



Meta-Evaluation and Analysis of Project Evaluations 2016

Final Report to Welthungerhilfe

Andrea Queiroz de Souza
Agronomist
MA Organizational Development
MSc Agricultural Extension
Independent Consultant

Bielefeld, 28.06.2017
andreasouza@online.de

Acknowledgements and disclaimer

I would like to thank all people contacted during the meta-evaluation and analysis process for their support:

I would like to express my special gratitude to the MELA evaluation advisors Leila Broich and Bernhard Hoepfer for their tireless support, their patience with the many requests for additional information, the interesting discussions and the encouraging and critical feedback on intermediate results as well as the joint learning we all went through while reflecting on the potentials and limitations of the meta-evaluation approach. Thanks also to Saskia Wegener for compiling the contact data for the survey and for scheduling the interviews.

I appreciate that Country Directors, MELA focal points and Heads of Projects took the time to participate in the survey and to be available for interviews.

The results of the evaluation and the report are the sole responsibility of the author and can in no way be taken to reflect the views of Welthungerhilfe.

Andrea Queiroz de Souza

June 2017

Table of contents

Abbreviations	5
I Summary	6
II Main text	9
1 Introduction	9
1.1 Background	9
1.2 Purpose and objectives of the meta-evaluation and analysis	9
1.3 Meta-evaluation and analysis subject/scope	10
2 Methodology	11
2.1 Methodological concept	11
2.2 Methods	11
2.2.1 Desk review/text analysis	12
2.2.2 Content analysis	12
2.2.3 Mail survey	13
2.2.4 Interviews	13
2.3 Methodological limitations	13
3 Findings	15
3.1 Evaluation standard 'Accuracy'	15
3.1.1 Description and analysis of 'Accuracy'	15
3.1.2 Factors influencing 'Accuracy'	26
3.1.3 Assessment and conclusions on 'Accuracy'	27
3.2 Evaluation standard process quality ('Feasibility and Fairness')	28
3.2.1 Description and analysis of process quality	29
3.2.2 Factors influencing process quality	31
3.2.3 Assessment and conclusions on process quality	32
3.3 Evaluation standard 'Utility'	32
3.3.1 Description and analysis of 'Utility'	33
3.3.2 Factors influencing 'Utility'	34
3.3.3 Assessment and conclusions on 'Utility'	34
3.4 Patterns of reoccurring findings and recommendations	34
3.4.1 General patterns	35
3.4.2 General management	35

3.4.3	Working with local partners	36
3.4.4	Working with government	37
3.4.5	Other topics	37
3.4.6	Humanitarian assistance	37
3.4.7	Strategic topics	38
3.4.8	Conclusions on patterns of reoccurring findings and recommendations	38
4	Overall conclusions	40
5	Recommendations	42
	Annex 1: Terms of Reference for the Meta-Evaluation and Analysis	46
	Annex 2: List of Evaluations Included in the Meta-Evaluation and Analysis	55
	Annex 3: Evaluation Matrix	59
	Annex 4: Evaluation Quality Assessment Checklist	70
	Annex 5: Individual Checklist Evaluation Report Quality	76
	Annex 6: Mail Survey Questionnaire	78
	Annex 7: Mail Survey Results	Fehler! Textmarke nicht definiert.
	Annex 8: List of Interviewees	83
	Annex 9: Bibliography	84
	Annex 10: Specific Assessment Criteria for Accuracy and Degree of Fulfilment	85
	Annex 11: Management Response Sheet Meta-Evaluation	89

List of Figures

Figure 1: Percentage of evaluation reports fulfilling the accuracy criteria	15
Figure 2: Fulfilling the accuracy criteria – Comparison of DE and CE	16
Figure 3: Percentage of evaluation reports fulfilling the criteria on general report features	16
Figure 4: Fulfilling the accuracy general features criteria – Comparison of DE and CE	17
Figure 5: Percentage of evaluation reports fulfilling the specific assessment criteria for general report features	17
Figure 6: Percentage of evaluation reports fulfilling the assessment criteria on methodology	19
Figure 7: Fulfilling the accuracy criteria on methodology – Comparison of DE and CE	19
Figure 8: Percentage of evaluation reports fulfilling the specific assessment criteria for methodology	21
Figure 9: Percentage of evaluation reports fulfilling the criteria on 'analysis along DAC criteria'	22
Figure 10: Fulfilling the accuracy criteria on DAC analysis – Comparison of DE and CE	23
Figure 11: Percentage of evaluation reports fulfilling the specific assessment criteria for the DAC criteria analysis	23
Figure 12: Percentage of evaluation reports fulfilling the criteria on 'quality conclusions and recommendations'	25
Figure 13: Fulfilling the accuracy criteria on quality of conclusions and recommendations; Comparison of DE and CE	25
Figure 14: Percentage of evaluation reports fulfilling the specific assessment criteria for the quality of conclusions and recommendations	26
Figure 15: Perception of the evaluators' performance	29
Figure 16: Perception of other fairness and feasibility criteria	30
Figure 17: Perception of adequacy of HO/CO support to the evaluation process	31
Figure 18: Perception of evaluation utility	33

Abbreviations

CBI	Cash-Based Interventions
CD	Country Directors
CE	Centralised Evaluation
CBO	Community-Based Organisations
CO	Country Office
DAC	Development Assistance Committee
DE	Decentralised Evaluation
DeGEval	German Association for Evaluation (<i>Deutsche Gesellschaft für Evaluation</i>)
ERT	Emergency Response Team
FGD	Focus Group Discussion
GIZ	<i>Gesellschaft für Internationale Zusammenarbeit</i>
HO	Head Office
HoP	Head of Project
M&E	Monitoring and Evaluation
MELA	Monitoring, Evaluation, Learning and Accountability
RD	Regional Director
RMS	Risk Management Systems
ROM	Results-Oriented Monitoring
SMART	Specific, Measurable, Achievable, Relevant/Responsible, Time-bound
ToR	Terms of Reference
WASH	Water, Sanitation and Hygiene
WHH	Welthungerhilfe

I Summary

Introduction

Welthungerhilfe (WHH) has conducted yearly meta-evaluations of its evaluation reports since 2015. This meta-evaluation/analysis was: i) to provide structured external feedback on the quality of evaluation reports as well as on the quality of the evaluation processes using the evaluation standards Accuracy, Feasibility, Fairness and Utility; ii) to contribute to improving advisory services and information offered by the Monitoring, Evaluation, Learning and Accountability (MELA) team of the Head Office to the newly established MELA focal points in the Country Offices and other staff managing evaluations; and iii) to document reoccurring evaluation findings and recommendations from evaluation reports in order to facilitate organisational learning and strategic decision making.

The meta-evaluation/analysis contained 28 of the 30 project evaluation reports approved in 2016. It included evaluations managed from the WHH Head Office (centralised evaluations) and evaluations managed from the Country Offices (decentralised evaluations). With the ongoing decentralisation of evaluation management processes, Country Offices gradually assume the responsibility for an increasing number of project evaluations. The role of the evaluation advisors in the Head Office is accordingly changing from evaluation management to advising and monitoring functions vis-à-vis the Country Offices.

The meta-evaluation/analysis used a text and content analysis of the reports, a mail survey among Country Directors, MELA focal points and Heads of Projects and telephone and skype interviews with different actors involved in the evaluation processes to generate findings.

Findings and conclusions

The meta-evaluation/analysis has shown that the WHH actors in the countries have an overall positive perception of the quality of the evaluation processes and the utility of evaluations. To them, evaluations are to a high degree fair, feasible and useful. Of the survey respondents, 83% declared that the evaluation they were involved in was conducted in a way that encouraged the acceptance and the utilisation of the evaluation findings. It was felt that the evaluations were inclusive, provided useful findings to the project and addressed the most important issues. Evaluators were perceived as technically and methodologically competent and the advisory services from Head Offices were appreciated as professional and helpful for the process. Some elements of the evaluation process, however, still have the potential to improve; for example, the quality of recommendations which are sometimes perceived as too vague and not sufficiently feasible, the follow-up of and learning from evaluations which could be more systematic, and the support materials for evaluations which are perceived as being too difficult to obtain and adapt.

While results on the evaluation process quality were very good, the quality of the evaluation reports was moderate. Only half of the reports fully or mainly fulfilled evaluation report quality criteria. While centrally managed evaluations produced good quality reports, the report quality of the decentralised evaluations was varied: some very good, some average, but a relatively high number of low-quality reports. This result can be regarded as an outcome of decentralising the responsibility for project evaluations to the Country Offices, while not all of the Country Offices already have the capacities to manage evaluations in a way that ensures that quality standards are met.

Another challenge for WHH is the identification and utilisation of findings and recommendations which are of relevance to the strategic level of the organisation and provide information for strategic decision making. Decentralised evaluations have a tendency to focus on project specific questions and rarely generate findings of interest beyond the project context. Although for centralised evaluations there are some mechanisms and processes to promote the inclusion of strategic questions in evaluations and to provide feedback at Head Office level, these processes are not yet

established for decentralised evaluations. Thus, it becomes increasingly difficult for the Head Office to learn from project evaluations.

The meta-evaluation/analysis was only of limited utility for identifying strategic findings and recommendations that could be of value for strategic decision making. Some patterns of findings and recommendations could be identified, but they were of a general nature and did not provide substantial insights. Currently, it is mainly the strategic evaluations and to a certain extent the central evaluations that provide strategic insights.

The meta-evaluation confirms the need to reorganise the responsibilities and tasks of the evaluation advisors in the MELA team, in conformity with the draft of a recently developed MELA framework. Until recently, the main focus of the evaluation advisors was the management of the strategic and centralised evaluations. Advisory services were and still are rather ad hoc and on demand and evaluation advisors strongly feel that (time) capacities do not allow for intensifying support on top of their current tasks. The monitoring and quality assurance of decentralised evaluation management processes in the Country Offices is still not strongly developed and it is unclear to what extent quality assurance for decentralised evaluations is perceived as a mandate of MELA.

Resulting from meta-evaluation findings, the focus of the evaluation advisory activities should shift from managing evaluations to developing the capacity of the Country Offices, to enable these to successfully manage evaluations. In addition, the evaluator is of the opinion that the assurance of quality standards and the monitoring of the Country Offices' evaluation management performance should also feature more clearly in the activity portfolio of the evaluation advisors.

The MELA framework draft is pointing in the right direction and with the approval and implementation of the framework, it is expected that the decentralisation of MELA processes will progress. However, some updates of the MELA framework would be required, e.g. on the role of MELA evaluation advisors in monitoring and quality assurance and on the meta-evaluation procedure. The capacity development process on evaluation management is a long-term process that will require several years to render convincing results.

Major recommendations

Recommendation 1: The MELA evaluation advisors should more strategically address the comprehensive capacity building of the MELA focal points on evaluation management. This would include:

- Designing a capacity development strategy on evaluation management for MELA focal points
- Updating support materials (including the finalisation of evaluation standards and requirements)
- Following through with updating the evaluation information on the O365 website
- Providing regional evaluation management trainings for all MELA focal points, presenting and working with the support materials (August 2018 – March 2019)

Recommendation 2: The MELA evaluation advisors should assume the responsibility for the (coordination of) quality assurance of evaluation management. This would include:

- Defining evaluation report quality criteria which are key for WHH
- Clarifying quality assurance processes and responsibilities
- Carrying out quality checks of evaluation reports and providing feedback to Country Offices
- Monitoring quality development of evaluation reports

Recommendation 3: The MELA evaluation advisors should take the lead in the coordination of the establishment of processes and mechanisms to generate strategic findings and

recommendations from decentralised evaluations and strategic evaluations and to communicate findings of strategic interest within the Head Office.

Recommendation 4: To free resources for the key activities under recommendations 1–3, the MELA evaluation advisors should seek to reduce the number of strategic evaluations and centralised evaluations to a minimum until the above-mentioned processes have been consolidated.

Recommendation 5: MELA should assume the responsibility for setting up a network of MELA focal points in all countries and develop them into a functional community of practice. This would include:

- Appointing a person within MELA who will take the lead in establishing a MELA community of practice
- That person would lobby with Regional Directors (RD) and Country Directors for the appointment of MELA focal points for each country. These should be appointed and known by December 2017
- Establishing rapport with the MELA focal points

Recommendation 6: Country Directors should strengthen the strategic use of evaluations at country level and seek to improve the evaluation management processes in their respective country. This would include:

- Appointing a qualified MELA focal point and provide his/her contact details to MELA Head Office
- Communicating evaluation requirements to country staff
- Making sure that evaluations are already planned for and budgeted during the development of the project proposal
- Making sure management response sheets are included as a deliverable in the evaluation Terms of Reference and following up the implementation of recommendations by participating in one discussion on the operationalisation of recommendations immediately after the evaluation and participate in a second discussion several months later to follow up on the implementation status.

Recommendation 7: Sector advisors should proactively seek to know which evaluations are planned within their sectors, offer advice for designing evaluation questions and seek to include a few strategic sector evaluation questions in the evaluation's Terms of Reference. They should further seek to have a debriefing to get details on the evaluation results.

Recommendation 8: The Executive Director Programmes should communicate mandatory requirements and standards on evaluation to regional directors and country directors and demand for feedback on compliance.

Recommendation 9: The Board should approve the MELA framework and allocate the resources for its implementation. This will formalise the MELA focal points and provide the necessary leverage for the MELA team to proceed with the decentralisation of MELA responsibilities.

II Main text

1 Introduction

1.1 Background

With the Welthungerhilfe (WHH) decentralisation process, Country Offices (CO) are strengthened and empowered to improve programme quality and intensify relations with national stakeholders and donors on the ground. Consequently, responsibilities for monitoring and evaluation (M&E) are being shifted from Welthungerhilfe Head Office (HO) to the CO. This is resulting in an increasing number of project evaluations being commissioned and managed directly by the CO.

The HO Monitoring, Evaluation, Learning and Accountability (MELA) team is to focus on the commissioning and management of strategic evaluations to respond to information needs for decision making for the Executive Director Programmes and the Board of Directors. However, MELA continues to commission project evaluations in exceptional cases.

With the ongoing decentralisation of the project evaluation processes, the roles and responsibilities for the staff responsible for evaluation at HO level are changing. Staff is, inter alia, supposed to develop standards for evaluations, to provide supporting material and advisory services related to evaluations to CO and projects, and to monitor the quality of evaluations. This is to make sure that WHH evaluations adhere to international standards for quality evaluations.

To ensure a cohesive and coherent MELA work across the entire organisation, MELA focal points are supposed to be established in all CO. These are to coordinate and facilitate the MELA processes in their respective countries. At project level, depending on the duration, size and complexity of the project, M&E officers take care of MELA processes with partners, beneficiaries and other relevant stakeholders.

In 2016, WHH conducted a total of 30 project evaluations. Out of these 30, 22 were commissioned and managed by the CO (decentralised evaluations), while 8 were commissioned and managed by the MELA HO team (centralised evaluations).

Meta-evaluations, that is evaluations of the project evaluations, have been carried out by WHH since 2015 to gain insight into the quality of evaluation reports and the quality of evaluation practices, and to obtain indications for relevant strategic evaluation topics. International evaluation standards encourage meta-evaluations as an integrated element of a quality evaluation system.

1.2 Purpose and objectives of the meta-evaluation and analysis

The overall purpose of the meta-evaluation/analysis is to provide a learning opportunity for the MELA team and MELA focal points in the CO to improve the WHH evaluation system, instruments and practices.

There are several specific objectives of the meta-evaluation/analysis:

- To provide structured external feedback on the quality of evaluation reports as well as the quality of the evaluation process.
- To contribute to improving advisory services and information offered by the MELA HO team to the MELA focal points and other project staff involved in the commissioning and managing of evaluations.
- To provide information on good practices in evaluation that will help MELA focal points to improve their evaluation practices.
- To document reoccurring evaluation findings and recommendations from evaluation reports in order to facilitate organisational learning and strategic decision making.

- To reflect on opportunities and limitations of meta-evaluations as part of the WHH evaluation system. This will enable MELA to rethink the existing evaluation system and its instruments with regard to the extent it serves the information and learning needs of the organisation.

1.3 Meta-evaluation and analysis subject / scope

The meta-evaluation included 28 of the 30 project evaluation reports approved in 2016. It was originally foreseen to include all 30 evaluations, but over the course of the meta-evaluation process, it turned out that two evaluations (SYR 1025 and NPL 1025) could not be assessed using the criteria developed for assessing the report quality. This was because these two were real-time evaluations working with different sets of evaluation questions and Terms of Reference (ToR) and, therefore, the assessment checklists developed using assessment criteria on the standard / recommended ToR for decentralised and centralised evaluations could not be applied for the quality check. For the analysis of reoccurring findings and recommendations, all 30 evaluations were assessed as this analysis was carried out independently from the assessment of report quality.

Out of the 28 evaluations included, 21 were decentralised evaluations (DE) and 7 centralised evaluations (CE). While 11 evaluations were mid-term evaluations, 17 were final evaluations.

Regionally the evaluation reports were clustered as follows:

- Africa/Caribbean/Middle East and North Africa: 17
- Asia/South America: 11

According to the thematic portfolio, the evaluation reports covered the following sectors:

- Agriculture and Environment: 10
- Civil Society and Social Empowerment: 7
- Water, Sanitation and Hygiene (WASH): 4
- Humanitarian Aid: 2
- Economic Development: 4
- Nutrition: 1

The meta-evaluation sought to answer the evaluation questions as stated in the ToR (annex 1), covering four major areas of interest to Welthungerhilfe:

- **Evaluation standard 'Accuracy':**¹ This looks at the quality of the evaluation reports. The accuracy standards are intended to ensure that an evaluation produces and conveys valid and comprehensible information and results on the respective evaluation object and the evaluation questions. In this context, scientific criteria of quality are of particular importance.
- **Evaluation standards 'Utility', 'Feasibility' and 'Fairness':** These standards reflect the process quality of the evaluation. In how far were evaluations able to ensure that they served the information needs of the intended users (utility)? To what extent were evaluations realistic, prudent, diplomatic and efficient (feasibility)? Were evaluations conducted in a legally and ethically acceptable way and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results (fairness)?
- **Patterns of reoccurring findings and recommendations:** This was to identify patterns of reoccurring findings and recommendations that could be relevant beyond the project

¹ The German Evaluation Association DeGEval has defined four evaluation standards: Accuracy, Utility, Feasibility and Fairness. For all four standards, there is a subset of criteria detailing and describing the main criteria (see: <http://www.degeval.de/de/degeval-standards/>).

context and bear learning potential for WHH as an organisation and serve as a basis for strategic decision making.

- **Methodological reflection of the meta-evaluation:** This was a reflection on the potentials and limitations of the meta-evaluation tool and on design options for future meta-evaluations. As in the past, most meta-evaluations tended to focus mainly on the quality of evaluation reports with a strong focus on formal requirements, with limited consideration of the other evaluation standards, the current meta-evaluation had a number of innovative elements that were to provide useful orientation to WHH for meta-evaluation as a concept.

2 Methodology

2.1 Methodological concept

Participation: As the overall purpose of the meta-evaluation/analysis was to provide a learning opportunity for the MELA team, the methodological concept followed a participatory approach, involving relevant stakeholders as much as possible within the given timeframe and with the given resources into all stages of the evaluation process.

Frequent feedback loops during all stages of the evaluation (briefing meeting, feedback on inception report and the data collection instruments, sharing and discussion on preliminary findings, feedback on the draft evaluation report, debriefing meeting, learning event meeting) provided opportunities for active participation for the MELA team, the MELA focal points and other interested stakeholders within the organisation. By creating a space for the joint work on documents in O365, MELA focal points were invited to join the discussion on the inception report, the tools developed for data collection, comment on the draft final report, give suggestions for recommendations, etc. This also contributed to increasing the transparency of the evaluation process.

Adherence to evaluation standards: The meta-evaluation/analysis itself sought to adhere to the evaluation standards of Accuracy, Utility, Feasibility and Fairness.²

Triangulation: Combining both qualitative and quantitative data collection methods as a basis for the evaluation (methodological triangulation) allowed for making use of the comparative strengths of both forms of inquiry, while compensating for the respective weaknesses. A triangulation of sources was applied by using different sources of information (i.e. that the views of all major stakeholders were included into the evaluation process).

2.2 Methods

In phase 1 (inception) of the meta-evaluation/analysis, the evaluator reviewed the existing documents, clarified the ToR and discussed options for designing the process during a briefing meeting with the MELA HO evaluation advisors. She elaborated an inception report presenting in detail the approach and methodology for the evaluation and the data collection tools.

Phase 2 focused on data collection and exploration. During this phase, the major data collection tools – desk review / text analysis, content analysis, interviews, mail survey – were implemented.

² For example, by using data collection methods that provided valid and reliable findings and basing conclusions on findings which were based on explicit criteria (accuracy), clarifying and taking into account information needs of relevant stakeholders with regard to the meta-evaluation/analysis (utility), choosing appropriate methods, balancing efforts and results (feasibility), providing a fair and balanced presentation of strengths and weaknesses of the evaluation subject (fairness).

In phase 3 (synthesis phase), the evaluator analysed and triangulated the results of the different assessments and brought them together in this comprehensive evaluation report. The synthesis phase was also used to prepare the additional deliverables.³

Phase 4 is to address the dissemination of results and the learning the meta-evaluation/analysis can provide for the organisation. It includes a debriefing meeting with MELA HO staff and a learning event for WHH HO staff.

The major methods used are presented in more detail in the following sub-chapters.

2.2.1 Desk review / text analysis

The main tool for analysing the quality of evaluation reports was a desk review using text analysis. The evaluator had developed a checklist for the assessment of the quality of evaluation reports (see annexes 4 and 5) based on the evaluation standard 'accuracy' of the German Association for Evaluation (DeGEval). All other DeGEval evaluation standards could not be assessed by analysing the evaluation reports, but needed other tools. The checklist was adapted to the specific requirements of WHH and the ToR for the meta-evaluation/analysis. It was developed drawing on experiences from other existing checklists, the checklist used by WHH in the past and the quality criteria WHH communicates through its evaluation ToR and documents on M&E within the organisation.

The checklist was pretested and adapted by the evaluator and one of the M&E advisors of the MELA HO team before applying it to all reports. It defined criteria for analysis with a scoring system for four major criteria areas: formal overall quality, presentation of the methodology, assessment of the evaluation criteria of the Development Assistance Committee (DAC), and the quality of conclusions and recommendations. The scoring system was explained in detail in the quality assessment checklist to allow for comparability and traceability. An individual assessment sheet was filled in for every report. The summary of all sheets allowed for an overall quantitative assessment of reporting quality.

As a second element, the desk review included an analysis of already existing data regarding the evaluation questions. For example, MELA had already used a survey for assessing the quality of the evaluation process and the utility of the evaluation for the centralised evaluations. A survey among MELA focal points had already provided useful information on the support needs regarding advisory services from HO. The results of these surveys were included in the meta-evaluation assessment with regard to the evaluation standards of Utility, Fairness and Feasibility and for developing recommendations.

2.2.2 Content analysis

Mayring's content analysis was used across all 30 evaluations' recommendations⁴ to identify patterns of reoccurring findings and recommendations that could be relevant beyond the project context and that could bear learning potential for WHH. This method is characterised by three basic steps: i) all statements (in this case all recommendations) are listed and grouped along defined topics; ii) the statements are generalised; and iii) the recommendations are summarised and reduced to common statements on the topic.

³ 'Good Evaluation Practices' to document good evaluation practice examples from the 28 evaluation reports as an orientation for MELA focal points. 'Scenarios for a Meta-Evaluation/-Analysis' to present findings and recommendations directed to the question of how meta-evaluations and -analysis could complement the Welthungerhilfe evaluation system. These deliverables were submitted as separate stand-alone documents.

⁴ As the content analysis was carried out before the decision was taken to focus the analysis of report quality only on 28 evaluations, evaluation recommendations were analysed for all 30 evaluations of 2016.

