure 5: Financial and cumulative size of contribution assessed by sample evaluations

The financial size of the contributions that were evaluated ranges from SEK 12 million to SEK 91 million, apart from one outlier, which is largest contribution – Three Millennium Development Goal Fund (SEK 308 million). The median contribution size is SEK 40 to SEK 48 million and the average is SEK 62 million.

Annex 5: Terms of Reference



Terms of Reference for the Mapping of Partnerled Evaluations during 2018-2019, and the Quality and Use of a selected selected number of these evaluations.

Date: 12 February 2020

1. Introduction

This assignment builds on a need for Sida to deepen its understanding with regard to the number, quality of and usefulness of **Partnerled evaluations financed by Sida**.

When a partner to Sida have been assessed as having satisfactory Monitoring and Evaluations (M&E) capacity and systems, it is the policy to support these partners to themselves commission and manage evaluations pertaining to the Sida funded intervention. Sida thereby encourages such partners to include also evaluation activities in the project/programme budget to be funded and many partners do commission and manage a number of evaluations which are financed by Sida.

Sida however does not have an overall picture of how many evaluations that are Sida financed but commissioned and managed by our cooperation partners. Nor does Sida have an overall picture of the quality and use of these evaluations. This means that currently, there is limited knowledge on the extent to which Sida's development cooperation support is subject to evaluations financed by Sida as this is handled at the level of individual interventions but not systemised. There is however information on decentralised evaluations commissioned and managed by Sida and Embassy Staff available.

In 2018, an Internal Audit of the evaluation function at Sida¹⁰ recommended that a selection of both Sida/Embassy and partnerledd decentralised evaluations should be subject to quality assessments. So far, Sida has been doing quality assessments only of the decentralised evaluations managed by Sida/Embassies, but agrees with the internal audit and will conduct such assessments also of a selection of the partnerled evaluations. The internal audit also recommended Sida to monitor how partner led evaluations are used after the evaluation has been concluded, if a management response/action plan based on the recommendations is elaborated and how the evaluation is used to improve operations. Again, Sida has a Management Respons mechanism

¹⁰ Granskningsrapport IR 18/03, Sida.

included in the contribution management system but as there is no overall picture of the partner-led evaluations it is hard to assess how effective it is.

This assignment is not an evaluation but falls under other evaluation related services as the per framework agreement.

2. Rationale

Given the shortcomings of the current system and the lack of an overall picture with regard to partnerled evaluations, Sida has a need to enhance its understanding with regard to the number, quality and use of partnerled evaluations financed by Sida. The information is to feed back into Sida's (and Evaluation Units) work to strengthen credibility, quality and use of evaluations in general and to get a better idea of how much of Sida's development cooperation support is subject to being evaluated (or not).

3. Assignment object: Partner led evaluations financed by Sida.

The evaluation object is partnerled evaluations financed by Sida and concluded during 2018 and 2019, to *map* them and to do a selected diagnostics with regard to their *quality* and *use*.

The partnerled evaluations will no doubt be very different in terms of purpose, scope, size and focus. And they have been conducted in very different contexts, encompass a wide variety of aid modalities and cooperation partners, where the demand and supply side of evaluation varies extensively. They are bound to vary in terms of quality and use.

One core hypothesis behind conducting evaluations is that IF there is an intended user who has an intended use of an evaluation and the evaluation is based on these identified needs, is conducted independently and with high quality, THEN the evaluation process and product will generate credible findings, conclusions and lessons learnt on what works and not, AND HENCE provide an evidence-base for how to improve operations and be better equipped to achieve the intended objectives.

For the purpose of this assignment and what evaluations that are to be mapped, *Partnerled evaluations financed by Sida* should be defined as *evaluations commissioned and managed by a Sida cooperation partner, financed under a contribution agreement between Sida/Embassy and the partner, directly or indirectly*.¹¹

The term quality and the diagnostic work related to evaluation quality should be based on and conform to OECD/DAC's Quality Standards for Development Evaluation.¹²

¹¹ Even when the budget for an evaluation is not directly detected in a specific project/programme budget, evaluations can be indirectly financed under a contribution agreement of core support to the partner. In the case of these types of agreements, i.e. core support agreements or agreements where the partner is redistributing funds to other partners, only evaluations relating to the agreement between Sida/Embassy and the partner with which Sida has signed the agreement should be considered in this assignment.

¹² DAC Quality Standards for Development Evaluation, OECD 2010.

The term use should be related to *use by whom*, where the partners will be the primary and Sida the secondary users. Are there any signs that the partners and/or Sida have used the selected evaluations? The term should also be related to *what type of use*. Have the evaluations been communicated and disseminated to relevant stakeholders, are they easily accessible, have the evaluations been followed up and/or actions been taken to improve operations (i.e. have Management Responses and/or action plans been elaborated, has implementation taken place).

