



**WFP EVALUATION**

WFP  
World Food Programme

SAVING  
LIVES  
CHANGING  
LIVES

# Evaluation Methods Advisory Panel at WFP

2022 in Review

December 2022



# Contents

Foreword .....	3
Introduction .....	4
1. Approaches and methodology .....	6
2. Use of theory-based evaluation .....	9
3. Evaluability assessment and linkages with evaluation design .....	11
4. Triangulation, clarity and transparency .....	13
5. WFP evaluation guidance.....	15
6. Lessons to strengthen WFP's evaluation function.....	17
Annex 1: Short biographies of members of the EMAP .....	19
Annex 2: Evaluation Documents Reviewed by the EMAP .....	20
Annex 3: Selection of evaluations for review by the EMAP .....	21
Annex 4: Guidance on development of evaluation questions by type of evaluation .....	22
Endnotes.....	23

## Foreword

Over the past few years, the work of the WFP Office of Evaluation (OEV) has grown significantly, both in terms of the number of evaluations and the increasingly complex and diverse contexts in which WFP operates. While taking major steps to ensure continued scale-up of its work, OEV has also devoted attention to diversify evaluation approaches and methods.

Among the recommendations of the independent peer review of the WFP evaluation function, conducted in 2020, was to “experiment with various evaluation approaches and methodologies and offer an expanded menu of evaluation tools”. In response, the updated WFP Evaluation Policy 2022 and WFP Corporate Evaluation Strategy 2022 committed to exploring innovative and adaptive evaluation approaches and methods for all evaluation types, ensuring that WFP evaluations remain relevant and useful.

This same year, 2022, OEV launched an Evaluation Methods Advisory Panel (EMAP) with the aim of playing a key role in fostering innovation in approaches and methods. During its 12-month pilot phase, the EMAP was composed of seven external evaluation advisers who provided expert and independent advice on a selection of diverse draft and completed evaluation products. The intention was to assist the Office of Evaluation and Regional Evaluation Units reflect and improve evaluation approaches and methods, identifying systemic and structural challenges and shed light on international best practices and innovations on methods in evaluation.

More broadly, the EMAP is designed as a new feature of WFP’s evaluation function with the aim of strengthening the credibility and utility of independent evaluations.



It is distinct but complementary to other quality assurance, support and assessment systems currently established, namely the work conducted by evaluation managers and second level quality assurers, the feedback provided by internal reference groups and external advisory groups, as well as reviews conducted by the Decentralized Evaluation Quality Support and the Post-Hoc Quality Assessment services.

This EMAP Annual Report brings together all the advisory support and reviews conducted during 2022, highlighting five key themes: (1) Evaluation approaches and methods; (2) Use of theory-based evaluation; (3) Linkages between elements of the evaluation design; (4) Triangulation, clarity and transparency; and (5) WFP evaluation guidance. Above all, it offers a route map for WFP’s evaluation function to review its current practices and remain innovative in a rapidly changing landscape while continuing to strengthen the quality and utility of future evaluations and guidance.

**Andrea E. Cook**  
Director of Evaluation

# Introduction

## The Evaluation Methods Advisory Panel

Given the increase in the number of evaluations and the complex and diverse contexts in which the World Food Programme (WFP) operates, the WFP Office of Evaluation (OEV) has created an Evaluation Methods Advisory Panel (EMAP) to provide advice on WFP's evaluations to improve evaluation methodology, approaches and methods, and reflect on international best practice and innovations in these areas.

The panel, composed of seven senior external evaluation experts (see Annex 1), was launched in January 2022. It complements provisions in the WFP evaluation quality assurance system (EQAS).

## The EMAP Annual Report

The aims of the review are to:

- reflect on the overall evaluation approaches and methods used across evaluations
- identify systemic and structural challenges
- derive lessons to increase quality and utility in future evaluations and guidance

The report covers all types of evaluations conducted by WFP's evaluation function - Policy Evaluations (PEs), Complex Emergency Evaluations (CEEs), Strategic Evaluations (SEs), Decentralized Evaluations (DEs), Country Strategic Plan Evaluations (CSPEs) - in 2021-2022, apart from impact evaluations, which are covered by another mechanism.<sup>1</sup> It is based on a set of specific reviews undertaken by EMAP members (referred to as "the reviewers" below), and online exchange with the reviewers and WFP, as detailed below.