2.2.3 Mail survey

A mail survey was conducted among the Heads of Project of all projects evaluated in 2016 and the Country Directors and MELA focal points of the respective countries. The focus of the online survey was on the quality of the evaluation process and the three evaluation standards 'utility, fairness and feasibility', which could not be explored through the text analysis of the evaluation reports. The survey provided quantitative as well as qualitative information on these topics. (See annex 7 for the questionnaire with the responses.) The survey questionnaire was mailed to 59 potential respondents with 26 answering the questionnaire. This response rate of 44% corresponds to the response rate of the meta-study conducted in 2015 and can be regarded as moderate in view of the targeted stakeholder groups.

2.2.4 Interviews

The mainly quantitative methods described above were complemented by 14 interviews with selected relevant stakeholders (the Executive Director Programmes, three MELA HO staff, one MELA focal point, two (former) Country Directors (CD), two Programme Coordinators, three other HO staff, one (former) Head of Projects (HoP), the Executive Director Programmes and an international evaluator). (See annex 8 list of interviewees.) Interviewees were picked by their functions, trying to cover the entire range of relevant stakeholders in the evaluation process. Within the sub-groups, it was often the availability of the interviewees that guided the choice. Interviews were particularly relevant for providing insights on the perceived quality of the evaluation process and on learning through evaluations at different levels. They were also a major source for the methodological reflection on the meta-evaluation as part of the Welthungerhilfe evaluation system. As stakeholder groups were very diverse, as well as the topics they could respond to, the different questionnaires for guiding the semi-structured interviews were drafted ad hoc during the evaluation process for each interviewee / stakeholder group.

2.3 Methodological limitations

There were several methodological limitations to the meta-evaluation.

Text analysis is always subject to a certain degree of subjectivity. Having two people assess each report could have reduced this aspect, but resources did not allow for it. However, the evaluator tried to minimise subjectivity as much as possible by defining comparable and traceable judgement criteria with clear indicators, by discussing the details for the different criteria with the MELA HO team and by conducting a joint pre-test.

For the DAC criteria, it turned out to be too difficult to measure the quality of the analysis. Therefore, the assessment focused on the question, if certain key elements of the DAC criteria had been covered at all.

Because of the overall low number of evaluation reports analysed and the high standard deviations caused by the wide range of results, the data base was not suitable for a statistical back-up of the results. Findings should therefore be interpreted as trends.

The analysis of reports allowed findings on the quality of the report, but was of limited value with regard to the quality of the evaluation itself. It is likely that some evaluators have assessed a certain topic or have reflected on the limitations of the methods used, but have failed to document it. To assess this, one would have had to go far beyond text analysis. A good or poor report quality in terms of accuracy does not allow for a judgement on the quality of the evaluation itself or on the other evaluation standards. However, a quality evaluation report is regarded as key for the credibility of an evaluation and, therefore, the majority of meta-evaluations focus to a high extent on accuracy.

The interpretations of the perception of the evaluation standards Feasibility, Fairness and Utility are likely to have a positive bias, as the evaluator interpreted full agreement and the category 'I rather agree' as positive perceptions on the respective question. However, during data analysis, she noticed that in some cultural contexts it does not seem common to use the full range of possible opinions. Relating the comments to the perception scale, many national colleagues seemed to use 'I rather agree' to express a rather negative perception / disagreement. Nevertheless, by comparing the answers to the different questions, it was possible to identify the strengths and weaknesses related to the process quality.

It was not possible to compare the results of the meta-evaluation with the results of meta-studies of previous years as the assessment criteria used were different. Even using the same quality assessment checklist would not have rendered comparable results, because in previous years the criteria for assessing evaluations as 'good' or 'poor' were not documented in a way to allow for replication.

Two real-time evaluations that initially were meant to be included had to be taken out, because the assessment criteria for the report quality did not match well with the specific design of those evaluations.

With a very broad spectrum to be covered in a limited timeframe, it was not possible to explore all topics in-depth. It was not possible to include all stakeholder groups (e.g. project staff and partner organisations who should be using the recommendations and therefore have an opinion on the utility of the evaluation) and to include all stakeholders' groups to the same extent. Opinions of some stakeholder groups featured more prominently than others.

The assessment of recommendation patterns was limited by the wide variety of projects from different sectors.

The inclusion of MELA focal points into the learning process of the meta-evaluation proved to be difficult for several reasons. MELA focal points were not included in the preparatory activities of the evaluation, so the first effort to include them was rather late at the time when the inception report was uploaded and comments were invited. As the MELA discussion forum on the WHH intranet is not yet well established as an exchange forum, postings there did not generate any comments and discussions. As in the beginning, there was no comprehensive contact list of MELA focal points available and there was not yet an established, functioning communication platform, communication was limited.

During the evaluation process, the evaluator and MELA HO were not successful in motivating the MELA focal points to participate more actively than just providing information and participating in the mail survey and the interviews.

3 Findings

3.1 Evaluation standard ‘Accuracy’

The evaluation standard ‘Accuracy’ looks at the quality of the evaluation reports. The accuracy standards are intended to ensure that an evaluation produces and conveys valid and comprehensible information and results on the respective evaluation object and the evaluation questions.

The following sub-chapters provide a description and analysis of the findings along the criteria and discuss possible influencing factors and conclusions.

3.1.1 Description and analysis of ‘Accuracy’

The evaluation standard accuracy was assessed using four major quality categories with 20 specific assessment criteria subdivided into 49 indicators (see annex 10).

Overall, the assessment of accuracy rated half of the reports as fully or mainly fulfilling the accuracy criteria, while half of the reports were rated as rather unsatisfying or unsatisfying.⁵

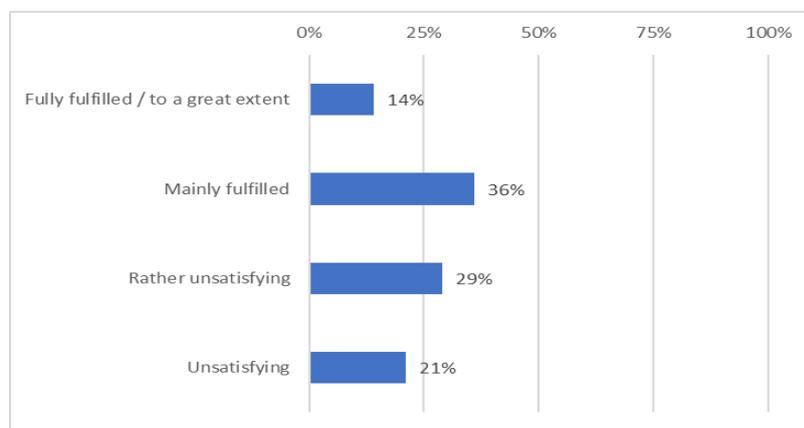


Figure 1: Percentage of evaluation reports fulfilling the accuracy criteria

There was a distinct difference between the DE and the CE reports fulfilling the accuracy criteria. While 85% of CE were assessed as fully or mainly fulfilling the accuracy criteria without any report being unsatisfactory, only 38% of the DE were assessed as fully or mainly satisfying.

⁵ The details of the scoring system are presented in annex 4 Evaluation Quality Assessment Checklist.

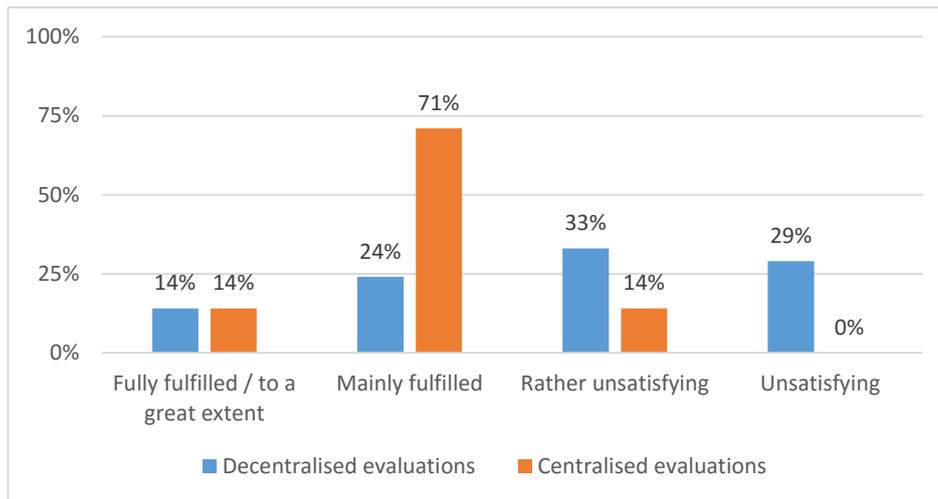


Figure 2: Fulfilling the accuracy criteria – Comparison of DE and CE

Although the CE report quality was rather uniform, ranging from 45 to 52 points in the overall scoring system, the quality of the DE reports showed a high variability, ranging from 17 to 56.5 points; that is, the best and the worst reports were DE.

3.1.1.1 General features

The first assessment category ‘general features’ shows to which extent general report quality features have been respected. The category used four specific assessment criteria. These measured, if the report structure was clear and coherent (1.1); the executive summary was a stand-alone section presenting the main information of the evaluation (1.2); the length of the report was adequate to cover the major aspects of the evaluation while at the same time being economic to read (1.3); and the report included an assessment of how the project addressed gender issues and how women / men benefited from project interventions (1.4). In total, nine indicators were assigned to the criteria.

The analysis showed a high degree of fulfilment for this category. Some 83% of the reports fully or mainly fulfilled the quality criteria on general features.

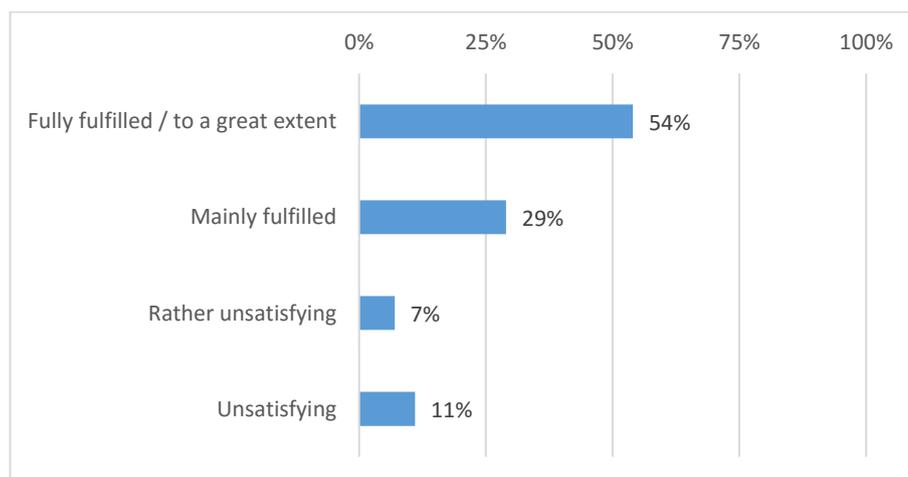


Figure 3: Percentage of evaluation reports fulfilling the criteria on general report features

Figure 4 compares performance of DE and CE in fulfilling the accuracy criteria on general features. While 100% of CE fulfilled the criteria, for DE it was 77% of the reports with full or main fulfilment.

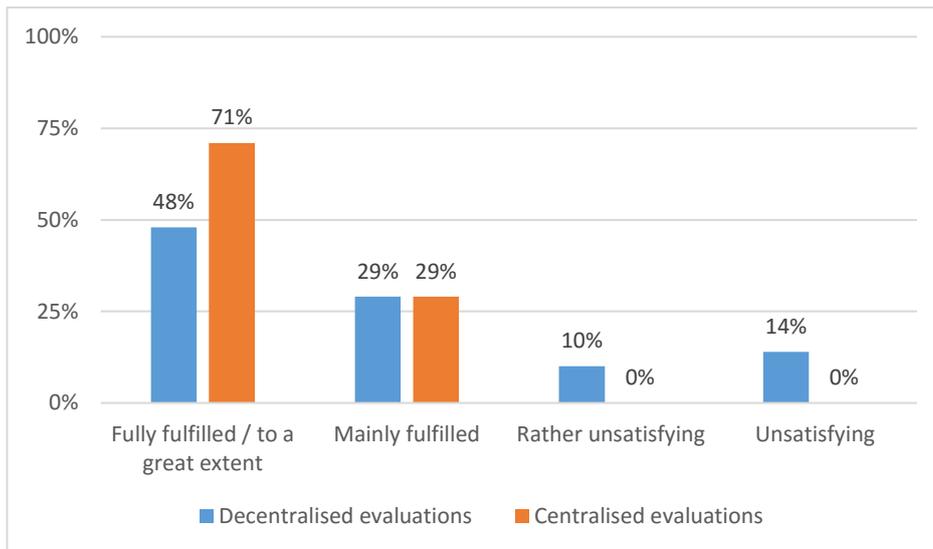


Figure 4: Fulfilling the accuracy general features criteria – Comparison of DE and CE

Clear and coherent report structure: This criterion looked at the clarity of the reporting structure. Was the sequencing of findings, conclusions and recommendations as suggested in the WHH standard ToR for evaluations respected? Were background and evaluation objectives presented before findings, findings before conclusions and conclusions before recommendations?

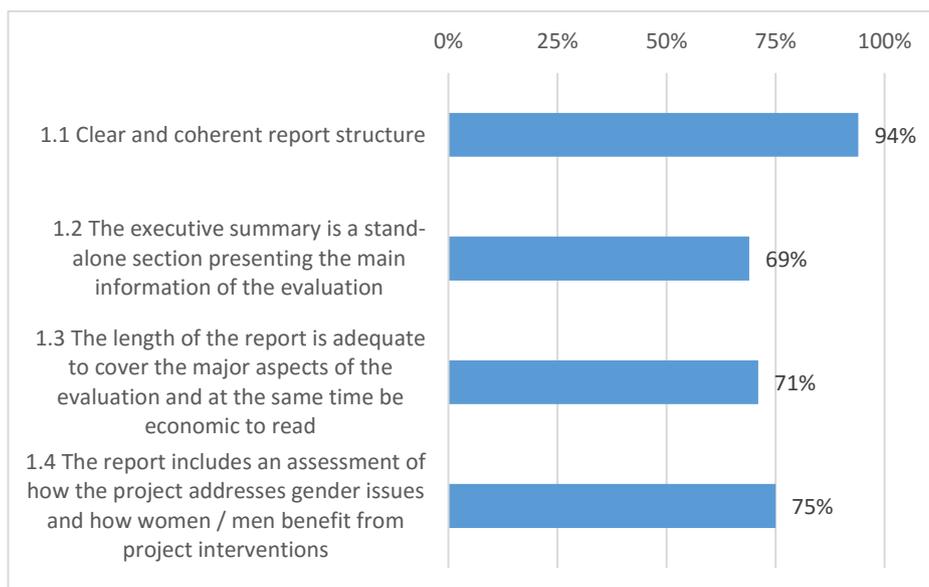


Figure 5: Percentage of evaluation reports fulfilling the specific assessment criteria for general report features

An average of 94% of the reports followed this logical sequencing with only a few reports presenting recommendations before the actual conclusions, which provided the rationale for developing the recommendations. Thus, the quality of the report structure was rated as very high.

Executive summary: The executive summary should be a stand-alone section presenting the main information of the evaluation. The indicators applied for rating the quality of the executive summary included the questions if it was comprehensible as a stand-alone document, if it presented the major elements of the main report, and if the language was adequate to allow publishing. This question is rather important for WHH, as currently the summaries of the central evaluations are published on the WHH homepage, while the summaries of the decentralised evaluations are not. Overall, the average degree of indicator fulfilment was 69%, so more than two thirds of the reports had a good quality executive summary and almost one third had not. There

was a distinct quality difference between the CE and the DE. Centralised evaluations fulfilled the quality criteria for the executive summary to almost 100%; the average fulfilment rate for the decentralised evaluations was only at 55%.⁶

Some 64% of the reports presented a summary that was comprehensible as a stand-alone document and contained all major elements of the evaluation. Factors reducing the stand-alone quality of summaries were mainly due to a failure to provide any general information on the project and initiating with the evaluation objectives, without briefly introducing the project with its objective, major results and basic key data. Thus, an external reader does not get any idea on what the project is about. More than half of the DE reports did not include information on all DAC criteria in the summary, even though the DAC criteria for the vast majority of evaluations formed the key structuring elements for presenting the evaluation findings.

Of the evaluation reports, 79% had a language quality that was sufficiently good to allow for publishing. Almost one third of the DE reports demonstrated that the authors' writing skills in the report language were insufficient.

Even though this was not part of the assessment criteria, the evaluator noted that many summaries were much longer than the three to four pages recommended in the WHH standard terms of reference. Long summaries tended to negatively influence readability.

Length of report: The third assessment criterion rated, if the length of the main report was adequate to cover the major aspects of the evaluation, while at the same time being economic to read: 20 to 40 pages were regarded as an acceptable range, as the standard ToR propose a maximum length of 35 pages. 71% of the evaluation reports' main section was between 20 and 40 pages long. The remaining 29% were either shorter or longer.

Addressing gender: The fourth criterion looked into the extent gender issues were addressed in the report, using two indicators. One assessing in how far the report discussed if and how the project addressed gender issues in its interventions; and the other to what extent effectiveness and outcome analysis included observations on how men and women benefited differently from project interventions. 75% of the evaluation reports addressed these gender dimensions, with a slightly higher inclusion of discussing gender specific interventions (79%) in comparison to a disaggregation of gender with regard to benefits and outcomes (71%). 25% of the evaluation reports failed to address gender issues.

3.1.1.2 Methodology / validity and reliability of findings

The second assessment category 'Methodology / validity and reliability of findings' explored in how far the methodology of the evaluation was clearly spelled out. It should be noted that the meta-evaluation could only look at the presentation of the methodology. This did not allow for any judgement on the quality of the applied methodology itself. The category used seven specific assessment criteria. These measured: if an inception report or minutes of the kick-off meeting replacing the inception report had been part of the methodology (2.1); if the report described the data collection methods and analysis and the rationale for selecting them (2.2); if the report described the data sources and the rationale for their selection (2.3); if the report included a discussion on how the mix of data sources was used to obtain a diversity of perspectives (2.4); if the report described the sampling for the evaluation process (2.5); if the report explained in how far the evaluation had avoided duplications in data collection by relying as far as possible on

⁶ Differences between centralised (CE) reports and decentralised (DE) reports: Comprehensible stand-alone document: CE reports 100%, DE reports 52%; major elements: CE reports 87% all major elements, 13% most major elements, DE reports 43% all major elements, 24% most major elements, 33% insufficient; adequate language: CE reports 100%, DE reports 71%.

existing data (2.6); and if the report explained if the evaluation was designed as a participatory process (2.7). In total, 16 indicators were assigned to the criteria.

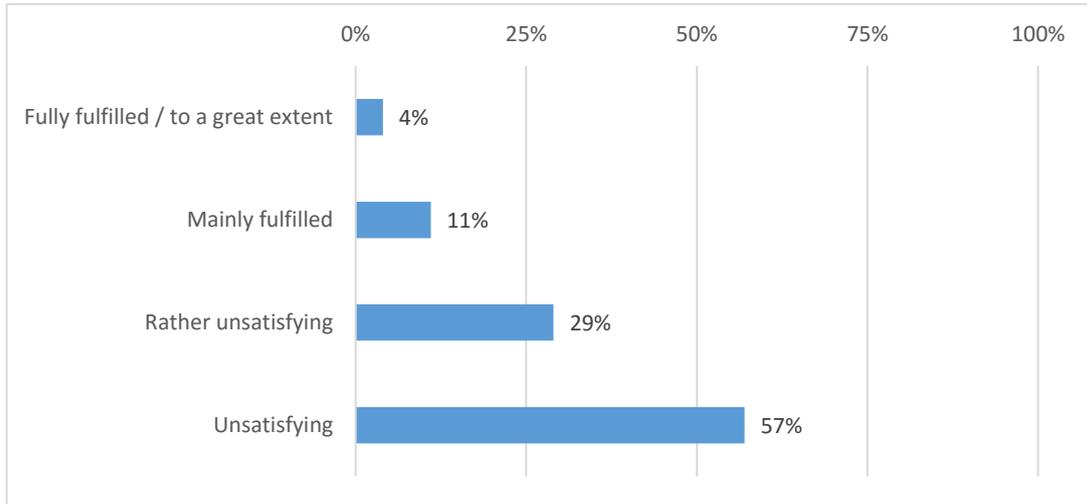


Figure 6: Percentage of evaluation reports fulfilling the assessment criteria on methodology

The analysis showed the lowest degree of fulfilment compared to the other three assessment categories. Only 15% of the reports fulfilled the methodology criteria fully or mainly; 86% were rather unsatisfying or unsatisfying. The majority of reports presented the methodology on half a page or less. This could be regarded as an indicator that up to now WHH has not emphasised the importance of a detailed methodology description. The ToR do not provide specific guidance on the topic.

Figure 7 compares performance of DE and CE in fulfilling the accuracy criteria on methodology. Here, the DE performed better than the CE. While 19% of DE fulfilled the methodology criteria fully or mainly, none of the CE did. All CE were ‘rather unsatisfying’ or ‘unsatisfying’.

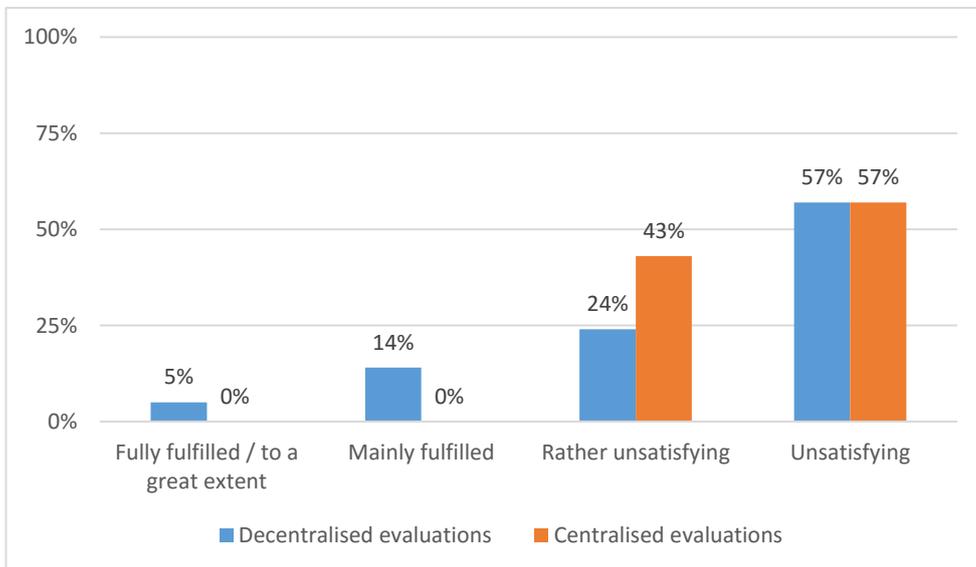


Figure 7: Fulfilling the accuracy criteria on methodology – Comparison of DE and CE

Inception report: This criterion looked at the existence of an inception report or, alternatively, the existence of minutes of the kick-off meeting as a ‘lighter’ form of an inception report. While an inception report was listed in the ToR for CE as a deliverable supposed to outline the methodology, operationalise the questions and plan the approach and methodology, the ToR for DE suggested

either an inception report or kick-off meeting minutes.⁷ Only three evaluations (11%) stated that an inception report was elaborated as part of the evaluation process. No minutes of kick-off meetings were mentioned in any of the reports.

Data collection methods: The second criterion considered the presentation of data collection methods using four indicators. The first assessed the mere description of methods; the second the description of the rationale for selecting them; the third the plausibility of choice with regard to the evaluation context and the resources available for the evaluation; and the fourth on how far the report discussed the limitations of the chosen methods. Overall, the rating of the data collection methods was 68% of reports fulfilling the indicators. Results for the specific indicators varied to a high degree.