4. Scope

The scope of the exercise is to conduct a mapping and list all partnerled evaluations financed by Sida that have been conducted and/or concluded during 2018 and 2019. Evaluations which have been completed in 2018, but where the evaluation process was started in 2017, should be included. Evaluations which have been started in 2019, but where the evaluation process is not concluded or the evaluation report completed only in 2020, should be included *but listed separately*.

From the total number of mapped evaluations a sample should be selected (approximately 10-15 evaluations depending on the total number of evaluations) from the list of completed evaluations and the evaluation ToR and report be subject to a quality assessment. In addition, the use of a similar sample size of evaluations should be documented and analysed.

5. Assignment purpose: Intended use and intended users

The purpose and intended use of the evaluation is for Sida/UTV to better understand how many partnerled evaluations financed by Sida that have been conducted during 2018 and 2019, and the quality and use of a selection of these. The purpose is for Sida/UTV, as the primary intended user of this mapping, to get an overview of the total number and scope of decentralised partnerled evaluations conducted within projects/programmes which are funded by Sida. And to get a better understanding of the quality and use of selected evaluation reports. This understanding will serve as an input to the development of Sida's systems for evaluation, evaluation quality assessment systems and evaluation use (such as for example the Management Response system).

Thus, the assignment is to be designed, conducted and reported to meet the needs of Sida /UTV.

6. Questions for the assignment

The questions for the assignment to answer are as follows:

1. **Mapping - How many partnerled evaluations financed by Sida were conducted and/or concluded during 2018 and 2019?** Data should tentatively be disaggregated by title of evaluation, country/region, sector/thematic area, type of partner, size, end-of-programme, mid term or final evaluation, what evaluation criteria that have been assessed, approach and methods. In addition and for the selected sample of mapped evaluations subject to quality assessment, abstracts should be included.
2. **Quality assessment - Based on a selected sample of the total number of mapped evaluations, what can be said in terms of the quality of the evaluation ToR's and the evaluation reports' rigour, quality and credibility of the assessed evaluations?** The assessments should be based on the OECD/DAC quality standards. Sida has checklists for the quality assessment for both the ToR and the evaluation reports, which the consultants are expected to develop further.
3. **Use – use by whom and what type of use. Based on a selected sample of the total number of mapped evaluations, have the evaluations been used (and by whom) and**

if yes, how have they been used? Have the evaluations been communicated and disseminated to relevant stakeholders, are they easily accessible, have the evaluations been followed up and/or actions been taken to improve operations (i.e. have Management Responses and/or action plans been elaborated, and has implementation taken place).

If relevant, the tenderer is welcome to develop further the questions for the assignment in the tender and in the inception note.

7. Approach and methods

It is expected that the evaluator describes and justifies an appropriate approach/method to answer the questions of the assignment. The design and methods for data collection and analysis are expected to be developed in a short inception note for the assignment.

The mapping should aim to capture *all* partnerledd evaluations concluded during 2018 and 2019. Sida foresees that a survey will be used as one tool to inform such a mapping exercise, together with follow-up activities as needed. Timely, informative and at times repetitive communication on the assignment will be vital to engage staff and partner organisations to ensure success in the mapping exercise.

A proposed sample for the quality assessment (approximately 10-15 evaluations depending on the total number of evaluations) and an assessment template for the ToR's and the reports should be approved by Sida prior to the start of the assessment work.

A proposed sample for the assessment of the use of the evaluations (approximately 10-15 evaluations depending on the total number of evaluations) and a detailed method on how to capture and document use should be approved by Sida prior to the start of the assessment work.

In this assignment, Sida does not foresee any field visits.

Limitations to the methodology and methods shall be made explicit and the consequences of these limitations discussed. Where relevant, a *gender responsive* methodology, methods and tools and data analysis techniques should be used.

In cases where sensitive or confidential issues are addressed, evaluators should ensure an evaluation design that do not put informants and stakeholders at risk during the data collection phase or the dissemination phase.

8. Organisation and management of the assignment

This assignment is commissioned by the evaluation unit at Sida HQ. The primary intended user is the evaluation unit at Sida. The assignment will serve as an input to Sida/the evaluation unit's efforts to strengthen the rigour, quality, credibility and use of evaluations funded by Sida.

A small steering group will be set up at the evaluation unit. The steering group is a decision-making body and will evaluate the tenders, participate in the start-up meeting, and approve the short inception note as well as the final mapping and final report.

9. Quality in conducting the assignment

Where relevant, this assignments shall conform to OECD/DAC's Quality Standards for Development Evaluation¹³. The evaluators shall specify how quality assurance will be handled by them during the assignment process.

10. Time schedule and deliverables

It is expected that a time and work plan is presented in the tender and further detailed in the short inception note. The assignment shall be carried out between 20 March to 30 June 2020.