Two approaches to the EMAP reviews were undertaken in 2022: (a) reviewers were given a number of completed CSPE and DE evaluation

reports (8 and 14 respectively), and (b) reviewers were given draft outputs - terms of reference (ToR), concept notes, inception reports, evaluation reports - of a smaller number of ongoing CEEs, PEs and SEs (the list of evaluations reviewed can be found in Annex 2 and the selection criteria in Annex 3). The process of preparing this annual report was conducted by two EMAP advisers and entailed:

- Review of the advice provided by EMAP on WFP evaluations during 2022
- Validation meeting with the other EMAP advisers to deepen and nuance findings
- Discussion of the draft annual report with OEV and Regional Evaluation Officers. This report incorporates key elements from these discussions.

The report faced a number of challenges: (i) the small number of some types of evaluations covered by the reviews (PEs, CEEs and SEs) compared with others (DEs and CSPEs), meant it was easier to pick up patterns in the latter; (ii) for some CSPEs, only final evaluation reports and for DEs, final evaluation reports and the corresponding terms of reference and inception reports were examined, whereas for others (SEs, PEs and CEEs) the reviews largely covered concept notes, inception reports and terms of reference; (iii) all reviews followed a structure provided by WFP which varied by evaluation type and did not address all the issues covered in this report in the same way; (iv) COVID-19 resulted in more remote work which may not be typical of the design and conduct of WFP evaluations outside of a pandemic, and; (v) only reviewing evaluation outputs presented challenges to explaining why something did or did not happen in an evaluation process.

## Structure of the EMAP Annual Report

The following sections will explore issues within a number of themes selected by OEV based on the initial feedback from the EMAP reviews of evaluation documents:

1. Evaluation approaches and methods
2. Use of theory-based evaluation
3. Linkages between elements of the evaluation design
4. Triangulation, clarity and transparency

5. WFP evaluation guidance
6. A final section will bring together lessons and tentative recommendations that may contribute to strengthening the WFP evaluation function.

Each section starts with the question set by OEV for the report, examines the current practice and challenges in response to the question, and identifies good practice that can be leveraged as well as avenues for further exploration by OEV.



WFP/Theresa Piorr

# 1. Approaches and methodology

To what extent are the approaches and methods applied across the evaluations similar or do they include innovative elements? What are some good practices which could be leveraged to enhance the design of evaluations?

## CURRENT PRACTICE AND CHALLENGES

The EMAP review found many common features across the WFP evaluation and inception reports and related ToRs. The WFP evaluations commonly use time-tested approaches and methods – generally, mixed-methods designs combining quantitative and qualitative data collection and analysis. The reviews identified a few good examples of mixed methods approaches adapted to the purpose, subject and context of the respective evaluation, generating new insights for WFP (while fulfilling accountability purposes as well), as specified in the section *Good practice to be leveraged* below. Innovative elements in terms of new developments in the field of evaluation or unconventional methods mixes have been rare, one exception being the application of Qualitative Impact Protocol (QulP), a method developed by the University of Bath (Regional Bureau Johannesburg DE).<sup>ii</sup> Generally, the questions asked in the reviewed evaluations can be answered by applying standard methods of social and economic research – methods that can also be used for reflection on systems, networks and other complex phenomena.