Almost all reports (96%) described in more or less detail the data collection methods. Some only listed them in one sentence; others provided short explanations on each method. The plausibility of choosing the data collection methods presented was equally high (96%). The evaluator did not detect any methods that might not have been appropriate for the context or that did not sufficiently consider the available resources.

Fewer reports (46%) critically reflected the chosen methods and discussed their limitations and even less (32%) described the rationale for the selection of the data collection methods.

Data sources: The criterion on data sources considered the extent to which the sources were presented and worked with two indicators: one stating whether the report described the data sources; and the second assessing if the rationale for the selection of the data sources was presented. Overall, the rating of the data collection methods was 64% of reports fulfilling the indicators. But again, the ratings for the first indicator varied widely from the ratings for the second.

⁷ Currently, WHH is using different standard / recommended ToR for DE and for CE. Overall, the two templates are similar, but the ToR for CE are more detailed than the ToR for the DE.

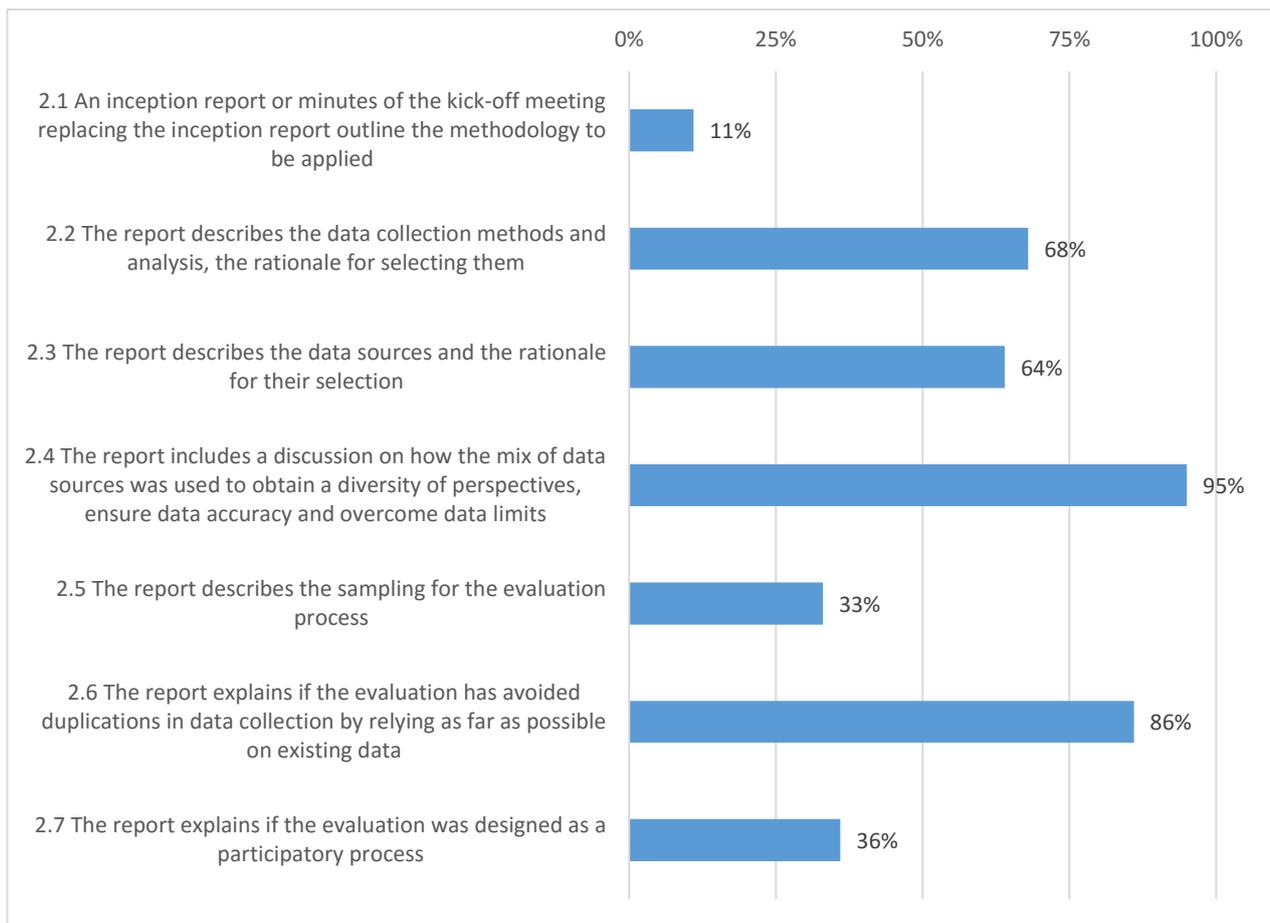


Figure 8: Percentage of evaluation reports fulfilling the specific assessment criteria for methodology

Almost all reports (96%) described in more or less detail the data sources. Some only listed them in one sentence, while others provided a more detailed explanation. Only 32% of reports described the rationale for selecting the data sources.

Triangulation: From almost all reports (95%) it was obvious that a triangulation of methods – i.e. using at least two different methods to obtain findings (93%) and a triangulation of data using at least two different sources and including the views of different stakeholder groups (96%) – was applied.

Sampling: Sampling rated lowest of all assessment criteria (33% fulfilment). The four indicators looked at the presentation of the sample frame, the description of the rationale for selecting the sample, the stating of numbers for the selected sample, and the description of sample limitations.

While 46% of the reports provided some description of the sample frame and some numbers, only 29% described the rationale for selection and only 11% reflected on the limitations of the sample. More detailed descriptions were usually provided by the few reports that focused on quantitative methods – e.g. surveys. Most of the evaluations focused on qualitative methods and apparently did not consider a description of the sampling as relevant for presentation.

Utilisation of existing data: This criterion provided information on the extent to which reports have indicated the utilisation of already existing data (e.g. baseline surveys, monitoring and end line data or data from partners). This criterion rated highest in the category – 86% of reports referred to existing data on which to base their findings. These were mainly project baselines and monitoring data.

Participatory methodology: Only a few reports (36%) provided an explanation on participatory methodologies. Again, the difference between indicators was high. While 57% of reports mentioned

the participatory nature of their approach, only 14% stated reasons for selecting the particular level of participation the evaluation applied.

3.1.1.3 Analysis along DAC criteria

The third assessment category was 'analysis along DAC criteria'. The DAC criteria are regarded as central elements of evaluations in the frame of development cooperation and should be adequately covered in each evaluation. The meta-evaluation assessed the category along six specific assessment criteria, mainly: analysing to which extent the reports adequately covered the five DAC criteria relevance (3.1); effectiveness (3.2); efficiency (3.3); outcomes and impact (3.4); and sustainability (3.5), using three indicators for each DAC criterion. The last criterion (3.6) considered if the report related findings to evidence derived from the data collection and analysis. In total, 16 indicators were assigned to the criteria.

The analysis showed a moderate degree of fulfilment, with 50% of the reports adequately covering the DAC criteria and 50% of the reports not.

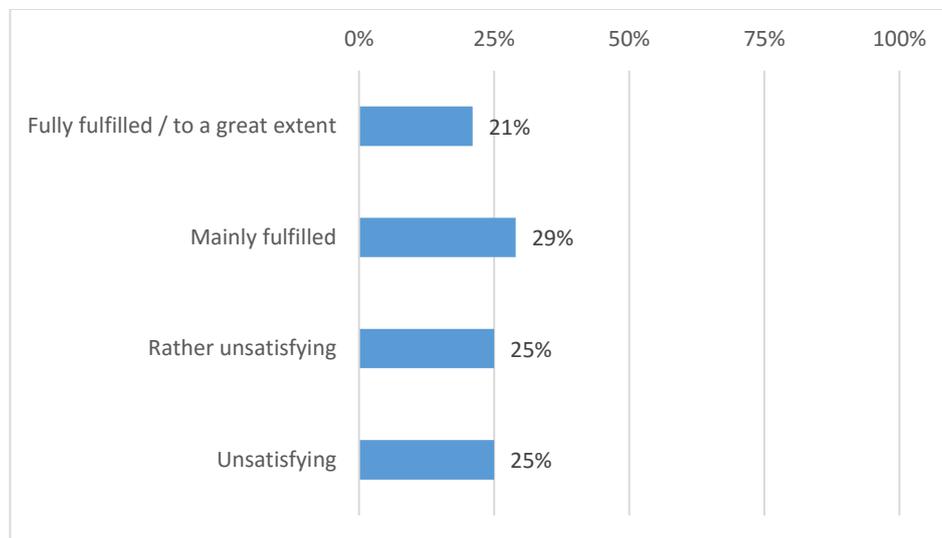


Figure 9: Percentage of evaluation reports fulfilling the criteria on 'analysis along DAC criteria'

Figure 10 compared performance of DE and CE in fulfilling the accuracy criteria on the analysis of the DAC criteria. While 86% of CE fulfilled the criteria fully or mainly, with only 14% of rather unsatisfying reports, 62% of DE reports were rather unsatisfying or unsatisfying. One third of DE reports were unsatisfying with regard to the analysis of DAC criteria.

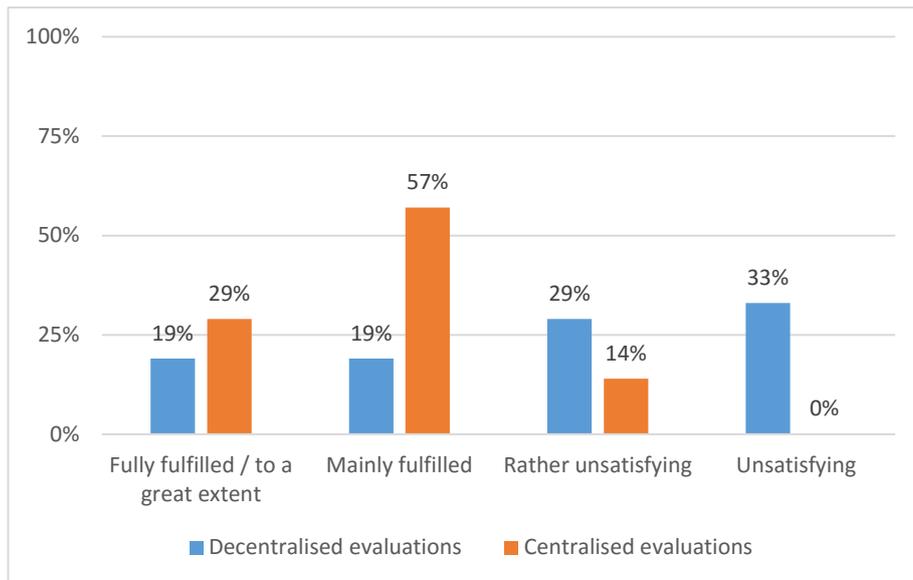


Figure 10: Fulfilling the accuracy criteria on DAC analysis – comparison of DE and CE

Relevance: The first criterion considered the adequate assessment of relevance using three indicators. The indicators explored aspects: Did the reports discuss to which extent the activities and outputs of the projects were consistent with the overall goal? Did the reports discuss to what extent the project addressed a core problem of the target groups? Did the reports discuss to what extent the project was in line with relevant strategies (e.g. WHH strategies or international / national / donor strategies) and objectives?

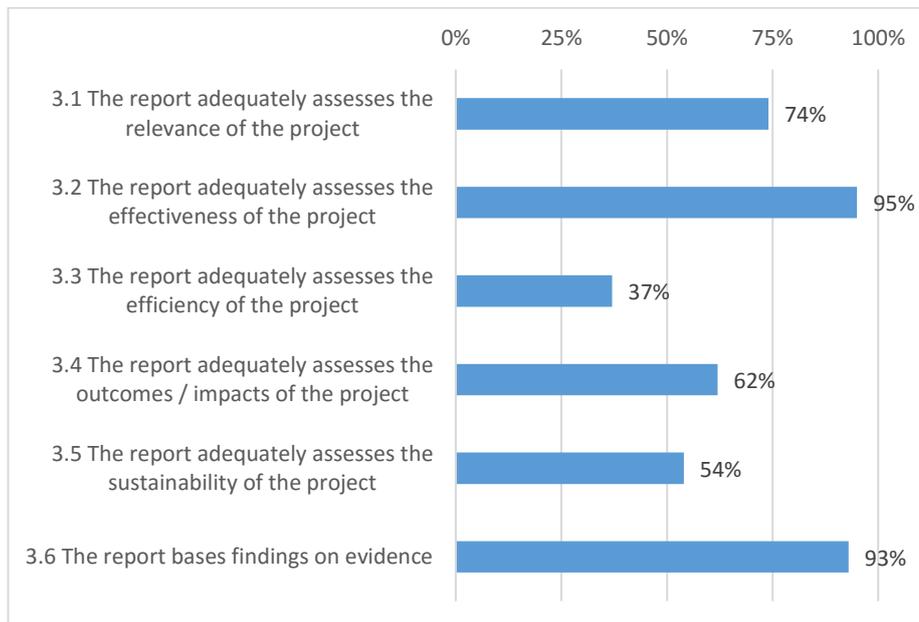


Figure 11: Percentage of evaluation reports fulfilling the specific assessment criteria for the DAC criteria analysis

Overall, the rating of the relevance criterion was 74% of reports fulfilling the indicators with some variations between indicators.

Some 50% of the reports included a reflection on the consistency of activities and outputs with specific objectives and the overall goal. Those reports which included a sub-chapter on the analysis of the results chain usually reflected on the results logic, including an analysis of the consistency of the theory of change. Half of the reports did not discuss this aspect at all; 89% of reports discussed the relevance of the project for the core problems of the target groups and 82% the

relevance in the context of strategies and objectives, showing that the questions 'relevant for whom?' was well anchored in the relevance reflections.

Effectiveness: The criterion assessed along three indicators, if the evaluation reports presented findings on the extent to which objectives had been achieved, if reports referred to the logframe indicators to assess the effectiveness of the project or explained why the indicators were not used for analysis, and if the reports explained which factors influenced the achievement or non-achievement of results.

The overall results on effectiveness scored highest among all criteria of the analysis (95%). Almost all reports (96%) discussed to which extent the project achieved its objectives and used the logframe indicators as a reference point for their analysis (89%) or explained the challenges and limitations to using the indicators. All reports (100%) reflected on factors influencing the achievement or non-achievement of the objectives. The high fulfilment reflects the importance of the effectiveness criterion for WHH. Evaluation objectives, questions and the ToR usually had a strong focus on the effectiveness criterion.

Efficiency: The analysis of efficiency reporting considered if reports discussed cost/output relations, if reports reflected on alternatives which would have allowed for a more efficient use of resources, and if reports presented plausible reasons for rating efficiency as high or low. This criterion scored lowest among the DAC criteria (38%).

A third of the reports (33%) attempted discussing examples on cost/output relations and reflected on the existence for the cost-efficient use of resources (33%). Almost half of the reports (48%) presented plausible reasons for their efficiency rating. Across the reports, the evaluator noted a high diversity of interpreting the efficiency criterion and thus different topics were addressed under the criterion. As evaluators found cost calculations in the framework of the evaluation and with the available data challenging, common topics of reflection were average costs per beneficiary, the degree of expenditures in relation to the timeframe of the project and the timeliness of delivery of project results. As the CE standard ToR efficiency chapter has a sub-chapter on the quality of the project's M&E system, many reports included a discussion on the project's M&E.

Outcomes and impacts: With regard to outcomes and impact, evaluations were supposed to reflect on the existence of evidence for outcomes and impact, the plausibility of the project contributing to long-term change and (positive or negative) outcomes which were unintended. Overall, 62% of the reports fulfilled the criteria on outcome and impact analysis.

Most of the reports (77%) reflected on the existence of evidence for outcomes and impact and on the plausibility of the project's contribution to long-term change (70%). Fewer reports (40%) included a reflection on unintended outcomes, indicating that evaluations focused on assessing achievement along intended changes as stated in the objectives and did not sufficiently consider possible negative outcomes of the intervention.

Apart from the indicators, the evaluator noted that there was a lot of overlapping with the effectiveness discussion, as there is no clear distinction between the achieved objectives and outcomes. While some evaluators solved this pragmatically by pointing to the effectiveness chapter, others presented the observed changes twice in detail.

Sustainability: Under the sustainability criterion, one would have expected a reflection on the likely degree of sustainability, on likely challenges for sustainability and on the existence and quality of exit strategies to enhance the likeliness of sustainability. Overall, the fulfilment of the sustainability criteria was rated with 54%.

Most reports reflected on the likeliness of sustainability for outputs / outcomes achieved (77%), while the reflection on likely challenges for sustainability was a bit lower (67%). Only very few reports (13%) discussed if projects took action and what projects did in order to enhance the likeliness of sustainability (e.g. through an exit strategy).

Findings based on evidence: 93% of the reports sought to base their findings on evidence. The degree to which they successfully identified existing evidence depended to a high degree on the quality of the project’s monitoring system.

3.1.1.4 Quality of conclusions and recommendations

The fourth assessment category ‘quality of conclusions and recommendations’ reflected the importance of the learning aspect of WHH evaluations. Good quality conclusions and recommendations are required to improve project performance and to learn lessons for the design of successor projects and similar projects in the region / sector. The three specific assessment criteria assessed to which extent conclusions were based on findings (4.1), conclusions related to the evaluation purpose (4.2) and good quality recommendations were provided (4.3). Overall, eight indicators were assigned to the criteria.

The analysis showed a moderate degree of fulfilment, with 54% of the reports with a good / satisfying conclusion and recommendation quality and 46% of the reports not.

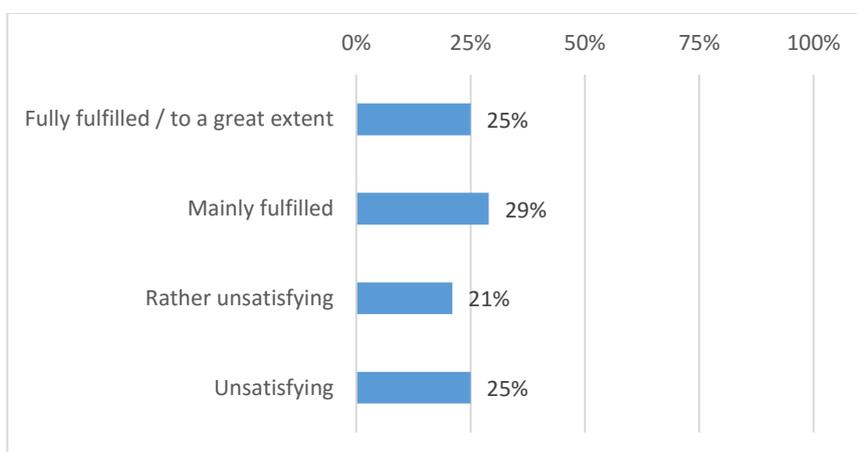


Figure 12: Percentage of evaluation reports fulfilling the criteria on ‘quality conclusions and recommendations’

Figure 13 compares performance of DE and CE in fulfilling the accuracy criteria on the quality of recommendations. While 85% of CE fulfilled the criteria fully or mainly, with only 14% of rather unsatisfying reports, 62% of DE reports were rather unsatisfying or unsatisfying. Furthermore, 29% of DE reports were unsatisfying with regard to the quality of conclusions and recommendations.

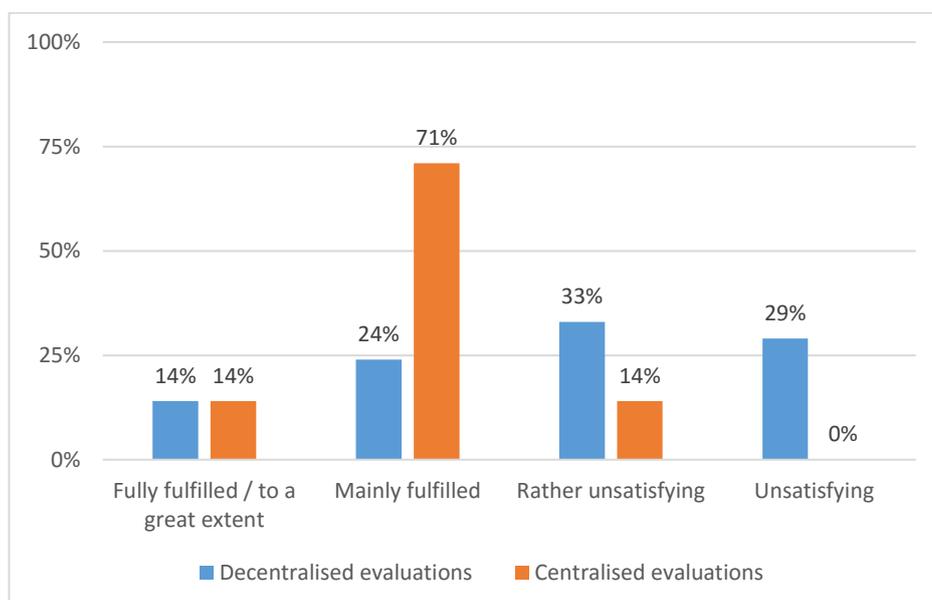


Figure 13: Fulfilling the accuracy criteria on quality of conclusions and recommendations – Comparison of DE and CE

Conclusions are based on findings: The first criterion considered the extent to which conclusions were based on findings presented in the findings section of the full report using two indicators. The indicators explored, if reports made a clear distinction between findings / analysis, conclusions and recommendations and if the majority of conclusions could be related to preceding findings. 94% of the reports fulfilled this criterion, with 97% clearly distinguishing findings, conclusions and recommendations and 90% basing most of the conclusions on findings.

Conclusions relate to the evaluation purpose: There was a strong linkage between the conclusions and the stated evaluation purpose, with 93% of the reports fulfilling this criterion.

Quality of recommendations: The meta-evaluation used five indicators to assess the quality of recommendations. The indicators explored: if the reports stated recommendations, if there was a clear link between findings, conclusions and recommendations, if recommendations were targeting at different actors, if recommendations were SMART⁸ and if the number of recommendations was adequate. Overall, fulfilment of the criterion was at 69% with high variations between the results for the specific indicators.

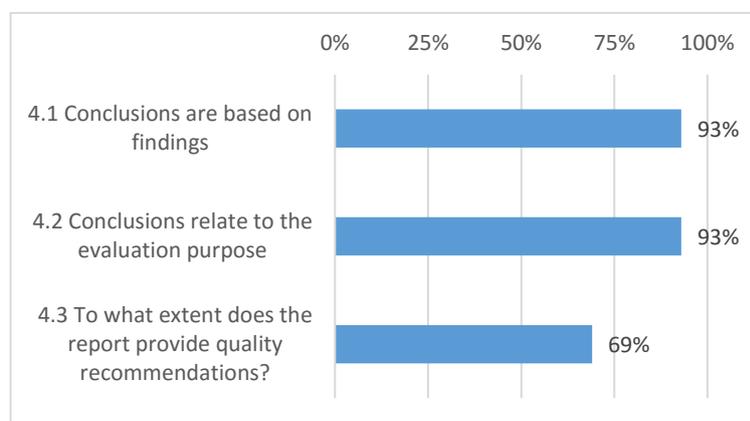


Figure 14: Percentage of evaluation reports fulfilling the specific assessment criteria for the quality of conclusions and recommendations

Almost all reports (97%) of the reports stated recommendations and in most reports (87%) it was obvious that recommendations were linked to the findings and conclusions and the number of recommendations was adequate (73%). Some reports provided only very few and general recommendations, or too many and in such detail that it made it very difficult to monitor implementation and it left too much space for the management to decide on which recommendations to follow up. The two indicators looking at the quality of recommendations showed a considerably lower fulfilment. Some 53% of the recommendations were targeted (e.g. they indicated who should take the responsibility for acting on the recommendation), while 47% of the reports stated the recommendations in a general way, normally addressing 'the project' as the only recipient without addressing specific actors in the project or other stakeholders (e.g. WHH HO, CO, donors, partners, etc.). Only 37% of recommendations were considered as SMART, including, for example, a prioritisation, the identification of the implementers of the recommendation or a period for implementation.