The table below lists key deliverables for the evaluation process. Deadlines for inception note and final report must be kept in the tender, but alternative deadlines for other deliverables may be suggested by the consultant in the inception note.

| Deliverables | Participants | Deadlines |
|---|-----------------------------------|------------------|
| 1. Start-up meeting | Consultant team Steering group | 23 March |
| 2. Inception note, including detailed methods and proposed templates for the mapping, quality assessment and the capturing of use, and including the criteria for selection of sample evaluations. | Consultant team | 2 April |
| 3. Inception approval | Steering group | 7 April |
| 4. Mapping and assessment phase. Before the quality assessment and assessment of use of selected evaluations (and based on the agreed selection criteria in the inception note), the proposed evaluation sample shall be submitted <i>and approved by Sida</i> prior to assessment. | Consultant team | 8 April – 27 May |
| 5. Draft report including mapping, Quality assessment and | Consultant team | 28 May |

¹³ DAC Quality Standards for development Evaluation, OECD, 2010.

| | | |
|-----------------------------------|------------------------------------|---------|
| documentation and analysis of use | | |
| 6. Comments 1) | Steering group | 4 June |
| 7. Final report | Consultant team | 11 June |
| 8. Seminar/presentation | Consultant team Evaluation unit | 17 June |

The short inception note will describe and justify the proposed approach to answer the questions of the assignment (including a time and work plan), which will form the basis for the work and be approved by Sida/UTV. In addition, a proposed mapping matrix, selection criteria, template and method for the quality assessment work and methods to capture use of the selected evaluations should be included.

After the mapping and related analysis have been conducted, and during step four in the time schedule above, a **proposed selection of evaluations** should be shared with the Evaluation unit for approval. The selection method should be described.

The final report shall be written in English and be professionally proof read.

The final report should have clear structure and follow the report format in the Sida Studies in Evaluation Template (see Annex A). Possible limitations to the methodology and methods shall be made explicit and the consequences of these limitations discussed. Findings shall flow logically from the data, showing a clear line of evidence to support the conclusions. Conclusions should be substantiated by findings and analysis.

The report should be no more than 30 pages excluding annexes (including Terms of Reference, short Inception note, the mapping and quality assessment matrix, and other relevant documents). The evaluator shall adhere to the Sida OECD/DAC Glossary of Key Terms in Evaluation¹⁴.

The evaluator shall, upon approval of the final report, insert the report into the Sida Studies in Evaluation and submit it to Nordic Morning (in pdf-format) for publication and release in the Sida publication data base. The order is placed by sending the approved report to sida@nordicmorning.com, with a copy to the responsible evaluation manager at Sida/UTV as well as Sida's Evaluation Unit's inbox (evaluation@sida.se). Write "Sida Studies in Evaluation" in the email subject field. The following information must always be included in the order to Nordic Morning:

1. The name of the consulting company.
2. The full title of the assignment.
3. The invoice reference "ZZ980601".
4. Type of allocation "sakanslag".

¹⁴ Glossary of Key Terms in Evaluation and Results Based Management, Sida in cooperation with OECD/DAC, 2014

5. Type of order "digital publicering/publikationsdatabas.

11. Team qualification

The qualifications already stated in the framework agreement for evaluation services shall apply and a Core team member should lead work on this assignment. The Evaluation unit foresees a small consultancy team but consisting of more than one consultant.

A CV for each team member shall be included in the call-off response. It should contain a full description of relevant qualifications and professional work experience.

12. Financial and human resources

The maximum budget amount available for the assignment is 600 000 SEK.

The contact person at Sida/UTV is Susanne Mattsson. The contact person should be consulted if any problems arise during the assignment.

The consultant will be required to arrange any logistics if needed.

13. Relevant documents and contact details

At the start of the assignment, an e-mail list to relevant staff at all Embassies and relevant departments at Sida will be shared the consultants. This will form the basis upon which to reach-out and gather data for the mapping of partnerled evaluations financed by Sida (e.g the survey and follow-up activities as needed). Sida/UTV will also share any information available from our planning and follow-up tool for decentralised evaluations, including partner led Sida financed evaluations

In addition, Sida/UTV will share its template for Terms of Reference for decentralised evaluations, and checklists for quality assessments of Terms of Reference and evaluation reports (all of which can also be found in Sida's evaluation handbook), which the consultants are expected to develop further.



Meta-study of Partner-led Evaluations 2019

Based on a recommendation from its internal audit, Sida/UTV commissioned NIRAS to undertake a stocktaking and meta-assessment of partner-led evaluations that were undertaken in 2019 and early 2020. The assignment consisted of i) a stocktaking of partner-led evaluations and reviews; ii) a quality assessment of a sample; and iii) an assessment of utility of the same sample. While there is scope for improvement, the quality of the evaluation reports is good. The average utility score for the evaluations is also relatively high. Management responses have been prepared by the respective partners for the entire sample. The terms of reference for the evaluations were, however, a weak aspect.

The data suggests that quality of a partner-led evaluation depends to some extent on whether it is commissioned as an end-term or mid-term evaluation, and the size of its budget – the latter of which is often dependent on the first. Correlations between utility and quality; and the quality of the ToRs and quality or utility of the evaluation reports are weak. A larger sample size might better reveal potential correlations.