Although the WFP Evaluation Policy supports a culture of learning, evaluation ToRs are frequently oriented mainly towards accountability objectives, covering all OECD-DAC criteria and long lists of evaluation questions with a focus on performance (“how well did WFP...”, “to what extent did...”). Most evaluations are expected to serve both accountability and learning purposes, but the accountability aspect tends to dominate. In

DEs, donor requirements, which sometimes specify pre-determined indicators and processes for monitoring and evaluation, often prioritise accountability. A focus on accountability risks limiting the potential of evaluations to investigate learning questions and generate recommendations that are important for learning and evidence-based strategy development. Therefore, more learning-oriented questions should be included in TOR. For example, an evaluation of a humanitarian intervention could probe into barriers against humanitarian access, strategies used to overcome them, and the effectiveness of those strategies in different contexts. It could focus more on describing evidence for changes and probing into helping and hindering factors. This happens to some extent in CSPEs and PEs, both including a standard evaluation question on explanatory factors, but has not been common in other evaluations.

Since there are many stakeholders in an evaluation process, including the donors of WFP programmes, **consultation** between OEV and stakeholder groups is also important in the preparatory phase of evaluations. The evaluation questions need to be developed through consultation and space needs to be made available for an effective **co-creation process**. At the same time, evaluation questions need to be prioritized over criteria to ensure the evaluation will address the most important issues within the time and resource constraints of the evaluation. Moreover, while it is good to link evaluation questions to relevant evaluation criteria, it is the questions that should drive the evaluations.

The evaluations reviewed often interpret effectiveness as the delivery of outputs and contributions to outcomes, without examining the (intended or unintended) distribution of results across different groups of stakeholders. As a result, equity, a key element in OECD-DAC's current definition of effectiveness, is rarely discussed, or limited to gender issues.

Some EMAP reviewers have pointed out a frequent focus on processes and outputs, producing limited evidence on higher results (outcome level) or on the validity of the hypotheses or assumptions underlying programme design. It is unclear which aspects of the evaluation systems contribute to this issue – apart from the fact that examining outputs is a more straightforward process than determining WFP's contribution to higher-level changes.

Divergences exist in the extent to which participatory approaches and equity orientation, including the use of feminist/gender-responsive or transformative methods, are implemented. Reviewers found participatory elements in some evaluations, but evidence for systematic consultation and participation of a spectrum of intended beneficiaries throughout the evaluation is rare. WFP and partner staff are often the evaluation teams' main interlocutors in key informant interviews, the perspectives of intended beneficiaries being captured mainly in surveys and some group discussions. Limited evidence has been found for deliberate efforts to consult with hard-to-reach groups, e.g., indigenous people and persons with disabilities.

Gender has often been an add-on in evaluation design, receiving systematic attention only in the decentralised thematic evaluations focusing on gender. At the same time, gender is often considered as equal to women's empowerment in the evaluations. Lack of gender-disaggregated monitoring data has been quoted as an issue in several DEs. Reviewers have found limited evidence of intersectional approaches, which would examine positive and negative effects (or lack thereof) of WFP interventions on population groups affected by intersecting conditions linked to marginalisation.

Furthermore, there seem to be divergences in the understanding of the idea of methodology, which in some CSPEs does not go beyond process. Terms such as "theory-based evaluation" and "contribution analysis" are often used in CSPEs without any explanation of what they mean in practice. Consequently, as described in section 4 of this report (Triangulation, clarity and transparency), the review found divergences in the quality of data collection, analysis and reporting.

An important requirement for ensuring appropriate and context specific evaluation approaches is selecting evaluation team with the right skills. This raises the question of how much of the design takes place in the terms of reference, how much in the proposal for undertaking the evaluation and how much in the inception report. No amount of guidance can replace the experience of team members. On the side of OEV, adequate time needs to be given in advance of the start of the evaluation for high quality teams to be put in place.

## **GOOD PRACTICES TO BE LEVERAGED**

The review identified good examples of participatory approaches to evaluation including participatory feminist evaluation (Burkina Faso Gender Thematic DE), and participatory data collection with the support of an experienced national NGO and the oversight of a team member (Bangladesh CSPE). Good practice examples regarding participatory approaches also include the Madagascar School Feeding DE (well-tailored mixed methods).