3.1.2 Factors influencing 'Accuracy'

With the available data, empirical evidence on the factors influencing accuracy could only be assessed for analysing differences between centralised and decentralised evaluations. All other statements on influencing factors were based on the findings of the mail survey and on statements from the interviews conducted. Apart from the statements on differences between CE and DE, all

⁸ Specific, Measurable, Achievable, Relevant/Responsible, Time-bound.

other statements should be read as hypotheses, founded on stakeholder perceptions and partly backed by empirical evidence from meta-evaluations conducted by other organisations.

As outlined under 3.1.1, there was a distinct difference in accuracy between CE and DE, with the CE performing much better in comparison. There could be several factors contributing to the differences between CE and DE:

- The MELA HO evaluation advisors, with their longstanding experience and evaluation expertise, exert a quality assurance function and ensure that WHH quality standards for evaluations are adhered to. This quality assurance function is either missing in (some of) the CO or the MELA focal points do not have the leverage and / or capacities to exert this function satisfactorily. There was no evidence that CO / MELA focal points make use of the quality checklist that was attached to the annotations to the template for DE ToR. MELA HO does currently not have any defined role in quality assurance for DE or does not fulfil its role.
- MELA focal points with clearly defined tasks and responsibilities are not yet fully functional in all countries, meaning there is no organised management of the evaluation process, but each country / project acts on its own account with varying results.
- WHH uses different standard / recommended ToR for DE and CE, with the ToR for CE being more detailed. It seems likely that the detailed orientation gives evaluators a clearer idea of the WHH quality standards for evaluations, while the DE recommended ToR are not substantial enough to provide sufficient guidance for quality. Both ToR do not clearly state WHH expectations with regard to the description of the methodology, which is quite likely the reason why methodology scoring was particularly low.

The categorisation into CE and DE does not give any hint on the nationality of the evaluators. While all CE used either international consultants or mixed teams, DE made use of national consultants, international consultants and mixed teams. Although there was no clear difference between the performance of national and international consultants, mixed teams clearly performed better. This corresponds with observations of some of the interviewees who felt that mixed teams provided an ideal combination of international experience, methodological expertise and a thorough understanding of the local context. This in turn corresponds with the findings of a Gesellschaft für Internationale Zusammenarbeit (GIZ) meta-evaluation (2016, page 20f.) that found a significant correlation between the number of evaluators on the team and the quality of reports. The same study pointed to a significant positive correlation between the numbers of working days the evaluators had and the report quality with a specific influence of the number of days provided for reporting. More days for reporting particularly influenced the reflection on the methodology, so it seems that an adequate time budget for reporting promotes the discussion of the evaluation data base and the methods for data collection and analysis.

3.1.3 Assessment and conclusions on 'Accuracy'

General report features: This category had the best performance rate compared to the other three assessment categories. A particular strength of the reports was their clear structure, mainly following the outlined standard structure in the ToR. CE overall rated better than DE.

Almost one third of the reports' summaries were not suitable for publication. This raises the question of how WHH will deal with the transparency issue in the future. WHH has repeatedly been praised for allowing the interested public access to evaluation results (of centralised evaluations) by publishing the executive summaries on its homepage. An evaluation system's assessment commissioned by BMZ in 2008 suggested to extend public access to the full reports; a discussion currently ongoing in the MELA team. While it seems important to maintain transparency and continue publishing executive summaries, the evaluator does not see how the lower quality

summaries, much less the 50% of less satisfying full reports, could be published without compromising WHH's professional reputation. On the other hand, selective publishing, distinguishing suitable from non-suitable summaries, does not seem a recommendable long-term option to handle transparency of evaluation results.

The length of reports and inclusion of gender issues were strengths, with most reports having a readability-friendly length. However, the considerably shorter reports tended to leave out relevant parts of the analysis, while the longer reports went into too much detail, with many reports repeating the same information in different sections. This made them difficult and tiresome to read. With regard to gender, the evaluator observed that it was often the gender specific or gender disaggregated logical framework indicators that triggered the observations on gender in the reports.

Methodology: This category had the lowest performance rate compared to the other three assessment criteria with the vast majority of reports presenting the methodology in a rather unsatisfying or unsatisfying way.

Interestingly, it was the DE that performed better than the CE.

As the methodology is to ensure the validity and reliability of findings and the evaluator assumes that the methodology applied in most evaluations was sound and of good quality, it is the presentation that should be improved.

DAC criteria analysis: Only half of the reports fulfilled (fully or mainly) the quality criteria for the analysis of the DAC criteria, even though these formed the structuring core elements of almost all evaluations. While the assessment of effectiveness and relevance was strong, evaluation reports struggled with the analysis of sustainability and particularly efficiency.

Quality of recommendations: Also, only slightly more than half of the reports fulfilled (fully or mainly) the quality criteria on the quality of recommendations. This was due to a tendency to formulate recommendations rather vaguely without targeting and prioritising them. As a result, quite a number of stakeholders had some reservations with regard to the usefulness and feasibility of recommendations (see 3.3.1).

Overall conclusions on accuracy: The fact that only half of the reports were rated as fulfilling the accuracy standard to a good or satisfying extent raises the question of how to improve report quality to maintain WHH's good reputation among donors and evaluation professionals regarding the implementation of quality evaluations.

The assessment results revealed how important currently the role of MELA HO is for the quality assurance of evaluation reports. There was a distinct difference between the quality of the CE and the quality of the DE. MELA HO does not have the staff resources to provide the same quality assurance for the decentralised evaluations. It should therefore become a priority for WHH to develop decentralised capacities for evaluation management in order to capacitate the CO to monitor and increase report quality.

3.2 Evaluation standard process quality ('Feasibility and Fairness')

The evaluation standards 'Feasibility' and 'Fairness' assess the quality of the evaluation process by reflecting on the extent to which evaluations are realistic, well designed, diplomatic and cost-efficient (feasibility) and if they are conducted in a way that ensures a fair and respectful interaction with the involved stakeholders (fairness).

The following sub-chapters provide a description and analysis of the findings along the criteria and discuss possible influencing factors and conclusions. Findings were based on the perceptions of HoP, MELA focal points and CD involved in the 2016 evaluations. Perceptions were further explored through a mail survey and telephone / skype interviews.

3.2.1 Description and analysis of process quality

The evaluation standards Feasibility and Fairness were assessed using 15 indicators that covered a range of topics (e.g. the competences and performance of the evaluators), stakeholder participation in the evaluation process, the timeliness of the process, access to findings and the quality of the support provided.

Overall, the satisfaction with the process quality was very high: 83% of the respondents declared that the evaluation was conducted in a way that encouraged the acceptance and the utilisation of the evaluation findings. It was felt that the evaluations were inclusive, provided useful findings to the project and addressed the important issues.

With regard to the perception of the evaluators, these were generally perceived as fair and unbiased (89% fully or rather agreed). It was felt that evaluators are open to discuss, include the different stakeholders and look at all aspects of the project without prejudices and preconceived opinions. Only a few respondents complained about fixed opinions and evaluators with a generally negative attitude towards the project.

Evaluators were perceived as competent, technically as well as methodologically (89% fully or rather agreed) with solid knowledge of the respective sector and a good command of the evaluation methods they applied. Criticism was limited to a few respondents who felt that the evaluators lacked technical qualification, were unable to relate to the country-specific context or generalised individual opinions.

Respondents also felt that the evaluators showed a good understanding of the project management reality (86%). The identification of project management key issues and the provision of options considering the potentials and limitations of the project management and the context provided a sound basis for developing conclusions and recommendations.

Compared to the other criteria, this criterion scored slightly lower, with some people stating that the evaluators were not able to understand and take into account the challenges faced by management.

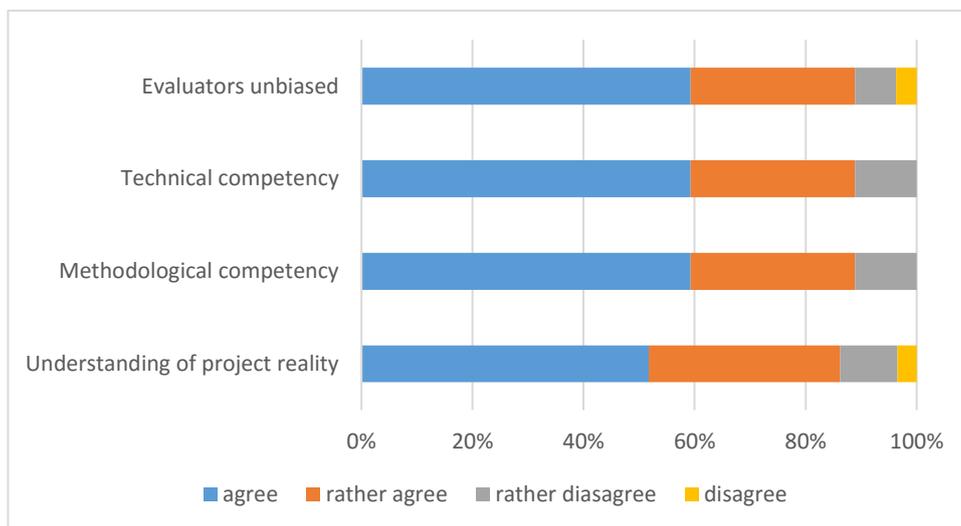


Figure 15: Perception of the evaluators' performance

The participation of relevant stakeholders in the preparation of the evaluation and the entire evaluation process was perceived as high. Particularly high was the consent that all the relevant stakeholders fully participated during the evaluation process, as different stakeholder groups were consulted and in many cases key stakeholders attended the debriefing meeting to discuss the findings. The perception of stakeholder involvement in the preparatory phase was slightly lower. In most cases, stakeholders were already involved in the planning phase of the evaluation, having a

chance to contribute to the ToR and thus shaping the scope and direction of the evaluation. In other cases, stakeholders were rather involved in the implementation of the evaluation and not in the preparation, as this turned out to be too time consuming and complex.

The majority of respondents thought that the resources allocated to the evaluation were adequate. It was felt, particularly during the evaluation process, everything needed (staff, transport) was provided for, so that the evaluation could be implemented smoothly. Reservations mainly related to an insufficient budget that did not allow for contracting highly qualified experts and that limited the time that could be dedicated to the evaluation. Insufficient time was mentioned as the main resource limitation.

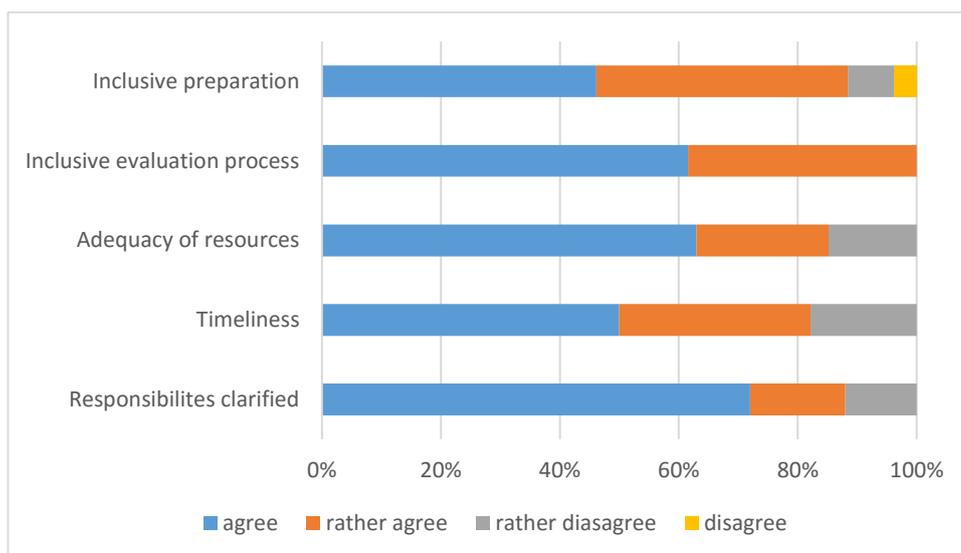


Figure 16: Perception of other fairness and feasibility criteria

The lowest scoring in this criteria group is related to the timeliness of the evaluation. While half of the respondents thought that the evaluation was conducted at the right time to serve its purpose, 32% had some reservations on the timeliness and 18% rather disagreed with regard to the timeliness.

Depending on the evaluation type, the evaluations were on time to give useful recommendations for the project management and how to better achieve the project objectives (mid-term evaluations) or on time to provide guidance for drafting a follow-up project or provide the donor with information on the project outcomes (final evaluation). While some of the delays were due to circumstances that could not be influenced (e.g. security issues, fuel crisis), other delays seemed to be due to a lack of strategic planning of when best to conduct the evaluation to serve its purpose. In some cases, the evaluations were too late to give orientation to the project management or to provide guidance for the planning of the next phase.

The vast majority of respondents considered the ToR and the contract with the evaluators as clear in stating the responsibilities and obligations of the parties involved in the evaluation process. From their point of view, all involved knew what was to be done, how, by whom and when. Only in a few cases, it was felt that ToR and contracts were not sufficiently clear to define the roles of the different actors.

The question on the ToR triggered some more general comments on the standard ToR, which some respondents perceived as too complex and overloaded to provide clear orientations.

An open question in the mail survey explored the extent to which all stakeholders had access to the evaluation findings. Projects usually ensured stakeholder access to evaluation findings by disseminating the report within the project, to partners and to donors and by inviting stakeholders

to the evaluation debriefing session and/or to a workshop to discuss the findings and the operationalisation of the recommendations. In some interviews, it was mentioned that the dissemination of findings is not yet satisfactory in all projects and that at times findings are not sufficiently shared. For CE, sector advisors and other interested HO staff were invited to the HO debriefing. In principle, all evaluation reports should be uploaded in O365 to provide overall organisational access to the findings, but not all projects did that and the way reports were uploaded, without clear headings on project and sector, did not facilitate access. The 'interested public' as another stakeholder currently has only access to the summaries of the CE as these are uploaded on the WHH website.

As WHH was interested in understanding to what extent its advisory services, formats, templates and materials are regarded as helpful for ensuring a good quality evaluation process, two questions in the mail survey asked for the perception of the support.

The majority of the respondents are satisfied with the support provided, but there is a distinct difference between satisfaction with the contentment regarding the advisory services provided and the available support material.

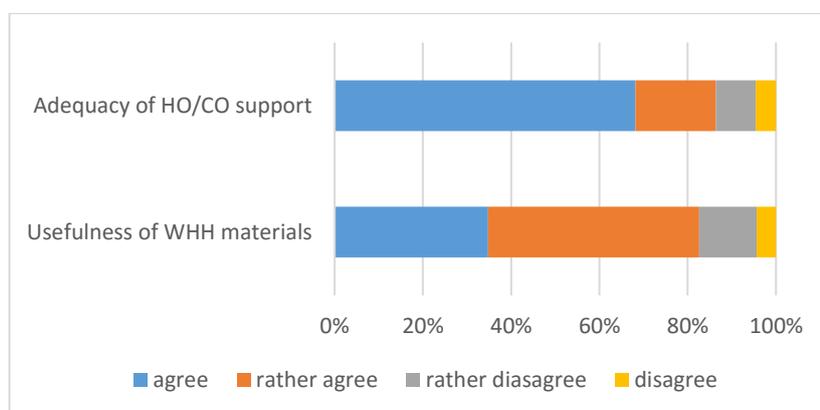


Figure 17: Perception of adequacy of HO/CO support to the evaluation process

Most respondents experienced support from HO and CO as excellent and helpful for the evaluation process. They underlined the high commitment of HO evaluation advisors and their availability for providing support on request. The degree to which HO or CO provided support depended on the type of evaluation (DE or CE). Only few respondents felt they would have needed more support.

The satisfaction with the materials provided is considerably lower. WHH support material in many comments was perceived as not being relevant, difficult to adapt to and difficult to obtain. Some respondents were not even aware of the existence of support materials.

One interviewee noted: 'The ToR for evaluations are too complicated. One third of the wording could not be easily understood.'

In comparison to the survey results, the interviews conducted with HoP, MELA focal points and CD presented a more critical view on the quality of the evaluation process. Several interviewees felt that HoP and CD did not have the capacities/experience to fully grasp what evaluations were about and how the process was to be managed. As a result, a lot of the potentials of evaluations could not be realised.

3.2.2 Factors influencing process quality

The presentation of the factors influencing process quality is based on stakeholder perceptions expressed through the mail survey and the telephone / skype interviews. From 42 comments on the factors influencing the process quality:

- 16 mentioned the choice of a competent and experienced evaluator

- 9 mentioned the proper planning of time and budget
- 7 mentioned the participation of all relevant stakeholders in the evaluation process
- 5 mentioned clear ToR with a precise definition of the evaluation objectives and a clarification of the intended users
- 4 mentioned the importance of choosing the right methodology
- 3 mentioned the thorough preparation of the evaluation process, including the coordination of different stakeholders, the discussion of evaluation questions and the communication with the evaluators to clarify the focus of the evaluation
- 3 mentioned the timing of the evaluation, taking into account seasonality, availability of the target group and the right time in the project cycle

There were no indications that there was a perceived difference between the process quality of CE and DE. It also seems that there was no correlation between the accuracy standard / reporting quality and the perception of the process quality. A good quality of reporting had no relation to the process quality being perceived as good. This confirmed findings from a GIZ meta-evaluation study (2016, page 31).

3.2.3 Assessment and conclusions on process quality

The degree of satisfaction of HoP, MELA focal points and CD with the process quality of the evaluations was high, with evaluations generally being perceived as fair and feasible.

As respondents repeatedly underlined the importance of the technical and methodological qualification of the evaluators and the identification of good and experienced evaluators as a key challenge for DE, the question on how to support the identification and selection of qualified evaluators might deserve some attention.

Resource allocation questions underlined the importance of allocating sufficient funds for the evaluation already in the planning phase and the need to provide enough time for preparation, field work and reporting. Time has to be considered as a key resource when contracting the evaluators, but also when considering the time resources the project itself needs to allocate for preparation, implementation and follow-up of the evaluation. The timeliness of an evaluation to be useful either for the management of the project and the result achievement or for lessons learnt for a next phase could be further improved.

The influence of clear ToR on the process quality has been repeatedly underlined.

Stakeholder access to evaluation findings seems to be rather well organised at project level, with open questions on how to secure access to findings at HO/organisational level and for the public for DE.

The appreciation of the HO advisory services could be interpreted as an indicator that direct bilateral support and support through relatively direct contacts is preferred to written support material and there is potential for the future in extending advisory services, probably through new formats. The critical remarks on the utility of support materials indicate that more practical and hands-on support materials are required.

3.3 Evaluation standard ‘Utility’

The evaluation standard is of particular importance as it looks at the question to which extent evaluations successfully achieve the evaluation objectives and in how far they satisfy the information needs of the intended users. The main question is, to what extent evaluations are useful or fail to be useful.

The sub-chapters follow the same structure as the sub-chapters of the other evaluation standards and use the same methodology to generate findings as applied for the process quality standard.

3.3.1 Description and analysis of 'Utility'

The evaluation standard utility was assessed using five indicators exploring perceptions on the relevance of the evaluation with regard to i) the evaluation purpose, ii) the usefulness of the recommendations for the project, iii) the feasibility of recommendations, iv) the usefulness of recommendations for the organisation (e.g. for similar projects and for strategic decision making), and v) the utilisation of project evaluation recommendations at higher levels of the organisation.

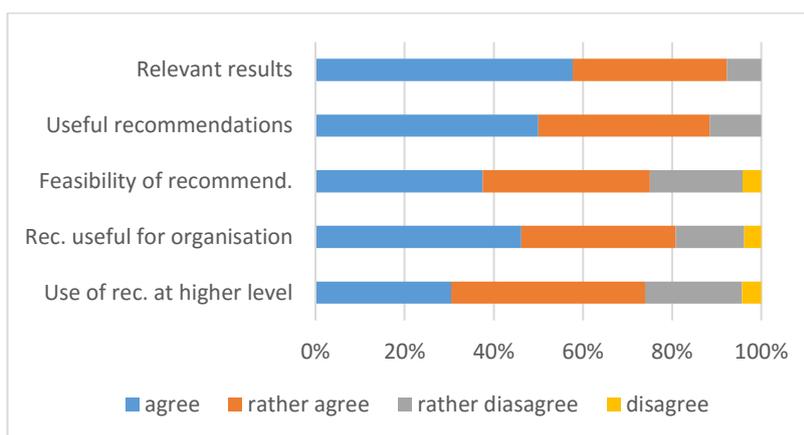


Figure 18: Perception of evaluation utility

For utility, the general satisfaction was also high with 93% of the respondents fully (58%) or rather (35%) agreeing that the evaluation has been relevant to the evaluation's purpose. Mid-term reviews were perceived as particularly useful for improving the effectiveness of the project and take corrective actions for achieving better results, while final evaluations were felt to be useful to guide the planning for a successor phase, to use findings for country programming and to fulfil the donor requirements on accountability. Only in a few cases, respondents felt that results were not relevant enough for improving the effectiveness of the project or that the evaluators should have taken a more critical view on the project, pointing out the weak spots that could be improved.

The recommendations were mainly perceived as being useful for the project. Most respondents stated that they were using the recommendations for the management and implementation of the project or for the planning of a successor project. Some respondents remarked that some recommendations were not clear, too general and not sufficiently targeted. Some complained about too many recommendations.

One of the lowest scores in the criteria group on utility was the feasibility of recommendations with a quarter of the respondents assessing recommendations as not being feasible or rather not being feasible and only 38% fully agreeing with the feasibility of the recommendations. While some respondents felt that recommendations were feasible, others thought them to be, at least partly, unrealistic and impossible to implement, reducing the usefulness of the evaluation to a certain extent.

Quite a high number of the respondents (81%) felt that the recommendations could be potentially useful beyond the project, for WHH as an organisation (e.g. for the planning of similar project) for the country strategy or for strategic decision making at HO level. Some respondents felt that the recommendations were mainly context and project specific and would not be of much use at other levels.

However, respondents were particularly sceptical with regard to the actual utilisation of recommendations at higher levels as they felt they rarely witnessed any strategic application within

the organisation. Issues mentioned in this context were a felt limited interest in HO sector units to discuss results and take them to higher levels, insufficient leverage of MELA to push developments, and a general lack of mechanisms/processes to identify, process, present, discuss and react on strategic recommendations from project evaluations. Furthermore, many evaluations, particularly DE, tended to focus on project specific recommendations and hardly considered the strategic level in analysis and recommendations. WHH management level felt that the strategic insights from project evaluations currently are insufficient. Some interviewees remarked that on the other hand management levels did not clearly define their information needs at strategic level and therefore expectations remained too vague to be addressed and project evaluations lacked a clear orientation on what strategic topics they should look at.

In general, interview partners felt that the learning from evaluations was still the weakest part of the evaluation process and that there was considerable potential to further increase the learning from evaluations and thus the utility of evaluations. While discussions of evaluation recommendations with staff and partners at project and country levels were already common, fewer projects used management response sheets to strategically discuss and agree on the follow-up on the implementation of recommendations; and less than a quarter of the projects confirmed using the management response sheet after some time to follow up the extent to which recommendations were being implemented.

3.3.2 Factors influencing 'Utility'

The factors influencing utility are basically the same as the factors influencing the process quality described in chapter 3.2.2.

In addition, it is the fulfilment of the above-mentioned quality criteria, i.e. the perceived usefulness and feasibility of recommendations that influences utility. The extent to which the projects/ countries reflect and follow-up the evaluation recommendations seems to be another major influencing factor.