The Burkina Faso thematic evaluation on gender stands out for its feminist, empowering evaluation approach. It has applied gender analysis to generate inter alia new, stereotype-challenging insights on women's roles in household productivity. Furthermore, it has used participatory analysis of WFP documents with WFP staff to sensitise, train and empower staff. Evaluations that are not focused on gender would greatly benefit from gender analysis to obtain more accurate picture of the benefits and limits of WFP interventions.



WFP should consider encouraging staff managing evaluations to devote careful attention to the development of evaluation questions<sup>iii</sup> that serve the objective and the context of the evaluation. DAC criteria can help inform evaluation questions but should not supersede them.

OEV should further examine the role of equity in its evaluations looking at who benefits or not from its interventions. The ongoing OEV work to ensure people with disabilities and indigenous populations are included in its evaluations is important. But a wider examination of the ways in which intersecting aspects of intended beneficiaries' identities (e.g., gender, age, social class, physical ability and other features) influence the effect of WFP interventions is still needed.

Correspondingly, evaluators should be encouraged to tailor the evaluation approach and method to the purpose, subject and context of the evaluation. This would include tailoring the evaluation questions to the specific context and issue being evaluated, something that is now standardized in CSPEs and Pes (see annex 4). The tailoring process would occur during preparation of the evaluation matrix which is where the questions, lines of enquiry, data sources and methods come together. Evaluators should be encouraged to include a focus on equity as part of their assessment of effectiveness.<sup>iv</sup>



WFP/Damilola Onafuwa

# 2. Use of theory-based evaluation

How could the theory-based evaluation approach and theories of change be more meaningfully designed, implemented and used to strengthen the evaluation design?

## CURRENT PRACTICE AND CHALLENGES

Theories of change (ToC) are found in most WFP evaluations of all types and there is frequent reference to theory-based approaches (especially in CSPEs and DEs), although unequal implementation in practice. In some cases, a theory of change has already been designed for a project, policy or strategy, in others they are reconstructed by the evaluation team. However, there does not appear to be a consistent understanding among evaluation teams of what a theory-based approach to evaluation is and/or the role a theory of change can play in the design of an evaluation. Linkages between the ToC and the evaluation design were often missing. Ideally, a ToC should surface the underlying assumptions or hypotheses that an evaluation should examine. Theory of change should have implications for evaluation design. ToC development is of limited value unless key causal linkages in the ToC are explored in evaluation design, data collection and analysis. In several cases, the ToC and the assumptions identified are poorly integrated into designing the evaluation approach and methodology, and do not appear to have been used in the analysis or reporting.

Many ToCs are poorly constructed in IRs and ERs. In large aggregations of diverse activities (e.g. CSPEs and possibly PEs and SEs), they are necessarily abstract, yielding no testable assumptions. If any assumptions are developed

in the ToC, they are not always treated as testable hypotheses and/or the relationship between the assumptions and the evaluation questions is not always visible. In some global evaluations covering many countries (SEs, PEs), ToCs don't address the heterogeneity of contexts, or, in CSPEs, a single ToC may not cover all components of what is being evaluated. It may be necessary to have several ToCs to take different contexts or different components into account, which could be nested in an overall ToC.

Finally, there is usually no recommendation to update the programme/project theory of change following the evaluation based on what has been learned.

## GOOD PRACTICES TO BE LEVERAGED

Reviews of DEs found that systematic engagement with theories of change tended to yield a clearer evaluation matrix and clearer statements regarding contribution to expected outcomes. Gender thematic DEs in El Salvador and Burkina Faso developed theories of change to surface gender mainstreaming issues. The Burkina Faso Gender DE was particularly compelling for (i) its analysis of problematic implicit assumptions in the existing DE, and (ii) its proposal to use specific gender-transformative, evidence-based assumptions in a revised theory of change.



WFP/Cesar Lopez

## AVENUES FOR EXPLORATION

Guidance should be developed to support the development and utilisation of more effective theories of change and to undertake, when appropriate, theory-based approaches to evaluation. ToC development or reconstruction should be commensurate with the evaluation approach and methods. For instance, an evaluation that uses contribution analysis needs a ToC that shows the underlying assumptions between all levels or the results framework – or at least for those causal pathways/ results chains that the evaluation will apply contribution analysis to.