3.3.3 Assessment and conclusions on 'Utility'

The degree of satisfaction of HoP, MELA focal points and CD with the utility of the evaluations was high, with evaluations generally being appreciated as useful for the projects.

The follow-up of evaluation results and recommendations and the learning from evaluations were identified as processes with considerable scope for improvement. A reflection on how WHH could strengthen these processes is required.

In addition, there is a perceived gap between the reality of DE that is characterised by a stronger focus on project specific recommendations and a perceived lack of mechanisms and processes ensuring the feeding up of strategic findings and recommendations within the organisation; and conversely, the expectations regarding the generation of strategic recommendations for HO management level. It would need further reflection on the realism of these expectations vis-à-vis project evaluations. Either additional efforts might be required to place strategic questions more prominently in the evaluation process or (and) strategic requirements should be addressed through strategic and sector evaluations.

3.4 Patterns of reoccurring findings and recommendations⁹

For identifying patterns of reoccurring findings and recommendations, the conclusions and recommendations of all 30 evaluations were clustered according to different topics¹⁰ and analysed with Mayring's content analysis (see 2.2.2).

⁹ The numbers in brackets show how many evaluation reports stated the respective recommendation.

¹⁰ For example, strategic recommendations, recommendations on general management, working with partners, gender, nutrition, WASH, income generating activities and agriculture. Under some topics,

3.4.1 General patterns

The most frequently reoccurring recommendation in the evaluation reports (10) was the recommendation to prolong the project, either by an additional three-year phase or at least by extending the project duration of the ongoing project. This recommendation was based on the conclusion that many intended outcomes could be achieved within the (mostly) three-year period of a project and that sustainable change would have required at least six years.

Other evaluations (4) stressed the need to define clear objectives with a defined strategy for all components. This went hand-in-hand with a need to focus on specific objectives rather than 'spreading out too thin' and trying to cover a wide range of objectives and topics.

Several evaluations (3) mentioned the need to apply a 'do no harm' approach to avoid contributing to conflicts.

3.4.2 General management

3.4.2.1 Quality of the ROM system¹¹

A relatively high number of evaluation reports addressed the quality of the projects' results-oriented monitoring systems (ROM), confirming in general the importance of having a functional M&E system in place, particularly for enabling the projects to report on outcomes.

Recurring recommendations were:

- Strengthen the utilisation of monitoring results for project steering/decision making (3)
- Develop staff capacity for results based M&E (3)
- Strengthen the participation of the target groups in the monitoring process (e.g. by involving them in defining indicators and collecting data) (3)

3.4.2.2 Visibility and communication¹²

Also, a relatively high number of evaluation reports addressed issues around visibility and communication:

- Visibility of the projects and its results were found to be insufficient, especially vis-à-vis donors, other implementers and state institutions. Evaluation reports (5) recommended enhancing visibility and communication efforts to this regard with a focus on results achieved and good practices
- Communication with partners and with other stakeholders should be intensified (4), particularly on project objectives and strategies to develop a joint understanding of the task; communication should be regular and strategic
- Communication to a wider public and communication in advocacy and campaigning (e.g. in nutrition, WASH) should increase the use of modern technologies, e.g. by using mobile equipment for public viewing and text messaging for mobile phones (3)

additional sub-categories were created – e.g. quality of M&E, administration and logistics, staff capacity development under general management.

¹¹ 10 out of 30 reports addressed the quality of the results oriented monitoring system in the conclusions/ recommendations section.

¹² 10 out of 30 reports addressed the topic of visibility / communication in the conclusions / recommendations section.

3.4.2.3 Management of human resources¹³

Findings on human resource management usually addressed staff shortages hampering a satisfying implementation of certain tasks and capacity development issues. Recommendations on the management of human resources included:

- Increasing human resources for specific responsibilities so that the related tasks can be carried out adequately (5). Particularly mentioned were staff shortages for administration, logistics, networking and advocacy, partnership relations, and coordination in the context of humanitarian assistance
- Increasing efforts for capacity building of staff on different topics, based on a needs analysis for capacity development (5)
- Ensure that all staff have a thorough understanding of the project concept, its objectives and strategy (2)

3.4.2.4 Administration, logistics, finance and procurement¹⁴

Most of the findings and recommendations on administration, logistics and finance were project specific. Findings on these topics mostly looked at delays caused by lengthy procurement procedures and partners weaknesses in adhering to WHH requirement in financial accounting and other administration issues. Two recommendations were mentioned twice:

- The time needed for tendering and procurement has to be realistically estimated and taken into account when planning for an activity (2)
- Implementation partners need support to strengthen their administration and financial procedures (2)

3.4.3 Working with local partners¹⁵

Evaluation findings generally confirmed the importance of working with/through local partner organisations as an important strategy to increase sustainability. The following recommendations were given in several reports:

- Partners' capacities should be assessed before starting the project and capacity development for partners should be included in the project strategy (6)
- It would be helpful to agree with partners on clear targets with indicators for their performance and for their capacity development (2)
- Capacity development needs are mostly identified in the areas of organisational development, management and administration and results-oriented monitoring (4)
- Apart from supporting specific partners, WHH should also seek to support networking and cooperation between the different partners (2)

One evaluation mentioned that particular importance should be given to strengthening partners in those countries where WHH is in the process of shifting from direct implementation to implementing through partners.

¹³ 11 out of 30 reports addressed the topic of human resources management in the conclusions / recommendations section.

¹⁴ 8 out of 30 reports addressed the topic of administration, logistics, finance and procurement in the conclusions / recommendations section.

¹⁵ 12 out of 30 reports addressed the topic of working with / through local partners in the conclusions / recommendations section.

For Afghanistan, one evaluation (AFG 1148) suggested to strategically build up new non-governmental organisations (NGOs) and to include security management as an aspect in the assessment of partners' capacities. The same evaluation recommended reflecting on WHH's definition of strategic partnerships and the specific requirements of a strategic partnership in a context like Afghanistan.

3.4.4 Working with government¹⁶

For projects cooperating with government stakeholders, some evaluation report findings stated the importance of government involvement and buy-in for improving sustainability perspectives. They came up with the following recommendations:

- Projects should seek to involve government officials more strongly in trainings and activities (on-the-job capacity development) with the aim of improving government service delivery (3)
- Projects should increase their advocacy activities on project topics aiming at increased ownership/buy-in of government institutions (3)
- Linking beneficiaries to the respective government institutions for specific topics/services (3)

Two evaluations underlined the importance of identifying the 'right' institutions; that is, all those with a stake in the intervention and thus emphasising the need for an actors' analysis.

3.4.5 Other topics

Findings and recommendations were too diverse and project specific or too few to identify patterns for the following categories:

- Working with communities
- Agriculture and livestock
- Value chain development
- WASH
- Gender

3.4.6 Humanitarian assistance

While identifying patterns across projects from all sectors provided few results (with a relatively high methodological effort), a comparison of projects from the same sector came up with some topics that might be worth to consider at a strategic level.

This was the case for the projects in the framework of humanitarian assistance, and in particular for several evaluation reports from Syria and Sierra Leone.¹⁷ These reports raised questions with implications for the WHH strategy on humanitarian assistance and – though coming up with different recommendations – addressed similar concerns: the strategic positioning of WHH as an

¹⁶ 9 out of 30 reports addressed the topic of working with government in the conclusions / recommendations section.

¹⁷ SYR1022: Lebensmittel- und Winterhilfe für syrische Bürgerkriegsopfer in Nordsyrien und der Türkei, SYR1024: Improving the livelihood for Syrian internally displaced persons (IDPs) and community members through vocational training measures and support of the medical infrastructure, SLE1033: Disaster preparedness and reactivation of agricultural production and trade in Ebola affected districts of Southern and Eastern Sierra Leone (DiPAT).

organisation with a strong engagement in fragile contexts; the question how best to act and deliver results in those contexts; and the question which support needs to be provided by HO.

Particularly the evaluations of SYR1022, SYR 1024 and SLE1033 pointed out that the COs would have required above all strategic HO support to improve emergency response delivery. All three evaluations underlined that specific expertise would have been required.

Several specific recommendations proposed strategies to provide COs with the support they require:

- With the increasing use of cash-based-interventions (CBI) (particularly e-vouchers), WHH should build up a pool of CBI experts to be contracted to write proposals and to be employed in projects. (SYR1022)
- The Emergency Response Team (ERT) provided valuable hands-on support in Sierra Leone. However, the CO would also have required more strategic support, for example: coaching of project coordinators, advice on how to coordinate WHH's different response interventions, how to act in response coordination, how to ensure a good visibility in the donor community and how to write humanitarian aid proposals. HO should consider building ERT's capacities to focus more strongly on strategic support. (SLE1033)
- Welthungerhilfe should build expertise for 'remote management' and provide investment and capacities. Experts need to work on risk management systems (RMS) that are tailored to the capacities, potentials and limitations of the organisation. (SYR1024)
- Planning of projects and activities in a crisis context should work with scenario planning to be prepared for context changes. (SYR1024)

3.4.7 Strategic topics

Content analysis showed that centralised evaluations were more likely to provide findings and recommendations beyond project level (in 88% of the reports) in comparison with decentralised evaluations (in 41% of the reports). Interviewees stated that this might be due to the stronger involvement of HO staff in designing the evaluation questions for the centralised evaluations, ensuring that strategic knowledge interests were included. COs and projects managing an evaluation have a stronger interest in specific findings and recommendations.

3.4.8 Conclusions on patterns of reoccurring findings and recommendations

Most findings and recommendations in the project evaluation reports were project specific. Owing to the wide variety of different sectors covered by the evaluations, the number of sector specific recommendations that could form specific patterns was comparatively low.

Overall, the findings from the content analysis provided some patterns of reoccurring findings and recommendations. However, these seem to be so general in nature that they express more of a 'joint common sense' or 'ABC' of good development cooperation practices than providing useful insights and guidance at a strategic level.

A reason for the overall low number of recommendations with strategic importance could be that evaluation questions focused on project specific questions rather than including questions of strategic interest. With an increasing number of decentralised evaluations, it can be expected that strategic questions in evaluation reports will be covered even less, as they are mainly of interest to HO.

Interpretation of the patterns posed several challenges. It is unclear whether a few recommendations on a certain topic indicate that in most projects the findings did not reveal any

shortcomings (and therefore no recommendations were required) or if the evaluators did not include the topic in their assessment.

The findings regarding the humanitarian assistance projects indicate that rather than looking overall at the identification of general patterns, it could be more effective to look at specific sectors. This would require strategic evaluations, covering specific sectors or topics of specific strategic interest to WHH or the continuation of mainstreaming 'annual evaluation topics' reflecting strategic knowledge interests of HO as obligatory for all project evaluations.

The identification of reoccurring findings and recommendations in the context of a meta-evaluation seemed not well suited for satisfying information requirements at a strategic level.

4 Overall conclusions

The meta-evaluation/analysis has shown that the WHH actors in the outer structure have an overall positive perception of the quality of the evaluation processes and the utility of evaluations. Thus, the degree of fulfilment with regard to the evaluation standards of fairness, feasibility and utility can be assessed as 'good', with some elements still having the potential to improve even more; for example, the inclusion of stakeholders in the preparation of the evaluation, the quality of recommendations and the follow-up of and learning from evaluations.

Less satisfactory was the fulfilment of the accuracy standards, with half of the reports (rather) not satisfying the quality standards for evaluation reports. This result can be regarded as an outcome of decentralising the responsibility for project evaluations to the CO, while not all of the CO are currently able to manage evaluations in a way that ensures that quality standards are met.

The relatively low quality of half of the reports is problematic mainly with regard to two aspects: the professional reputation of WHH vis-à-vis its donors and the interested public; and the shortcomings in the validity of findings that to some extent put the evaluation results in question.

Another challenge for WHH is the identification and utilisation of findings and recommendations requiring attention at a strategic level of the organisation. By decentralising evaluations, there is a tendency for evaluations to focus entirely on project specific evaluation questions. As mechanisms/processes to identify and communicate findings with strategic importance from DE to higher levels are missing, HO structures (management and sector advisors) feel that it is increasingly difficult to strategically learn from project evaluations. The meta-evaluation has proved to be of very limited utility for identifying strategic findings and recommendations that could be of value for strategic decision making. Currently, it is the strategic evaluations and the CE that provide strategic insights.

The meta-evaluation confirms the need to reorganise the responsibilities and tasks of the evaluation advisors in the MELA team, in conformity with the draft of the recently developed MELA framework.

Until recently, the main focus of the evaluation advisors was the management of the strategic and centralised evaluations. Advisory services were, and still are, rather ad hoc and on demand and evaluation advisors strongly feel that (time) capacities do not allow for intensifying support on top of their current tasks. There is a need to either enhance staff resources and/or prioritise capacity development and quality assurance activities, thus focusing on evaluation management processes in the CO. With the same resources at hand, that would mean reducing the number of evaluations managed by HO.

The monitoring and quality assurance of DE management processes in the CO is still not strongly developed and it is rather unclear to what extent quality assurance for DE is perceived as a mandate of the evaluation advisors in MELA HO.

As the meta-evaluation has pointed out, there is currently a problem with the evaluation report quality and the application of key evaluation tools (e.g. the management response) is lagging behind. The process of shifting evaluation responsibilities to CO has been going on for several years already, while the process of appointing MELA focal points in the CO is still in its beginning. Resulting from these findings and developments, the focus of the evaluation advisory activities should be shifted from managing evaluations to the capacity development of the CO to enable these to successfully manage evaluations. In addition, the evaluator is of the opinion that the assurance of quality standards and the monitoring of CO evaluation management performance should also feature more clearly in the activity portfolio of the evaluation advisors.

The evaluator noticed a certain gap in the responsibility for developing the MELA focal point approach. It seems that currently, MELA is still missing a clear mandate to go ahead with the

development of a MELA community of practice with focal points in every CO. However, this meta-evaluation has shown the need for expertise and responsibility of evaluation management processes in the CO, if evaluation quality standards are to be adhered to.

From the evaluator's point of view, there are several topics in need of more 'ownership':

- The responsibility for establishing MELA focal points and develop their capacities (one specific advisor in MELA?)
- The responsibility for including evaluations and their budgets in project planning and proposal documents (CD? RD? Donor Unit?)
- A clear definition and communication of mandatory evaluation requirements (Evaluation advisors? Management level? RD?)
- The responsibility for quality assurance of evaluations (Evaluation advisors? MELA focal points?)

The MELA framework is pointing in the right direction and quite a number of recommendations to be stated in the next chapter are already included in the framework with some activities already on its way. Still, they shall be mentioned, to point out vis-à-vis all stakeholders their relevance for achieving the objective of continuously developing and improving evaluation processes and thus fully utilise evaluations as key tools for learning. Some updates of the MELA framework would be required, e.g. on the role of MELA evaluation advisors in monitoring and quality assurance and on the meta-evaluation procedure.

All stakeholders ought to be aware that particularly the capacity development process is a long-term process that will require several years to render convincing results.

5 Recommendations

Recommendations are based on the findings and conclusions presented above. The evaluator sought to come up with specific recommendations, stating the actors to which the recommendations are targeted to, the priority level for implementation (priority levels 1–3, with 3 being the highest priority level) and a tentative timeframe. Recommendations for the MELA evaluation advisors are grouped according to the three areas of activity where the evaluator thinks priority action is required: capacity development for CO, quality assurance of DE management processes and the coordination of the establishment of processes and mechanisms to generate strategic findings and recommendations from DE and to communicate findings of strategic interest within HO. The recommendations to the other stakeholders are meant to provide a favourable framework and structures for a successful implementation of evaluation management processes and to clarify the responsibilities of the different actors.

Many of the activities are related to costs, that is, the allocation and availability of funding is a pre-condition for implementation. It was not possible to provide only no-cost recommendations. Some recommendation key activities do not necessarily have to be implemented by WHH staff, but could be outsourced to consultants or to interns and student assistants. However, time for the support and supervision of external support actors will still be required.

Recommendation 1: The MELA evaluation advisors should more strategically address the comprehensive capacity building of the MELA focal points on evaluation management (to MELA evaluation advisors, priority level 3).

This would include the following key activities:

- Design a capacity development strategy on evaluation management for MELA focal points, including elements like i) support material update/development, ii) follow through with updating the evaluation information on the O365 website, iii) provide regional evaluation management trainings for all MELA focal points (design until December 2017)
- Update support materials, including the finalisation of evaluation standards and requirements, bring together the different (partly outdated) materials into one comprehensive document on evaluation management guidelines. Make it a very practical and hands-on document with step-by-step explanations, checklists, templates, one standard ToR for all project evaluations, and short good practice examples. A short report (two pages max.) should inform the CD on evaluation requirements and their supervision duties. Use the same materials for the website (until July 2018)
- Follow through with updating the evaluation information on the O365 website, including an updated data base of consultants with the possibility of MELA focal points also uploading details on national evaluators and providing short feedbacks about experiences with the evaluators. The MELA evaluation advisors should also seek to increasingly use individual support requests (with the consent of the requesting party) as examples and post the discussion on the website instead of dealing with it on an individual basis (until July 2018)
- Provide regional evaluation management trainings for all MELA focal points, presenting and working with the support materials (August 2018 – March 2019)

It is not recommended to focus on specific countries for capacity development. A complete roll-out with less intensive advisory services seems more suited.

Recommendation 2: The MELA evaluation advisors should assume the responsibility for the (coordination of) quality assurance of evaluation management processes (to MELA evaluation advisors, priority level 3).

This would include the following key activities:

- Define evaluation report quality criteria which are key for WHH (probably the checklist developed for the meta-evaluation could be used / adapted) (until December 2017)
- Clarify quality assurance processes and responsibilities (until December 2017)
- Carry out quality checks of evaluation reports and provide feedback to CO (from January 2018)¹⁸
- Monitor evaluation coverage and quality development of evaluation reports (from January 2018)¹⁹

Recommendation 3: The MELA evaluation advisors should take the lead in the coordination of the establishment of processes and mechanisms to generate strategic findings and recommendations from DE and strategic evaluations and to communicate findings of strategic interest within HO (to MELA evaluation advisors, priority level 2).

This would include the following key activities:

- Facilitate a discussion on feasible mechanisms/processes to generate strategic findings and recommendations from DE and to communicate findings of strategic interest within HO (until March 2018)
- Test agreed mechanisms/processes and revise based on results (April to December 2018)
- Establish and consolidate processes for generating strategic findings and recommendations from DE and for communications the findings within HO

During the meta-evaluation interviews, stakeholders already provided a number of ideas. For example, it became quite clear that specific strategic questions are needed, so the management level needs to clarify their 'Erkenntnisinteresse' (strategic information needs). MELA evaluation advisors could facilitate this process, or strategic questions could be derived from WHH strategic objectives. Similar to the 'annual evaluation theme' that was used for CE in the past and that provided specific evaluation questions on a topic of strategic interest to WHH which had to be covered by each CE conducted in a certain year, strategic questions could be formulated and introduced as mandatory evaluation questions for all project evaluations for a certain year.

Similarly, sector advisors could formulate a few strategic questions for their sector and – after receiving the annual evaluation plans from the CO – MELA evaluation advisors could link sector advisors with sector projects to encourage the inclusion of the sector advisors with their specific questions into the evaluation process.

For communicating DE results of strategic interest within HO, different formats could be tested, from providing lists of planned/conducted evaluations with minimum information on the sectors covered to providing three-page summaries on DE.

Recommendation 4: To free resources for the key activities under recommendations 1–3, the MELA evaluation advisors should seek to reduce the number of strategic evaluations and CE to a minimum until the above-mentioned processes have been consolidated (to MELA evaluation advisors, priority level 3, January 2018 – December 2019).

¹⁸ This task could be outsourced to 1-3 consultants with a long-standing WHH evaluation experience and could be linked to other services related to recommendation 3 on the communication of findings within HO. Consultants could assess the quality of the report using a quality assessment checklist, write a German or English 3-page summary for the management level and the public and indicate to whom in HO an alert on the evaluation should be provided.

¹⁹ Data compilation could be carried out by an intern / a student assistant.

Recommendation 5: MELA should assume the responsibility for setting up a network of MELA focal points in all countries and develop them into a functional community of practice (to MELA, priority level 3).

This would include the following key activities:

- Appoint one person within MELA who will take the lead in establishing a MELA community of practice (until October 2017)
- That person would lobby with regional directors (RD) and CD and demand the appointment of MELA focal points for each country. MELA focal points for each country should be appointed and known by December 2017
- Establish rapport with the MELA focal points (October 2017 to March 2018) by contacting all on an individual basis (phone calls)

Recommendation 6: CD/CO should strengthen the strategic use of evaluations at country level and seek to improve the evaluation management processes in their respective country (to CD, priority level 3).

This would include the following key activities:

- Appoint a qualified MELA focal point and provide his/her contact details to MELA HO (until December 2017)
- Communicate evaluation requirements to country staff (until September 2018)
- Make sure that evaluations are already planned for and budgeted during the development of the project proposal (from now onwards)
- Make sure management response sheets are included as a deliverable in the evaluation ToR and follow-up the implementation of recommendations by participating in one discussion on the operationalisation of recommendations immediately after the evaluation and participate in a second discussion several months later to follow up on the implementation status.

Recommendation 7: Sector advisors should proactively seek to know which evaluations are planned within their sectors, offer advice for designing evaluation questions and seek to include a few strategic sector evaluation questions into the ToR. They should further seek to have a phone debriefing with the HoP, MELA focal point or CD to get details on the evaluation results (to sector advisors, priority level 3, from now onwards)

Recommendation 8: The Executive Director Programmes should communicate mandatory requirements and standards on evaluation to RD and CD and demand for feedback on compliance (to Executive Director Programmes, priority level 3, right after mandatory requirements have been defined and agreed on).

Suggestion: Limit mandatory requirements to a minimum to increase acceptance and chance for implementation. Mandatory requirements should include the appointing of MELA focal points, mandatory evaluations for projects of three years or more or for projects with a budget above €2.5 million, use and follow-up of management responses, uploading of reports and responses in O365, adherence to standard ToR with a possible inclusion of evaluation questions on an annual topic provided by HO

Recommendation 9: The Board should approve the MELA framework and allocate the resources for its implementation. This will formalise the MELA focal points and provide the necessary leverage for the MELA team to proceed with the decentralisation of MELA responsibilities (to the Board, priority level 3, as soon as possible, latest before December 2017).

The evaluator did not provide any specific recommendations to the MELA focal points as she assumed that these are to execute evaluation management according to the standards that currently exist and will further improve their performance when requirements will have been more clearly defined, capacity development will have taken place and Communities of Practice will have been established. However, reading the findings and conclusions in this report can provide them with numerous practical ideas for improving evaluation management without waiting for outside support or instructions.

Annex 1: Terms of Reference for the Meta-Evaluation and -Analysis

I. Contracting Party

Deutsche Welthungerhilfe e.V. is one of the largest Non-Governmental Organisations in Germany operating in the field of Humanitarian Assistance and Development. It was established in 1962, as the German section of the 'Freedom from Hunger Campaign', one of the world's first initiatives aimed at the eradication of hunger. Welthungerhilfe's work is still dedicated to the following vision: All people have a right to a self-determined life in dignity and justice, free from hunger and poverty. By 2015, Welthungerhilfe and its partner organisations ran 387 international projects in 40 countries with an overall financing volume of 192 Million Euros, comprised of private donations, public national and international funds.

II. Background and Context

1. Introduction and Context

In 2013, Welthungerhilfe's decentralisation process took momentum. This process aims at strengthening and empowering Country Offices to improve programme quality and to intensify relations with national stakeholders, existing donors as well as to develop relations with new donors and partners.

Consequently, M&E responsibilities were also shifted from Welthungerhilfe Head Office (Monitoring, Evaluation, Learning and Accountability Team (MELA)) to the Country Offices, resulting into an increasing number of project evaluations being commissioned and managed in the various project countries. Yet, the MELA team remains to focus on the commissioning and management of strategic evaluations, responding to Welthungerhilfe's Board of Director's/Executive Director Programme's information needs for decision making. However, in a transitional phase (2015–2017) MELA still commissioned project evaluations in exceptional cases. Furthermore, MELA as a part of Welthungerhilfe's support structure for Country Offices is occupied with the technical backstopping for M&E, i.e. project evaluations. This is ensured by (remote) advisory services and information provision.

In 2015, the integration of the former executive department 'Evaluation' into the programme department underlined Welthungerhilfe's aspirations to use evaluations as learning opportunities for programme improvement, while still responding to accountability requirements.

The upcoming meta-evaluation and -analysis is one of the strategic evaluations carried out on a regular basis, meant to fulfil the following purposes:

- Meta-evaluations are an instrument for quality assurance with regard to (reporting) quality of project evaluations carried out either by MELA or by the Country Offices. Furthermore, meta-evaluations build the basis for the subsequent meta-analysis by ensuring good quality reports only to be included into the analysis. Hence, the reliability of findings is increased.
- Meta-analysis provide Welthungerhilfe management with patterns of re-occurring evaluation findings and recommendations that could be relevant beyond the project context and are thus adduced for (strategic) decision making and organisational learning.

2. Description of the Evaluation Subject

The meta-evaluation includes 30 evaluation reports approved in 2016, of which

- 22 evaluations were commissioned and managed by the Country Offices;

- 8 were commissioned and managed by the MELA team.

Regionally the evaluation reports are clustered as follows:

- Africa/Caribbean/MENA: 18
- Asia/South America: 12

The evaluation reports fall into the following sector categories according to the Welthungerhilfe thematic portfolio:

- Water, Sanitation and Hygiene: 4
- Agriculture and Environment: 10
- Humanitarian Aid: 4
- Economic Development: 4
- Nutrition: 1
- Civil Society and Social Empowerment: 7

Out of the 30 evaluation reports

- 18 are reports of final evaluations
- 11 are reports of mid-term evaluations
- 1 is a report of a Real-time evaluation²⁰

Furthermore, the evaluation is extended by 5 management responses handed in by the Country Offices in the aftermath of the MELA-commissioned evaluations.

III. Aim, Objectives and Users of the Evaluation and Learning Exercise

1. Objectives

The overall objective of this evaluation is a learning exercise primarily for the MELA team and MELA focal points in the respective Country Offices to improve the Welthungerhilfe evaluation system, instruments and practices. The specific objectives are:

Welthungerhilfe obtains structured, external feedback on the quality of evaluation reports as well as the quality of the evaluation process. Evaluation findings allow for the improvement of advisory services and information offers provided by the MELA team for MELA focal points and other project staff involved in the commissioning and managing of evaluations. Furthermore, through the provision of 'good evaluation practices' MELA focal points and other project staff can improve their evaluation practices.

Welthungerhilfe is provided with a user-friendly documentation of reoccurring-occurring evaluation findings and recommendations (in the evaluation reports) in order to facilitate organisational learning beyond the project context, as well as (strategic) decision making.

The opportunities and limits of meta-evaluations as part of the Welthungerhilfe evaluation system are reflected and documented. The related evaluation findings enable MELA to scrutinize in how far the existing evaluation system, with its different instruments, serves the information and learning needs of the organisation.

²⁰ Real-time evaluations can be conducted in Humanitarian Assistance projects.

2. Users

The primary users of this evaluation's results will be internal stakeholders. On one hand, these are the MELA team as sub-team of the Sector Strategy, Knowledge and Learning Unit and MELA focal points in the respective Country Offices. On the other hand, the Executive Director Programmes, management personnel of the projects and programme department and generally staff involved in programming are intended users of this evaluation.

The wider distribution of selected results of this evaluation remains to be decided by the MELA team.

IV. Key Questions of the Meta-Evaluation/ -Analysis

1. Key Questions

a. Reporting Quality: Completeness, Accuracy

1. To what extent are the evaluation reports logically structured with clarity and coherence (e.g. background and objectives are presented before findings, and findings are presented before conclusions and recommendations)?
2. To what extent is the **executive summary** a stand-alone-section that presents the main information of the evaluation?
3. To what extent are the DAC criteria covered in the report?
4. Are the log frames, results chains or other documentations of the **theory of change** (incl. indicators) used as a reference document for the assessment of outcomes and impact?
5. To what extent do the evaluation reports represent a transparent description of the **methodology** applied to the evaluation that clearly explains how the evaluation was specifically designed to address the evaluation criteria,²¹ yield answers to the evaluation questions and achieve evaluation purposes?

Specifically, to what extent are the following criteria met:

- The report describes the data collection methods and analysis, the rationale for selecting them, and their limitations.
 - The report explains in how far existing data (i.e. monitoring data) was used.
 - The report describes the data sources, the rationale for their selection, and their limitations.
 - 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.
 - The report describes the sampling frame, rationale for selection, mechanics of selection, numbers selected out of potential subjects, and limitations of the sample.
 - The report presents evidence that adequate measures were taken to ensure data quality, including evidence supporting the reliability and validity of data collection tools (e.g. interview protocols, observation tools, etc.).
6. To what extent do the evaluation reports give a complete description of stakeholder's **participation process** in the evaluation, including the rationale for selecting the particular level and activities for participation?

²¹ OECD-DAC criteria and facultatively other criteria.

7. To what extent do **findings** respond directly to the evaluation criteria and questions detailed in the objectives section of the report and are based on evidence derived from data collection and analysis methods described in the methodology section of the report?

Specifically, to what extent are the following criteria met:

- Gaps and limitations in the data and/or unanticipated findings are reported and discussed.
 - Reasons for accomplishments and failures, especially continuing constraints, were identified as much as possible.
8. To what extent do **conclusions** present reasonable judgments based on findings and substantiated by evidence, and provide insights pertinent to the purpose of the evaluation?
 9. To what extent do **recommendations** clearly identify the target group for each recommendation; clearly state priorities; are actionable and reflect an understanding of the commissioning organization and potential constraints to follow-up?
 10. To what extent do the reports generally document the impartial and unbiased position of the evaluator?
 11. Do the reports comply with the **Welthungerhilfe reporting requirements** as laid down in the ToRs? Do the existing **Welthungerhilfe formats, templates and materials** suffice to ensure good quality evaluation reports? What should be included in future?
 12. Are there major **differences** with regard to the reporting quality between project **evaluations commissioned by the MELA team and Country Offices**?
 13. What are the **main influencing factors** on the reporting quality of evaluation reports?

b. Process Quality: Utility, Feasibility, Fairness²²

1. To what extent were obligations of the formal parties to the evaluations (what is to be done, how, by whom, when) agreed to in writing, so that these parties are obliged to adhere to all conditions of the agreement or to renegotiate it?
2. Did the evaluations ensure that interests of the persons or groups involved in or affected by the evaluation were identified, so that their interests can be clarified and taken into consideration when designing the evaluation?
3. To what extent did the evaluations ensure that the evaluation is guided by both the clarified purposes of the evaluation and the information needs of its intended users?
4. To what extent were the person(s) conducting an evaluation trustworthy as well as methodologically and professionally competent, so that the evaluation findings achieve maximum credibility and acceptance?
5. To what extent were the evaluations initiated and completed in a timely fashion, so that its findings can inform pending decision and improvement processes?
6. Is the time dedicated to the evaluations sufficient for the evaluand?
7. To what extent were the evaluations planned, conducted and reported in ways that encouraged the acceptance and ultimately utilization of the evaluation finding?
8. Did all stakeholders, to the extent possible, have access to the evaluation findings?
9. Are there major **differences** with regard to the process quality between project **evaluations commissioned by the MELA team and Country Offices**?

²² Based on DeGEval-Standards (http://www.degeval.de/images/stories/Publikationen/DeGEval_-_Standards_short.pdf)

10. Do the existing **Welthungerhilfe formats, templates and materials, advisory services** suffice to ensure good quality evaluation processes?

11. What are the **main influencing factors** on the quality of evaluation processes?

c. Meta-Analysis: Patterns of reoccurring-occurring findings and recommendations

12. Which patterns of reoccurring-occurring findings and recommendations could be relevant beyond the project-context and thus, bear a learning-potential for Welthungerhilfe as an organisation and can serve as a basis for (strategic) decision-making?

d. Methodological Reflection: Meta-Evaluation/-Analysis as part of the Welthungerhilfe evaluation system

13. Can the meta-evaluation (conceptualized as is; with the current reporting quality and quantity) generate reliable data on trends and patterns that allow for organisational learning and (strategic) decision-making? If no, which preconditions would be needed to allow for data on trends and patterns?

V. Methodology

1. Evaluation Approach

The evaluation team must adopt a consultative and participatory approach. The evaluation encompasses three parts:

A) Briefing meeting with MELA to further clarify and discuss ToR, which will be followed by an inception report. Existing checklists to assess the reporting quality used in former meta-evaluations will be used and further developed if need be.

B) A desk review of the relevant, evaluation-related documents and templates as well as the review of all evaluation reports in order to mainly answer evaluation questions a), c) and d). An in-depth analysis of selected evaluation reports, individual or group interviews with relevant stakeholders and other methods will be conducted to collect data mainly on the evaluation question b).

C) Debriefing with the MELA team to present preliminary findings and to discuss further presentations and ways of utilization of the findings. Presentation of findings in a learning-oriented, actionable way for Executive Director Programmes and other (preferably) management staff of the Programme Department.

VI. Deliverables and Reporting Deadlines

a. Inception Report, incl. Methodology and Work Plan (max. 8 Pages)

The inception report should set out the planned approach and methodology to meet the above-mentioned objectives and to answer the related questions, as well as a reflection of the limits of the suggested approach and methodology. It should provide a description on how data will be collected, an evaluation matrix, drafts of suggested data collection tools such as questionnaires and interview guidelines as well as a tentative evaluation schedule.

Deadline: 17 May 2017

The inception report needs the approval of the Contracting Party.

b. Draft Evaluation Report

For the structure, refer to final evaluation report (see section f.).

Deadline: 28 June 2017

c. Format for and Documentation of ‘Good Evaluation Practices’

A format to document selected good evaluation practices (pertaining the reporting, as well as the evaluation process), i.e ToR, Inception Report, Methodology, Management Response etc. for MELA focal points has to be developed.

Deadlines: The format suggestion is to be handed in with the inception report; the documentation is to be handed in with the draft report.

d. Concept, Presentation(s) and Documentation of the ‘Learning Event’

Predominantly the findings of the meta-analysis should be presented in a learning-conducive and interactive presentation format for a wider Welthungerhilfe audience, including representatives of the Board of Directors and other relevant actors responsible for programming. A short concept note for the ‘learning event’ including agenda, objectives, methodologies; pdf-file of ca. 2–3 pages, presentations, e.g. in the format of PowerPoint files.

Documentation, esp. of comments, conclusion etc. incl. photos of flipchart-papers etc.; pdf-file

Deadlines: The concept note and presentation/material is to be handed in with the draft evaluation report; the documentation is to be handed in 7 working days after the ‘Learning Event’.

e. Scenarios ‘Meta-Evaluation/-Analysis’ at Welthungerhilfe

Findings and recommendations directed to the evaluation question should be presented in a separate document, possibly presenting different scenarios on how meta-evaluations and -analysis could complement the Welthungerhilfe evaluation system.

Deadline: To be handed in with the draft evaluation report.

f. Final Evaluation Report including an **Executive Summary and other Evaluation Products**

The report (max. 35 pages excluding annexes) should include (but is not limited to) the following, when answering evaluation questions a) to c):

- Executive summary (max. 2–3 pages)
- Evaluation purpose, objectives and scope
- Methodology (reflection and linking to the ToR and possible constraints leading to deviations from the ToR)
- Findings (related to the objectives of the ToR)
- Conclusions
- Recommendations
- Lessons learnt

- Appendices (including ToR, questionnaires, list of interviewees, and bibliography)

Deadline: To be agreed upon the debriefing meeting, latest 14 July 2017.

g. Management Response

The management response outlines the recommendations, clearly stating the targeted organisational entities and stating the priority of recommendations. The management response is an annex to the evaluation report.

Deadline: To be agreed upon in the debriefing meeting.

VII. Language

All documents should be written in English. The report shall be presented according to the above described Terms of Reference and will include practical recommendations. An executive summary of the report is required.

VIII. Confidentiality

All documents and data acquired from documents as well as during interviews and meetings are confidential and to be used solely for the purpose of the evaluation. Interview partners will not be quoted in the report without their prior permission.

The draft and final report as well as all material linked to the evaluation (produced by the evaluator or the organisation itself) is confidential and remains at all times in the property of the contracting parties.

IX. Budget and Travel Arrangements

Offers should include a proposed budget for the complete evaluation. The budget should present daily consultancy fees according to the number of expected working days over the entire period.

It is anticipated that the evaluation will last around 30 working days, see X.2.

There are no field trips foreseen; travel costs will be limited to incurring costs for meetings in Welthungerhilfe Head Office.

X. Planning / Timeframe

1. Key Dates of the Meta-Evaluation/-Analysis

Closing date for applications	19 April 2017
Recruitment of evaluation team	latest 26 April 2017
Briefing with MECLA	Calendar week 18/19, 2017
Submission of inception report	latest 17 May 2017
Submission of draft report	30 June 2017

Debriefing with MELA	03 or 04 July 2017
Conduct Learning Event, Germany	TBA
Submission of the final report:	Calendar Week 28, 2017

The evaluator is requested to immediately inform Welthungerhilfe if serious problems or delays are encountered. Any significant changes to the evaluation timetable must be approved by the contracting parties.

2. Estimated work volume ('working days')

Preparation and reviewing of documents	2 days
Briefing with contracting party, in Bonn, incl. traveling there and back	1 day
Inception report	3 days
Desk review	13 days
Selected interviews	2 days
Documentation of selected 'good evaluation practices'	1day
Methodological reflection, scenario development	0,5 day
Debriefing	1 day
Final report	5 days
Preparation of learning event	0,5 day
Conduct learning event in Bonn, incl. travelling there and back	1 day
Total	30 days

XI. Expertise of the Evaluator

It is foreseen that this evaluation will be carried out by an experienced evaluator.

The evaluator is required to have

- proven experience in evaluating development projects, preferably in the sector of Sustainable Food and Nutrition Security
- experience in multi-methodological evaluation approaches of development projects and preferably meta-evaluations and -analysis
- experience with innovative, learning-oriented documentation of evaluation findings and learnings
- experience with non-governmental organisations, especially Monitoring, Evaluation, Learning and Accountability within NGOs
- experience in facilitating lessons learnt/learning events

- demonstrated cross-cultural skills
- demonstrated very good oral and written communication skills
- proven language skills: fluent in English and French (orally and written)

XII. Requirements for the Expression of Interest

1. Requirements

The offer must include the following documents:

- Description of methodology incl. reference to perceived feasibility of the ToRs
- Work plan
- Financial offer, fees/honorarium per working day. Kindly note that travel costs, accommodation and per diems will be covered by Welthungerhilfe additionally in line with German legislation ('Bundesreisekostenrecht')
- Proof of professional registration and taxation
- CV and references

2. Expression of Interest

Expressions of interest will be accepted by individual consultants as well as from commercial companies, NGOs or academics.

Contact details:

Expressions of interest shall be submitted via email to Welthungerhilfe, contact person: Leila Broich (Monitoring & Evaluation Advisor), leila.broich@welthungerhilfe.de

Bonn, April 4, 2017

Annex 2: List of Evaluations Included in the Meta-Evaluation and - Analysis

List of Decentral Evaluations in 2016

No	Regional Directorate	Project Title	Sector	Type of Eval.
1	RD 1: West Central Africa / Caribbean	PINAF (Projet Intensification de l'Assainissement Familiale dans 12 communes rurales des région des Hauts Bassins et de la Boucle du Mouhoun)	4. Water, Sanitation and Hygiene 4.1 Water Supply	mid-term evaluation
2		BRACED (Building Resilience of 620000 women, children and men in Burkina Faso, to Climate Extremes in improving household food security and incomes)	2. Agriculture and Environment 2.1 Sustainable Agriculture ?	mid-term evaluation
3		Amélioration de la sécurité alimentaire et des moyens de subsistance des populations déplacées du Nigeria et de leurs hôtes dans la région de Diffa, au Niger	1. Humanitarian Aid 1.1 Emergency food aid, emergency nutrition	final evaluation
4		Projet de « Amélioration de la situation alimentaire et prévention du choléra dans les départements de Téra, Gotheye et Bankilaré dans la région de Tillabéri, Niger»	3. Nutrition 3.2 Nutrition linked health interventions	final evaluation
5		Amélioration de la sécurité alimentaire et nutritionnelle dans la Province Ngozi	2. Agriculture and Environment 2.1 Sustainable Agriculture	mid-term evaluation
6		Reintegration and Recovery Programme (RRP IV)	5. Economic Development 5.4 Communal Infrastructure	mid-term evaluation
7		Food Security and sustainable livelihoods improvement of vulnerable communities of conflict and drought affected population in Agig Locality, Red Sea State (RSS)	2. Agriculture and Environment 2.1 Sustainable Agriculture	mid-term evaluation
8	RD 2 : East & Southern Africa/ MENA	Climate Change in Education Programme	6. Civil Society and Social Empowerment	final evaluation

No	Regional Directorate	Project Title	Sector	Type of Eval.
			6.1 Education and Skills Development	
9		Improving the livelihood for Syrian IDPs and community members through vocational training measures and support of the medical infrastructure	6. Civil Society and Social Empowerment 6.1 Education and Skills Development	mid-term evaluation
10		Rehabilitation of basic infrastructure to secure access to water and wastewater as well as renovation of school buildings in northern Iraq	5. Economic Development 5.4 Communal Infrastructure	RTE
11		Sustainable intensification of market based agriculture (SIMBA)	2. Agriculture and Environment 2.1 Sustainable Agriculture	mid-term evaluation
12	RD 3 : Asia /South America	Improved food-security by enhancing local NGOs in four provinces in north/east and central Afghanistan	6. Civil Society and Social Empowerment 6.2 Civil Society Cooperation	final evaluation
13		Rural Development in the provinces Jawzjan, Faryab, Samangan and Nangarhar	2. Agriculture and Environment 2.1 Sustainable Agriculture	final evaluation
14		Strengthening the capacity of organisations for poverty reduction (SCOPE)	6. Civil Society and Social Empowerment 6.2 Civil Society Cooperation	final evaluation
15		Strengthening Farmers' Organisations' Voices in Policy Dialogue on Food Security and Nutrition	6. Civil Society and Social Empowerment 6.2 Civil Society Cooperation	final evaluation
16		Capacity development of networks for food security and social safety nets at the community level in Htan Tabin Township'	6. Civil Society and Social Empowerment 6.2 Civil Society Cooperation	mid-term evaluation
17		Strengthening Community Based Biodiversity Management in the Sacred	2. Agriculture and Environment	mid-term evaluation

No	Regional Directorate	Project Title	Sector	Type of Eval.
		Himalayan Landscape of Nepal	2.4 Natural Resource Management and Environmental Protection	
18		Consolidation of Sewage Disposal Activities in Anbyon and Kujang city	4. Water, Sanitation and Hygiene 4.4 Waste Management	final evaluation
19		Verbesserung der ökologischen Nachhaltigkeit und der Wettbewerbsfähigkeit der Agroforstwirtschaft in drei Wassereinzugsgebieten der Provinz Satipo	2. Agriculture and Environment 2.4 Natural Resource Management and Environmental Protection	final evaluation
20		Project on Sustainable Solution to Food and Nutrition Insecurity in Khyber Pakhtunkhwa	2. Agriculture and Environment 2.1 Sustainable Agriculture	final evaluation
21		Promotion of Food and Nutrition Security in Muzafargarh District	2. Agriculture and Environment 2.1 Sustainable Agriculture	final evaluation
22		Poverty Reduction of Coastal Communities of District Badin through Inclusive and Integrated Development	6. Civil Society and Social Empowerment 6.4 Empowerment of Local Governments	final evaluation

List of Central Evaluations in 2016

No	Regional Directorate	Project Title	Sector	Type of Eval.
23	RD 1: West Central Africa/Caribbean	Sustainable Waste Management in Bo city	4. Water, Sanitation and Hygiene 4.4 Waste Management	mid-term evaluation

No	Regional Directorate	Project Title	Sector	Type of Eval.
24		Disaster preparedness and reactivation of agricultural production and trade in Ebola affected Districts of southern and Eastern Sierra Leone (DiPAT)	1. Humanitarian Aid 1.5 Disaster Risks Management, Disaster Preparedness	final evaluation
25		Rehabilitation Runway Goma Airport	5. Economic Development 5.4 Communal Infrastructure	final evaluation
26	RD 2 : East & Southern Africa/ MENA	Promote and Strengthen Enterprises and Market Systems in Drought-Prone ASAL Areas	5. Economic Development	final evaluation
27		Ernährungssicherung und Katastrophenvorsorge durch Schutz und Verbesserung der Produktions- und Lebensbedingungen	4. Water, Sanitation and Hygiene 4.4 Waste Management	mid-term evaluation
28		Enhancement of Disaster Resilience in Mozambique	1. Humanitarian Aid 1.5 Disaster Risks Management, Disaster Preparedness	final evaluation
29		Lebensmittel- und Winterhilfe für syrische Bürgerkriegsopfer in Nordsyrien und der Türkei	1. Emergency operations 1.3 Basic needs	final Evaluation
30	RD 3 : Asia /South America	Emergency Relief for earthquake victims in Nepal	1. Humanitarian Aid 1.1 Emergency food aid, emergency nutrition	final evaluation

Annex 3: Evaluation Matrix

(based on Welthungerhilfe ToR)

Key areas	Evaluation questions and sub-questions	Assessment criteria / fields of observation (FoO)	Methods of data collection and analysis	Sources of information
	a. Reporting quality			
Accuracy	1.) To what extent are the evaluation reports logically structured with clarity and coherence (e.g. background and objectives are presented before findings and findings are presented before conclusions and recommendations)?	FoO: report structure / table of contents: <ul style="list-style-type: none"> - Background / objectives before findings - Findings before conclusions - Conclusions before recommendations 	Text analysis Use of assessment checklist	Evaluation reports
Accuracy	2.) To what extent is the executive summary a stand-alone-section that presents the main information of the evaluation?	FoO: report executive summary <ul style="list-style-type: none"> - There is an executive summary - The executive summary is a stand-alone section - The executive summary contains the main information of the evaluation 	Text analysis Use of assessment checklist	Evaluation reports
Accuracy	3.) Is there a balance of questions in the ToR that allow for an overall assessment of the project according to the OECD/DAC criteria and the assessment of project-relevant, specific questions?	This question should be removed from the ToR, because the vast majority of evaluation reports does not include the specific ToR of the assignment in the annex. So with the available data, it is not possible to answer the question		
Accuracy	4.) To what extent do the ToR focus on the assessment of (unintended) outcomes and impacts?	This question should be removed from the ToR, because the vast majority of evaluation reports does not include the specific ToR of the assignment in the annex. So with the		

Key areas	Evaluation questions and sub-questions	Assessment criteria / fields of observation (FoO)	Methods of data collection and analysis	Sources of information
		available data, it is not possible to answer the question. One could change it to ‘To what extent does the report focus on the assessment of (unintended) outcomes and impacts’? And then include it under the effectiveness questions (EQ 8)		
Accuracy	5.) Are the log frames, result chains or other documentations of the theory of change (incl. indicators) used as a reference document for the assessment of outcomes and impact?	This question allows two interpretations: FoO: Discussion of the theory of change / results chain / logframe / indicators in terms of quality assessment of the results logic Or FoO: Effectiveness (and possibly impact) are assessed with a clear reference to the logframe and the indicators. Both interpretations will be evaluated	Text analysis Use of assessment checklist	Evaluation reports
Accuracy	6.) To what extent do the evaluation reports represent a transparent description of the methodology applied to the evaluation that clearly explains how the evaluation was specifically designed to address the evaluation criteria, yield answers to the evaluation questions and achieve evaluation purposes?	FoO: Methodology As the question is rather general, it will be answered through the answers to the specific sub-questions. The overall answer will be given by adding / summarising the answers to the sub-questions	Text analysis Use of assessment checklist	Evaluation reports
Accuracy	6a.) The report describes the data collection methods and analysis, the rationale for selecting them	FoO: Methodology <ul style="list-style-type: none"> - The report describes the data collection methods - The report describes why the respective data collection methods have been chosen for the evaluation 	Text analysis Use of assessment checklist	Evaluation reports

Key areas	Evaluation questions and sub-questions	Assessment criteria / fields of observation (FoO)	Methods of data collection and analysis	Sources of information
		<ul style="list-style-type: none"> - It is plausible that the data collection methods have been chosen taking into account the framework conditions (e.g. resources for the evaluation, answering the specific evaluation questions, characteristics of the project or the target group) 		
Accuracy	6b.) The report describes the data sources, the rationale for their selection, and their limitations	FoO: Methodology <ul style="list-style-type: none"> - The report describes the data sources - The report describes the rationale for the selection of the data sources It is recommended to rather assess if the limitations of the chosen methods are discussed (not the limitations of the sources)	Text analysis Use of assessment checklist	Evaluation reports
Accuracy	6c.) The report includes a discussion on how the mix of data sources was used to obtain a diversity of perspectives, ensure data accuracy and overcome data limits	FoO: Methodology <ul style="list-style-type: none"> - A triangulation of methods is applied (using different methods to obtain findings) - A triangulation of data is applied (using different sources, including the views of different stakeholders) 	Text analysis Use of assessment checklist	Evaluation reports
Accuracy	6d.) The report describes the sampling frame, rationale for selection, mechanics of selection, numbers of selected potential subjects, and limitations of the sample	FoO: Methodology <ul style="list-style-type: none"> - The report describes the sample frame - The report describes the rationale for selecting the sample - The report describes the numbers of the selected sample - The report describes the limitations of the sample 	Text analysis Use of assessment checklist	Evaluation reports
Accuracy	6e.) The report presents evidence that adequate measures were	This question is rather vague. To complement the sub-questions on evaluation	Text analysis	Evaluation reports

Key areas	Evaluation questions and sub-questions	Assessment criteria / fields of observation (FoO)	Methods of data collection and analysis	Sources of information
	taken to ensure data quality, including evidence supporting the reliability and validity of data collection tools (e.g. interview protocols, observation tools, etc.)	methodology, the evaluator proposes to use another question: The report explains in how far existing data (e.g. from baselines, partners, the monitoring system) was included into the assessment	Use of assessment checklist	
Accuracy	7) To what extent do the evaluation reports give a complete description of stakeholder's participation process in the evaluation, including the rationale for selecting the particular level and activities for participation?	FoO: Methodology <ul style="list-style-type: none"> - The report explains how the evaluation design addressed the participation of stakeholders in the evaluation process - The evaluation report states at least one reason for the level of participation selected 	Text analysis Use of assessment checklist	Evaluation reports
Accuracy	8.) To what extent do findings respond directly to the evaluation criteria and questions detailed in the objectives section of the report and are based on evidence derived from data collection and analysis methods described in the methodology section of the report? Specifically...	FoO Findings / DAC Evaluation criteria: This question is too broad and very difficult to operationalise. The evaluator proposes to specify the question much more, also in order to be able to cover the different evaluation criteria which form a core part of the assessment: Relevance: <ul style="list-style-type: none"> - In the relevance chapter, the report discusses to what extent the activities and outputs of the programme are consistent with the overall goal It would be good to have a second and third criterion on relevance, but the explanations of what should be covered under relevance are too different in the template explanation of centralised and decentralised evaluations to come up with a strong second criterion that could be valid for both	Text analysis Use of assessment checklist	Evaluation reports

Key areas	Evaluation questions and sub-questions	Assessment criteria / fields of observation (FoO)	Methods of data collection and analysis	Sources of information
		<p>Effectiveness:</p> <ul style="list-style-type: none"> - The evaluation report presents findings on the extent to which objectives are being achieved / have been achieved - The evaluation report refers to the logframe indicators to assess the effectiveness of the project or does he/she explain, why the indicators were not used - The report explains which factors influenced the achievement or non-achievement of the objectives? <p>Efficiency:</p> <ul style="list-style-type: none"> - The report allocates at least for one example costs to outputs - The report discusses if outputs or activities could have been implemented with less resources - The report discusses the advantages and disadvantages of different options for the use of resources or explains why different options cannot be considered <p>Impact</p> <ul style="list-style-type: none"> - The report reflects on the existence of evidence for impact - The report assesses the plausibility of the project contributing to long-term change 		

Key areas	Evaluation questions and sub-questions	Assessment criteria / fields of observation (FoO)	Methods of data collection and analysis	Sources of information
		<ul style="list-style-type: none"> - The report reflects on unintended (positive or negative) outcomes <p>Sustainability:</p> <ul style="list-style-type: none"> - The report reflects on the existence and quality of exit strategies to increase sustainability - The report reflects on the likely challenges for sustainability - The report reflects on the likely degree of sustainability for at least two activities / outputs / outcomes 		
Accuracy	9.) To what extent do conclusions present reasonable judgements based on findings and substantiated by evidence, and provide insights pertinent to the purpose of the evaluation?	<p>FoO: Quality of Triad Findings – Conclusions – Recommendations</p> <ul style="list-style-type: none"> - The report differentiates between findings/analysis, conclusions and recommendations - The report bases at least 80% of its conclusions on findings - At least two of the conclusions relate to the objectives of the evaluation as stated in the chapter on the evaluation purpose 	Text analysis Use of assessment checklist	Evaluation reports
Accuracy / Utility	10.) To what extent do recommendations clearly identify the target group for each recommendation; clearly state priorities; are actionable and reflect an understanding of the commissioning organization and potential constraints to follow up?	<p>FoO: Recommendations</p> <ul style="list-style-type: none"> - The report states recommendations, i.e. it does provide advice on what should be done to improve project performance / the achievement of objectives - There is a clear link of recommendations to findings / conclusions - The recommendations are targeted at different actors 	Text analysis Use of assessment checklist Mail survey Interviews	Evaluation reports CD, MELA focal points, HoP

Key areas	Evaluation questions and sub-questions	Assessment criteria / fields of observation (FoO)	Methods of data collection and analysis	Sources of information
		<ul style="list-style-type: none"> - The recommendations are SMART (i.e. address specific actors, prioritised, realistic) - The number of recommendations is adequate (not too broad and not too detailed, adequate number to be specified with MELA before using the assessment checklist) <p>If recommendations are</p> <ul style="list-style-type: none"> - Feasible - Reflect an understanding of the commissioning organization and its potential constraints to follow up <p>Will be explored through mail survey and interviews</p>		
Fairness	11.) To what extent do the reports generally document the impartial and unbiased position of the evaluator?	<p>Should be removed as this is impossible to judge from the reports. It could, however be rephrased and put under the process quality as:</p> <p>Are the evaluated perceiving the evaluators as impartial and unbiased?</p>	Mail survey Interviews	HoP, to a lesser degree CD and MELA focal points
Accuracy	12a.) Do the reports comply with the Welthungerhilfe reporting requirements as laid down in the (standard) ToRs?	<p>FoO: Overall reporting quality / completeness</p> <p>As the assessment checklist has been developed to a high extent along the standard ToR, the answer to this question will be through adding the results of the specific questions</p> <p>Classification of excellent / good / rather satisfying, etc. reports will have to be discussed with MELA</p>	Text analysis Use of assessment checklist	Evaluation reports
Process quality / Feasibility	12b.) Do the existing Welthungerhilfe formats, templates and materials suffice to ensure	<p>FoO: MELA support quality and requirements</p> <p>This sub-question is almost identical with b10 under process quality. Remove here and</p>	See b10	See b10

Key areas	Evaluation questions and sub-questions	Assessment criteria / fields of observation (FoO)	Methods of data collection and analysis	Sources of information
	good quality evaluation reports? What should be included in the future?	add 'What should be included in the future?' to b10		
Accuracy	13.) Are there major differences with regard to the reporting quality between project evaluations commissioned by the MELA team and Country Offices?	FoO: Comparison of centralised and decentralised evaluations	Text analysis Use of assessment checklist Data comparison between centralised and decentralised evaluations	Evaluation reports
Accuracy	14.) What are the main influencing factors on the reporting quality of evaluation reports?	FoO: factors influencing reporting quality It will not be possible to explore this question based on empirical evidence, as many factors that could potentially influence the quality of reports (e.g. experience and technical expertise of evaluators, resources dedicated to the evaluation) cannot be collected in the framework of the meta-evaluation. Therefore, the focus will be on exploring opinions / impressions of stakeholders	Mail Survey Interviews	All stakeholders
	b.) Process quality (Utility, feasibility, fairness)			
Fairness	1.) To what extent were obligations of the formal parties to the evaluations (what is to be done, how, by whom, when) agreed to in writing, so that these parties are obliged to adhere to all conditions of the agreement or to renegotiate it?	FoO: Clarity of division of tasks and responsibilities	Mail Survey Interviews	MELA, HoP? Or CD? (clarify)
Utility	2.) Did the evaluations ensure that interests of the persons or groups	FoO: Involvement of different stakeholders in the preparation of the evaluation and its design	Mail Survey Interviews	MELA, HoP, CD

Key areas	Evaluation questions and sub-questions	Assessment criteria / fields of observation (FoO)	Methods of data collection and analysis	Sources of information
	involved in or affected by the evaluation were identified, so that their interests can be clarified and taken into consideration when designing the evaluation?			
Utility	3.) To what extent did the evaluations ensure that the evaluation is guided by both the clarified purposes of the evaluation and the information needs of its intended users?	FoO: Results orientation of the evaluation	Mail Survey Interviews	MELA, HoP, CD
Utility	4.) To what extent were the person(s) conducting an evaluation trustworthy as well as methodologically and professionally competent, so that the evaluation findings achieve maximum credibility and acceptance?	FoO: Competences of evaluators (technically and methodologically)	Mail Survey Interviews	MELA, HoP, CD
Utility	5.) To what extent were the evaluations initiated and completed in a timely fashion, so that its findings can inform pending decision and improvement processes?	FoO: Timing of the evaluation (too late to take into account for strategic decisions or design of successor project, too early to measure what needed to be measured)	Mail Survey Interviews	MELA, HoP, CD
Feasibility	6.) Is the time dedicated to the evaluations sufficient for the evaluand?	FoO: Efficiency (enough time to carry out a thorough and professional evaluation, adequate cost / benefit relation)	Mail Survey Interviews	MELA, HoP, CD
Utility	7.) To what extent were the evaluations planned, conducted and reported in ways that encouraged the acceptance and	FoO: Acceptance of evaluation findings / willingness to utilize findings	Mail Survey Interviews	MELA, HoP, CD

Key areas	Evaluation questions and sub-questions	Assessment criteria / fields of observation (FoO)	Methods of data collection and analysis	Sources of information
	ultimately utilization of the evaluation findings?			
Fairness	8.) Did all stakeholders, to the extent possible, have access to the evaluation findings?	FoO: Transparency / Dissemination of evaluation results	Mail Survey Interviews	MELA, HoP, CD
	9.) Are there major differences with regard to the process quality between project evaluations commissioned by the MELA team and Country Offices?	FoO: Comparison of centralised and decentralised evaluations	Data comparison between centralised and decentralised evaluations (summary of results from mail survey and interviews)	MELA, HoP, CD
	10.) Do the existing Welthungerhilfe formats, templates, materials and advisory services suffice to ensure good quality evaluation processes?	FoO: Quality / completeness / utility of supporting material and advisory services Here, complete: What should be included in the future?	Mail Survey Interviews	All stakeholders
	11.) What are the main influencing factors on the quality of the evaluation process?	FoO: factors influencing process quality It will not be possible to explore this question based on empirical evidence, as many factors that could potentially influence the quality of the process (extent to which all stakeholders had access to findings, timeliness for decision making) cannot be collected in the framework of the meta-evaluation. Therefore, the focus will be on exploring opinions / impressions of stakeholders	Mail Survey Interviews	All stakeholders
	c.) Meta-Analysis: Patterns of reoccurring findings and recommendations			
	12.) Which patterns of reoccurring findings and recommendations	FoO: Contents of recommendation, strategic learning / decision-making interests	Text analysis	Evaluation reports

Key areas	Evaluation questions and sub-questions	Assessment criteria / fields of observation (FoO)	Methods of data collection and analysis	Sources of information
	could be relevant beyond the project-context and thus bear a learning potential for Welthungerhilfe as an organisation and can serve as a basis for a (strategic) decision making?		Content analysis (by Mayring method)	
	d.) Methodological Reflection: Meta-Evaluation /-Analysis as part of the Welthungerhilfe evaluation system			
	13.) Can the meta-evaluation (conceptualized as is; with the current reporting quality and quantity) generate reliable data on trends and patterns that allow for organisational learning and (strategic) decision-making? If no, which preconditions would be needed to allow for data on trends and patterns?	FoO: Résumé, suggestions / recommendations for future meta-evaluations Critical discussion / reflection on the process and the results, potentials and limitations	Reflection of meta-evaluation process and results	Evaluator, MELA HQ

Annex 4: Evaluation Quality Assessment Checklist

(based on DeGEval standards, the Welthungerhilfe ToR of April 2017, old Welthungerhilfe assessment checklist)

Criterion / Quality indicator	Description / Definition	Grading / Rating
1 General Features (Overall Quality)		Maximum points achievable: 9 Section overall quality: Fully fulfilled / to a great extent: 8-9 Mainly fulfilled: 6-7 Rather unsatisfying: 4-5 Unsatisfying: < 4
1.1 The structure of the report is clear and coherent (EQ 1)	1.1.1 The background / evaluation objectives are presented before the findings 1.1.2 The findings are presented before the conclusions 1.1.3 The conclusions are presented before the recommendations	1.1.1 yes: 1, no: 0 1.1.2 yes: 1, no: 0 1.1.3 yes: 1, no: 0 Maximum points 3
1.2 The executive summary is a stand-alone section presenting the main information of the evaluation (EQ 2)	1.2.1 The executive summary is comprehensible as a stand-alone document 1.2.2 a) The executive summary presents a summary of all major elements of the evaluation (i.e. description of project and framework conditions (1), findings on all five DAC criteria – relevance (1), effectiveness (1), efficiency (1), outcomes / impact (1), sustainability (1), conclusions / lessons learnt (1), major recommendations (1)) OR - b) The executive summary presents a summary of at least five (out of the eight) of the major elements of the evaluation OR c) The executive summary presents a summary of less than five (out of eight) of the major elements of the evaluation 1.2.3 The language of the executive summary is sufficiently adequate to allow for publishing	1.2.1 yes: 1, no: 0 1.2.2 a) 1 b) 0,5 c) 0 1.2.3: yes: 1, no: 0 Maximum points: 3

Criterion / Quality indicator	Description / Definition	Grading / Rating
1.3 The length of the report is adequate to cover the major aspects of the evaluation and at the same time be economic to read (additional)	1.3.1 a) The main section of the report (excluding table of content, abbreviations, executive summary and annexes) is between 20 to 40 pages long OR - b) The main section of the report (excluding table of content, abbreviations, annexes and executive summary) is shorter than 20 pages or longer than 40 pages	1.3.1 yes: 1, no: 0 Maximum points: 1
1.4 The report includes an assessment of how the project addresses gender issues and how women / men benefit from project interventions (additional)	1.4.1 The report discusses if / how the project addresses gender issues 1.4.2 The assessment of effectiveness / outcomes / impacts includes observations on how women / men benefit from project interventions	1.4.1 yes: 1, no: 0 1.4.2 yes: 1, no: 0 Maximum points 2
2 Methodology / Validity and reliability of information		Maximum points achievable: 14 Section methodology: Fully fulfilled / to a great extent: 13-14 Mainly fulfilled: 11-12 Rather unsatisfying: 9-10 Unsatisfying: < 9
2.1 An inception report or minutes of the kick-off meeting replacing the inception report outline the methodology to be applied (additional)	2.1.1 The report indicates that an inception report / minutes of the kick-off meeting has documented the methodology to be applied	2.1.1 yes: 1, no: 0 Maximum points 1
2.2 The report describes the data collection methods and analysis, the rationale for selecting them (EQ 6a)	2.2.1 The report describes the data collection methods 2.2.2 The report describes the rationale for selecting the data collection methods selected 2.2.3 It is plausible that the data collection methods have been chosen taking into account the framework conditions (e.g. resources for the evaluation, answering the specific evaluation questions, characteristics of the project or the target group)	2.2.1 yes: 1, no: 0 2.2.2 yes: 1, no: 0 2.2.3 yes: 1, no: 0 2.2.4 yes: 1, no: 0 Maximum points 4

Criterion / Quality indicator	Description / Definition	Grading / Rating
	2.2.4 The report describes the limitations of the chosen methods	
2.3 The report describes the data sources and the rationale for their selection (EQ 6b)	2.3.1 The report describes the data sources 2.3.2 The report describes the rationale for the selection of the data sources	2.3.1 yes: 1, no: 0 2.3.2 yes: 1, no: 0 Maximum points 2
2.4 The report includes a discussion on how the mix of data sources was used to obtain a diversity of perspectives, ensure data accuracy and overcome data limits (EQ 6c)	2.4.1 The report applies a triangulation of methods (using at least two different methods to obtain findings) 2.4.2 The report applies a triangulation of data (using at least two different sources and including the views of different stakeholder groups)	2.4.1 yes: 1, no: 0 2.4.2 yes: 1, no: 0 Maximum points 2
2.5 The report describes the sampling for the evaluation process (EQ 6d)	2.5.1 The report describes the sample frame 2.5.2 The report describes the rationale for selecting the sample 2.5.3 The report describes the numbers of the selected sample 2.5.4 The report describes the limitations of the sample	2.5.1 yes: 0.5, no: 0 2.5.2 yes: 0.5 no: 0 2.5.3 yes: 0.5 no: 0 2.5.4 yes: 0.5 no: 0 Maximum points 2
2.6 The report explains if the evaluation has avoided duplications in data collection by relying as far as possible on existing data	2.6.1 The report explains in how far existing data (e.g. from baselines, partners, the monitoring system) was included into the assessment	2.6.1 yes: 1, no: 0 Maximum points: 1
2.7 The report explains if the evaluation was designed as a participatory process (EQ 7)	2.7.1 The report explains how the evaluation design addressed the participation of stakeholders in the evaluation process 2.7.2 The evaluation report states at least one reason for the level of participation selected	2.7.1 yes: 1, no: 0 2.7.2 yes: 1, no: 0 Maximum points 2
3 Analysis along DAC criteria (EQ 8)		Maximum points achievable: 19 Section DAC criteria assessment: Fully fulfilled / to a great extent: 17-19 Mainly fulfilled: 14-16 Rather unsatisfying: 11-13

Criterion / Quality indicator	Description / Definition	Grading / Rating
		Unsatisfying: < 11
3.1 The report adequately assesses the relevance of the project	3.1.1 The report discusses to what extent the activities and outputs of the project are consistent with the overall goal 3.1.2 The report discusses to what extent the project addresses a core problem of the target group 3.1.3 The report addresses to what extent the project is in line with relevant strategies (e.g. by Welthungerhilfe, major international / national strategies)	3.1.1 yes: 1, no: 0 3.1.2 yes: 1, no: 0 3.1.3 yes: 1, no: 0 Maximum points 3
3.2 The report adequately assesses the effectiveness of the project	3.2.1 The evaluation report presents findings on the extent to which objectives are being achieved / have been achieved 3.2.2 The evaluation report refers to the logframe indicators to assess the effectiveness of the project or does he/she explain, why the indicators were not used (EQ 5) 3.2.3 The report explains which factors influenced the achievement or non-achievement of the objectives	3.2.1 yes: 2, no: 0 3.2.2 yes: 2, no: 0 3.2.3 yes: 2, no: 0 Attention: Because of the high importance of effectiveness, all points are weighed with a factor of 2, therefore a "yes" counts for 2 points, instead of 1 Maximum points 6
3.3 The report adequately assesses the efficiency of the project	3.3.1 The report allocates at least for one example costs to outputs 3.3.2 The report discusses if outputs or activities could have been implemented with less resources 3.3.3 The report provides plausible reasons why efficiency is rated high / medium / low	3.3.1 yes: 1, no: 0 3.3.2 yes: 1, no: 0 3.3.3 yes: 1, no: 0 Maximum points 3
3.4 The report adequately assesses the outcomes / impacts of the project	3.4.1 The report reflects on the existence of evidence for outcomes and impact 3.4.2 The report assesses the plausibility of the project contributing to long-term change 3.4.3 The report reflects on unintended (positive or negative) outcomes	3.4.1 yes: 1, no: 0 3.4.2 yes: 1, no: 0 3.4.3 yes: 1, no: 0 Maximum points 3
3.5 The report adequately assesses the sustainability of the project	3.5.1 The report reflects on the existence and quality of exit strategies to increase sustainability	3.5.1 yes: 1, no: 0 3.5.2 yes: 1, no: 0

Criterion / Quality indicator	Description / Definition	Grading / Rating
	3.5.2 The report reflects on the likely challenges for sustainability 3.5.3 The report reflects on the likely degree of sustainability for at least two activities / outputs / outcomes	3.5.3 yes: 1, no: 0 Maximum points 3
3.6 The report bases findings on evidence	3.6.1 The report relates findings to evidence derived from data collection and analysis methods	3.6.1 yes: 1, no: 0
4 Quality of conclusions and recommendations		Maximum points achievable: 16 Attention: Because of the high importance of conclusions and recommendations, all points are weighed with a factor of 2, therefore a “yes” counts for 2 points, instead of 1 Section conclusions and recommendations: Fully fulfilled / to a great extent: 15-16 Mainly fulfilled: 13-14 Rather unsatisfying: 11-12 Unsatisfying: < 11
4.1 Conclusions are based on findings (EQ 9)	4.1.1 The report differentiates between findings/analysis, conclusions and recommendations 4.1.2 The report bases at least 80% of its conclusions on findings	4.1.1 yes: 1, no: 0 4.1.2 yes: 1, no: 0 Maximum points 2
4.2 Conclusions relate to the evaluation purpose (EQ 9)	4.2.1 At least two of the conclusions relate to the objectives of the evaluation as stated in the chapter on the evaluation purpose	4.2.1 yes: 1, no: 0 Maximum points: 1
4.3 To what extent does the report provide quality recommendations? (EQ 10)	4.3.1 The report states recommendations, i.e. it does provide advice on what should be done to improve project performance / the achievement of objectives (or for final evaluations recommendations for future similar projects)	4.3.1 yes: 1, no: 0 4.3.2 yes: 1, no: 0 4.3.3 yes: 1, no: 0 4.3.4 yes: 1, no: 0 4.3.5 yes: 1, no: 0 Maximum points 5

Criterion / Quality indicator	Description / Definition	Grading / Rating
	<p>4.3.2 There is a clear link of recommendations to findings / conclusions</p> <p>4.3.3 The recommendations are targeted at different actors</p> <p>4.3.4 The recommendations are SMART (i.e. address specific actors, prioritised, realistic)</p> <p>4.3.5 The number of recommendations is adequate (not less than 5, not more than 20)</p>	
Total: Report Quality		<p>Maximum points achievable: 58</p> <p>Section overall quality: Fully fulfilled / to a great extent: 51-58 Mainly fulfilled: 43-50 Rather unsatisfying: 35-42 Unsatisfying: < 35</p>

Annex 5: Individual Checklist Evaluation Report Quality

Criterion / quality indicator	Grading / rating	Comments
Project Title:		
Project Number:		
Evaluator(s):		
Responsible for WHH:		
Date:		
2 General Features (Overall quality)	of 9	
1.1 The structure of the report is clear and coherent	(of 3)	
1.2 The executive summary is a stand-alone section presenting the main information of the evaluation	(of 3)	
1.3 The length of the report is adequate to cover the major aspects of the evaluation and at the same time be economic to read	(of 1)	
1.4 The report includes an assessment of how the project addresses gender issues and how women / men benefit from project interventions	(of 2)	
2 Methodology / Validity and reliability of information	of 14	
2.1 An inception report or minutes of the kick-off meeting replacing the inception report outline the methodology to be applied	(of 1)	
2.2 The report describes the data collection methods and analysis, the rationale for selecting them	(of 4)	
2.3 The report describes the data sources and the rationale for their selection	(of 2)	
2.4 The report includes a discussion on how the mix of data sources was used to obtain a diversity of perspectives, ensure data accuracy and overcome data limits	(of 2)	
2.5 The report describes the sampling for the evaluation process	(of 2)	
2.6 The report explains if the evaluation has avoided duplications in data collection by relying as far as possible on existing data	(of 1)	
2.7 The report explains if the evaluation was designed as a participatory process	(of 2)	
3 Analysis along DAC criteria	of 19	
3.1 The report adequately assesses the relevance of the project	(of 3)	
3.2 The report adequately assesses the effectiveness of the project	(of 6)	
3.3 The report adequately assesses the efficiency of the project	(of 3)	

Criterion / quality indicator	Grading / rating	Comments
3.4 The report adequately assesses the outcomes / impacts of the project	(of 3)	
3.5 The report adequately assesses the sustainability of the project	(of 3)	
3.6 The report bases findings on evidence	(of 1)	
4 Quality of conclusions and recommendations	of 16 (sum*2)	
4.1 Conclusions are based on findings	(of 2)	
4.2 Conclusions relate to the evaluation purpose	(of 1)	
4.3 To what extent does the report provide quality recommendations?	(of 5)	
Total: Report Quality	of 58	

Annex 6: Mail Survey Questionnaire

Dear Country Directors, MELA Focal Points and Heads of Project,

your project (or projects in your country) has/have been evaluated in 2016 and this is why we would like to include you in our survey for evaluating the evaluations.

Since 2014, Welthungerhilfe is carrying out yearly meta-evaluations of all evaluation reports. So far, the meta-evaluation was mainly a desk study, looking at the (formal) quality of the reports. But important aspects like the utility of the evaluation process cannot be explored by analysing a report. Maybe the report is formally perfect, but the recommendations seem useless for project steering? Maybe you got very useful recommendations, but the credibility of the report is low, because the evaluator has not been able to present the methodology according to standards? If the aim is 'learning from evaluations', then the aspect of usefulness is key.

This years' meta-evaluation will therefore include an assessment of the quality of the evaluation process. Please, support us by taking approximately 20-30 minutes to answer the following questions.

If you find it difficult to answer the questions or some of the questions, because you have not been involved in the evaluation or in some aspects of it, please, skip the question(s) and simply state 'not involved' in the space for specifications. If you have been involved in different evaluations in 2016, you can provide an overall opinion in the 'I agree – I disagree' section, and give more specific answers on each evaluation also in the space for specifications, for example if you feel that one evaluation has been much more useful than another one.

As this is a mail survey, your answers are anonymous and confidential and results will only be displayed in a general way. Please, fill in this Word document and send it to: andreasouza@online.de

We apologize that the questionnaire is only provided in English. Time resources did not allow to prepare questionnaires also in German, French, Spanish and Portuguese. We trust that your English will be sufficient to understand the questions. Of course, you can answer in a major international language you feel most comfortable with (English, German, French, Spanish, Portuguese).

If you have any further comments or questions, please contact the external evaluator, Andrea Queiroz de Souza:

[Deadline for returning the questionnaire: Friday, 2nd of June 2017](#)

Thank you very much for your cooperation!

Type of evaluation:

- Final evaluation:
- Mid-term evaluation:
- Other: If other, please specify:
- Not sure:

Responsibility for the evaluation

- Centralised evaluation (managed by MELA staff from headquarters)
- Decentralised evaluation (managed directly by country office or project itself)
- Not sure

The usefulness and utilisation of the evaluation

- 1.) Overall, the results of the evaluation have been relevant to the evaluation's purpose.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

- 2.) The recommendations of the evaluation have been useful for the project.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

- 3.) The recommendations have been feasible (e.g. reflected a good understanding of Welthungerhilfe and its potentials and constraints to follow up on recommendations).

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

- 4.) Some of the recommendations could be useful beyond the project for Welthungerhilfe as an organisation (e.g. for the strategy, for similar projects).

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

- 5.) Welthungerhilfe is making an effort to use recommendations from project evaluations at other levels of the organisation.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

- 6.) How did you use the recommendations? (Several answers possible)

Discuss with staff and agree on action

Discuss with Country Office and agree on action

Fill in management response sheet

Fill in management response sheet and documenting the implementation of the recommendations

Other Please specify:

Not at all

If not, please state the reasons

The evaluation process

- 1.) The evaluator(s) was /were fair / unbiased / impartial.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

- 2.) The evaluator(s) were technically competent for the evaluation topic.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

- 3.) The evaluator(s) were methodologically competent for carrying out the evaluation.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

- 4.) The evaluator(s) showed a good understanding of project management reality.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

5.) Adequate resources were dedicated to the evaluation.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

6.) The evaluation was carried out at the right time to fulfil its objectives.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

7.) Key stakeholders (e.g. partners, donors, target groups) were adequately involved in preparing for the evaluation.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify (e.g. how, why not):

8.) Key stakeholders (e.g. partners, donors, target groups) were adequately involved in the entire evaluation process.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify (e.g. how, why not):

9.) The Terms of Reference / the contract with the evaluator(s) clearly stated the responsibilities and the obligations of all parties involved (who was to do what and when).

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify (e.g. how, why not):

10.) The support provided by headquarters and/or the country office throughout the evaluation process was adequate.

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify (e.g. how, why not):

11.) Welthungerhilfe formats, templates, materials and advisory services have been helpful for promoting good quality evaluation processes?

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify (e.g. how, why not):

12.) Is there anything you would like to see included in the format, templates, materials and advisory services in the future?

13.) What did you do to provide access to the evaluation findings for the different stakeholders? (e.g. dissemination of report to whom, inclusion of partners in debriefing workshop)

14.) The evaluation was conducted in a way that encouraged acceptance and utilisation of the evaluation findings

I agree	I rather agree	I rather disagree	I disagree

Please, give reasons for your answer / specify:

15.) In your opinion, what are the main factors influencing the quality of an evaluation?

16.) Which recommendations would you give to your colleagues who are preparing for an evaluation of their projects? What has to be taken into account when planning and implementing an evaluation to ensure useful results?

Annex 8: List of Interviewees

WHH Head Office	
Mathias Mogge	Executive Director Programmes
Stephan Simon	Acting Head Sector Strategy, Knowledge & Learning Unit Advisor Basic Infrastructure, WASH
Bärbel Mosebach	Head of Humanitarian Directorate
Margrit Röhm	Sector Advisor Civil Society
Ursula Wüst	MELA Team, Advisor M&E
Bernhard Hoeper	MELA Team, Advisor Evaluation
Leila Broich	MELA Team, Advisor M&E
WHH Countries	
Jochen Moninger	Former Country Director Sierra Leone (today: innovation manager)
Dirk Reber	Country Director Cambodia and Laos
Susana Zschocke	Programme Coordinator Madagascar
Surendra Gautam	Programme Coordinator Nepal
Estery Madavo	M&E coordinator Zimbabwe
Jesco Weickert	Former Head of Project Syria (today: MENA coordinator)
Others	
Hendrik Hempel	Evaluator (international)

Annex 9: Bibliography

Note: The 30 evaluated evaluation reports are not listed in the bibliography, they are listed in annex 2

CEval and HWWI (2008): Systemprüfung der Evaluation in der deutschen Entwicklungszusammenarbeit, Fallstudie im Auftrag des BMZ: Welthungerhilfe. Hamburg / Saarbrücken

GIZ (2016): Meta-Evaluierung der Projektevaluierungen, Bonn und Eschborn.

GIZ (2014): Querschnittsauswertung Bildung: Meta-Evaluierung und Synthese, Bonn, Eschborn und Zug.

GIZ (2011): Synthese und Meta-Evaluierung Berufliche Bildung, Bonn und Eschborn.

Mayring, P. (2007): Qualitative Inhaltsanalyse. Grundlagen und Techniken, Weinheim.

Misereor (2016): Meta-Evaluierung der Projektevaluierungen aus den Jahren 2014 – 2015, Report, Aachen.

Welthungerhilfe (2017): MELA Framework, internal document.

Welthungerhilfe (2016): Meta-Evaluation. Welthungerhilfe Decentralized Evaluations 2015, Study, internal document.

Welthungerhilfe (2015): Reviews – Concept, Utilization, Strengths and Weaknesses, Study, Bonn.

Welthungerhilfe (2015): Annotation to the Template, Terms of Reference: Decentralised Evaluation, internal document.

Welthungerhilfe (2010): Wirkungsorientierte Evaluation der Arbeit der Welthungerhilfe. Fachkonzept. Bonn

Welthungerhilfe (2008): Outcome and Impact Orientation in the Projects and Programmes of Welthungerhilfe. Guidelines Part I and II, Bonn

Welthungerhilfe (without date): Template: Decentralised Evaluation, internal document.

Welthungerhilfe (without date): Standards Terms of Reference, Central evaluation of individual projects in the field of development cooperation, internal document.

Widmer, T. (1996): Meta-Evaluation. Kriterien zur Bewertung von Evaluationen, Bern, Stuttgart, Wien, Verlag Paul Haupt.

Web Documents

DeGEval – Gesellschaft für Evaluation (2016): Standards für Evaluation. Erste Revision auf Basis der Fassung 2002, online:

http://www.degeval.de/fileadmin/Publikationen/DeGEval_Standards_fuer_Evaluation_-_Erste_Revision__2016_.pdf (accessed 16.05.2017) and :

<http://www.degeval.de/de/degeval-standards/> (accessed 28.05.2017)

Annex 10: Specific Assessment Criteria for Accuracy and Degree of Fulfilment

Degree of fulfilment of specific assessment criteria

Evaluation standard 'accuracy'	Degree of fulfilment
Assessment criteria	
1 Overall Quality (general features)	
1.1 Clear and coherent report structure	94%
AC1 The background / evaluation objectives are presented before the findings	100%
AC2 The findings are presented before the conclusions	96%
AC3 The conclusions are presented before the recommendations	86%
1.2 The executive summary is a stand-alone section presenting the main information of the evaluation	69%
AC4 The executive summary is comprehensible as a stand-alone document	64%
AC5 The executive summary presents a summary of all major elements of the evaluation	64%
AC6 The language of the executive summary is sufficiently adequate to allow for publishing	79%
1.3 The length of the report is adequate to cover the major aspects of the evaluation and at the same time be economic to read	71%
AC7 The main section of the report (excluding table of content, abbreviations, executive summary and annexes) is between 20 to 40 pages long	71%
1.4 The report includes an assessment of how the project addresses gender issues and how women / men benefit from project interventions	75%
AC8 The report discusses if / how the project addresses gender issues	79%
AC9 The assessment of effectiveness / outcomes / impacts includes observations on how women / men benefit from project interventions	71%
2 Methodology / Validity and reliability of information	
2.1 An inception report or minutes of the kick-off meeting replacing the inception report outline the methodology to be applied	11%
AC10 The report indicates that an inception report / minutes of the kick-off meeting has documented the methodology to be applied	11%
2.2 The report describes the data collection methods and analysis, the rationale for selecting them	68%
AC11 The report describes the data collection methods	96%
AC12 The report describes the rationale for selecting the data collection methods selected	32%

AC13 It is plausible that the data collection methods have been chosen taking into account the framework conditions (e.g. resources for the evaluation, answering the specific)	96%
AC14 The report describes the limitations of the chosen methods	46%
2.3 The report describes the data sources and the rationale for their selection	64%
AC15 The report describes the data sources	96%
AC16 The report describes the rationale for the selection of the data sources	32%
2.4 The report includes a discussion on how the mix of data sources was used to obtain a diversity of perspectives, ensure data accuracy and overcome data limits	95%
AC17 The report applies a triangulation of methods (using at least two different methods to obtain findings)	93%
AC18 The report applies a triangulation of data (using at least two different sources and including the views of different stakeholder groups)	96%
2.5 The report describes the sampling for the evaluation process	33%
AC19 The report describes the sample frame	46%
AC20 The report describes the rationale for selecting the sample	29%
AC21 The report describes the numbers of the selected sample	46%
AC22 The report describes the limitations of the sample	11%
2.6 The report explains if the evaluation has avoided duplications in data collection by relying as far as possible on existing data	86%
AC23 The report explains in how far existing data (e.g. from baselines, partners, the monitoring system) was included into the assessment	86%
2.7 The report explains if the evaluation was designed as a participatory process	36%
AC24 The report explains how the evaluation design addressed the participation of stakeholders in the evaluation process	57%
AC25 The evaluation report states at least one reason for the level of participation selected	14%
3 Analysis along DAC criteria	
3.1 The report adequately assesses the relevance of the project	74%
AC26 The report discusses to what extent the activities and outputs of the project are consistent with the overall goal	50%
AC27 The report discusses to what extent the project addresses a core problem of the target group	89%
AC28 The report addresses to what extent the project is in line with relevant strategies (e.g. by Welthungerhilfe, major international / national strategies)	82%
3.2 The report adequately assesses the effectiveness of the project	95%
AC29 The evaluation report presents findings on the extent to which objectives are being achieved / have been achieved	96%

AC30 The evaluation report refers to the logframe indicators to assess the effectiveness of the project or does he/she explain, why the indicators were not used	89%
AC31 The report explains which factors influenced the achievement or non-achievement of the objectives	100%
3.3 The report adequately assesses the efficiency of the project	37%
AC32 The report allocates at least for one example costs to outputs	32%
AC33 The report discusses if outputs or activities could have been implemented with less resources	32%
AC34 The report provides plausible reasons why efficiency is rated high / medium / low	46%
3.4 The report adequately assesses the outcomes / impacts of the project	62%
AC35 The report reflects on the existence of evidence for outcomes and impact	79%
AC36 The report assesses the plausibility of the project contributing to long-term change	68%
AC37 The report reflects on unintended (positive or negative) outcomes	39%
3.5 The report adequately assesses the sustainability of the project	54%
AC38 The report reflects on the existence and quality of exit strategies to increase sustainability	14%
AC39 The report reflects on the likely challenges for sustainability	68%
AC40 The report reflects on the likely degree of sustainability for at least two activities / outputs / outcomes	79%
3.6 The report bases findings on evidence	93%
AC41 The report relates findings to evidence derived from data collection and analysis methods	93%
4 Quality of conclusions and recommendations	
4.1 Conclusions are based on findings	93%
AC42 The report differentiates between findings/analysis, conclusions and recommendations	96%
AC43 The report bases at least 80% of its conclusions on findings	89%
4.2 Conclusions relate to the evaluation purpose	93%
AC44 At least two of the conclusions relate to the objectives of the evaluation as stated in the chapter on the evaluation purpose	93%
4.3 To what extent does the report provide quality recommendations?	69%
AC45 The report states recommendations, i.e. it does provide advice on what should be done to improve project performance / the achievement of objectives (or for final evaluations recommendations for future similar projects)	96%
AC46 There is a clear link of recommendations to findings / conclusions	86%
AC47 The recommendations are targeted at different actors	54%

AC48 The recommendations are SMART (i.e. address specific actors, prioritised, realistic)	36%
AC49 The number of recommendations is adequate (not less than 5, not more than 20)	75%

Annex 11: Management Response Sheet Meta-Evaluation

Meta-evaluation/analysis of Project Evaluations 2016				
Responsible for completing management response: MELA				
Evaluation recommendation 1: The MELA evaluation advisors should assume the responsibility for a comprehensive capacity building of the MELA focal points on evaluation management				
Recommendation to: MELA evaluation advisors			Priority level 3	
Management response – Agree / partially agree / disagree (if recommendation is rejected or partially accepted, pls. provide an explanation):				
Key action(s)	Timeframe (or deadline)	Responsible unit(s)	Tracking (or monitoring)	
			Comments or action taken	Status
1.1 Design a capacity development strategy on evaluation management for MELA focal points	Until December 2017	MELA evaluation advisors		
1.2 Update support materials: Bring together the different (partly outdated) materials into one comprehensive document on evaluation management guidelines (including the finalisation of evaluation standards and requirements)	Until July 2018	MELA evaluation advisors		
1.3 Follow through with updating the evaluation information on the O365 website	Until July 2018	MELA evaluation advisors		
1.4 Provide regional evaluation management trainings for all MELA focal points, presenting and working with the support materials	August 2018 – March 2019	MELA evaluation advisors		

Evaluation recommendation 2: The MELA evaluation advisors should assume the responsibility for the (coordination of) quality assurance of evaluation management processes.				
Recommendation to: MELA evaluation advisors			Priority level 3	
Management response – Agree / partially agree / disagree (if recommendation is rejected or partially accepted, pls. provide an explanation):				
Key action(s)	Timeframe (or deadline)	Responsible unit(s)	Tracking (or monitoring)	
			Comments or action taken	Status
2.1 Define evaluation report quality criteria which are key for WHH	Until December 2017	MELA evaluation advisors		
2.2 Clarify quality assurance processes and responsibilities	Until December 2017	MELA evaluation advisors		
2.3 Carry out quality checks of evaluation reports and provide feedback to CO	From January 2018	MELA evaluation advisors		
2.4 Monitor evaluation coverage and quality development of evaluation reports	From January 2018	MELA evaluation advisors		
Evaluation recommendation 3: The MELA evaluation advisors should assume the responsibility for the coordination of the establishment of processes and mechanisms to generate strategic findings and recommendations from DE and strategic evaluations and to communicate findings of strategic interest within HO				
Recommendation to: MELA evaluation advisors			Priority level 2	
Management response – Agree / partially agree / disagree (if recommendation is rejected or partially accepted, pls. provide an explanation):				
Key action(s)	Timeframe (or deadline)	Responsible unit(s)	Tracking (or monitoring)	
			Comments or action taken	Status
3.1 • Facilitate a discussion on feasible mechanisms / processes to generate strategic findings and	Until March 2018	MELA evaluation advisors		

recommendations from DE and to communicate findings of strategic interest within HO				
3.2 Test agreed mechanisms / processes and revise based on results	April to December 2018	MELA evaluation advisors		
3.3 • Establish and consolidate processes for generating strategic findings and recommendations from DE and for communications the findings within HO	From December 2018	MELA evaluation advisors		
Evaluation recommendation 4: To free resources for the key activities under recommendations 1–3, the MELA evaluation advisors should seek to reduce the number of strategic evaluations and CE to a minimum until the above-mentioned processes have been consolidated				
Recommendation to: MELA evaluation advisors		Priority level 3		
Management response – Agree / partially agree / disagree (if recommendation is rejected or partially accepted, pls. provide an explanation):				
Key action(s)	Timeframe (or deadline)	Responsible unit(s)	Tracking (or monitoring)	
			Comments or action taken	Status
-	January 2018 – December 2019	MELA evaluation advisors		
Evaluation recommendation 5: MELA should assume the responsibility for setting up a network of MELA focal points in all countries and develop them into a functional community of practice				
Recommendation to: MELA		Priority level 3		
Management response – Agree / partially agree / disagree (if recommendation is rejected or partially accepted, pls. provide an explanation):				
Key action(s)	Timeframe (or deadline)	Responsible unit(s)	Tracking (or monitoring)	
			Comments or action taken	Status
5.1 Appoint one person within MELA who will take the lead in establishing a MELA community of practice	Until October 2017	MELA		

5.2 That person would lobby with regional directors (RD) and CD and demand the appointment of MELA focal points for each country		MELA		
5.3 MELA focal points for each country should be appointed and known by December 2017	Until December 2017	MELA		
5.4 • Establish rapport with the MELA focal points by contacting all on an individual basis (phone calls)	October 2017 to March 2018	MELA		
Evaluation recommendation 6: CD / CO should strengthen the strategic use of evaluations at country level and seek to improve the evaluation management processes in their respective country				
Recommendation to: CD / CO		Priority level 3		
Management response – Agree / partially agree / disagree (if recommendation is rejected or partially accepted, pls. provide an explanation):				
Key action(s)	Timeframe (or deadline)	Responsible unit(s)	Tracking (or monitoring)	
			Comments or action taken	Status
6.1 • Appoint a qualified MELA focal point and provide his/her contact details to MELA HO (until December 2017)	Until December 2017	CD		
6.2 Communicate evaluation requirements to country staff	Until September 2018	CD		
6.3 • Make sure that evaluations are already planned for and budgeted during the development of the project proposal	From now onwards	CD		
6.4 Make sure management response sheets are included as a deliverable in the evaluation ToR and follow-up the implementation of recommendations	From now onwards	CD		

by participating in one discussion on the operationalisation of recommendations immediately after the evaluation and participate in a second discussion several months later to follow up on the implementation status				
Evaluation recommendation 7: Sector advisors should proactively seek to know which evaluations are planned within their sectors, offer advice for designing evaluation questions and seek to include a few strategic sector evaluation questions into the ToR. They should further seek to have a phone debriefing with the HoP, MELA focal point or CD to get details on the evaluation results				
Recommendation to: Sector advisors			Priority level 3	
Management response – Agree / partially agree / disagree (if recommendation is rejected or partially accepted, pls. provide an explanation):				
Key action(s)	Timeframe (or deadline)	Responsible unit(s)	Tracking (or monitoring)	
			Comments or action taken	Status
7.1 Seek to know which evaluations are planned within their sectors	From now onwards	Sector Advisors		
7.2 Offer advice for designing evaluation questions to MELA focal points	From now onwards	Sector Advisors		
7.3 Seek to include a few strategic sector evaluation questions into the ToR.of project evaluations within the respective sector	From now onwards	Sector Advisors		
7.4 Have a phone debriefing with the HoP, MELA focal point or CD to get details on the evaluation results	From now onwards	CD		
Evaluation recommendation 8: The Executive Director Programmes should communicate mandatory requirements and standards on evaluation to RD and CD and demand for feedback on compliance				
Recommendation to: Executive Director Programmes			Priority level 3	

Management response – Agree / partially agree / disagree (if recommendation is rejected or partially accepted, pls. provide an explanation):

Key action(s)	Timeframe (or deadline)	Responsible unit(s)	Tracking (or monitoring)	
			Comments or action taken	Status
8.1 Communicate mandatory requirements and standards on evaluation to RD and CD	Right after mandatory requirements have been defined and agreed on)	Executive Director Programmes		
8.2 Demand for feedback on compliance	6 months later, 12 months later	Executive Director Programmes		

Evaluation recommendation 9: The Board should approve the MELA framework and allocate the resources for its implementation. This will formalise the MELA focal points and provide the necessary leverage for the MELA team to proceed with the decentralisation of MELA responsibilities (to the Board, priority level 3, as soon as possible, latest before December 2017).

Recommendation to: Board | **Priority level 3**

Management response – Agree / partially agree / disagree (if recommendation is rejected or partially accepted, pls. provide an explanation):

Key action(s)	Timeframe (or deadline)	Responsible unit(s)	Tracking (or monitoring)	
			Comments or action taken	Status
9.1 Approve the MELA framework	As soon as possible, latest before December 2017	Board		
9.2 Allocate the resources for its implementation	As soon as possible, latest before December 2017	Board		
9.3 Communicate decision within the organisation	As soon as possible, latest before December 2017	Board		