DE reviewers have recommended applying an equity or intersectional gender lens in theory of change development, surfacing assumptions that are based on gender stereotypes, or that seem to be oblivious of gendered realities in the field of intervention. In view of WFP's commitment to gender mainstreaming, this recommendation should be considered across all types of theory-based evaluations.

It would be helpful for WFP to encourage evaluators to enhance their critical engagement with the theory of change, and to provide more information on the way the theory of change has informed evaluation design.

To make the most of evaluation findings related to the programme/project theory of change, OEV could explore introducing a process to update the ToC based on what has been learned through the evaluation.

# 3 • Evaluability assessment and linkages with evaluation design

How could the linkages between the evaluability assessment (EA), the theory of change and the selection of methods and data collection tools be further strengthened?

## CURRENT PRACTICE AND CHALLENGES

Evaluability assessments are usually undertaken but are not always clearly linked to other elements of the evaluation design. Moreover, the linkages between elements of the evaluation design (objectives, questions, theory of change, stakeholder mapping, evaluation matrix, evaluability assessment, data collection strategy, etc.) also need to be made clear.

There is no systematic use of EAs across the different types of WFP evaluations. When evaluability assessments are undertaken, they are often narrowly focussed on data availability and even then, some do not cover the range of data that is needed. Other elements, such as the theory of change, and stakeholder demands and needs, are generally not addressed. Generally, it is the evaluation team that builds on the initial evaluability assessment found in the ToR, a situation that may generate conflicts of interest, especially if the more detailed EA is negative.

Scope and construct validity rarely receive explicit attention, for example in the form of a discussion of the focus of the evaluation and its rationale, and what that means for its design. Implicit definitions of key constructs (e.g., how WFP defines the theme of a thematic evaluation) are rarely examined and

challenged. Evaluations tend to focus on the question “are we doing things right” instead of “are we doing the right things”.

The selection of data collection and analysis methods is not always linked to addressing needs and challenges identified in the evaluability assessment. Although most evaluations formally adopted a mixed-methods approach, the EMAP reviews found few examples where the method mix was clearly targeted to the purpose, subject and context of the evaluation (see above, section 1). Equally, a more comprehensive evaluability assessment in the ToR may make it easier for evaluation companies to prepare a better bid for the evaluation and select a more appropriate team.

## GOOD PRACTICES TO BE LEVERAGED

The review of the Climate Change PE found the findings of an evaluability assessment preceding one specific evaluation very important. The evaluability assessment developed recommendations of its own, which could be acted upon during the evaluation process. This good practice contrasted with other evaluability assessments which focussed on determining which data were available.

## AVENUES FOR EXPLORATION

Apart from determining the availability of data, EAs should dedicate attention to evaluation stakeholder analysis and programme/project theory of change. This would also make it easier to identify the constituency for each evaluation question and design the evaluation accordingly. The guidance on evaluability assessment should be widened to incorporate evaluability of the Theory of Change, stakeholders' interests and requirements of the

evaluation, and the surrounding physical institutional and political constraints and opportunities on implementation of an evaluation. A useful source of inspiration for OEV could be the Austrian Development Agency's 2022 [guidance document on evaluability assessment](#).

Also, evaluations could include suggestions regarding potential areas for future assessment, i.e., evaluations could contribute to future evaluability assessments.



WFP/Fredrik Lerneryd

# 4 • Triangulation, clarity and transparency

How could data collection and analysis be further strengthened by enhanced triangulation, clarity and transparency

## CURRENT PRACTICE AND CHALLENGES

Triangulation: All WFP evaluations reviewed have used a rather standard mix of data collection methods, generally combining review of documentation (including monitoring data), key informant interviews (KIIs) and surveys, and sometimes focus group discussions (FGDs) and workshops. Triangulation, especially triangulation of data collection methods, is often noted in the methodology section, while triangulation of data analysis methods and of data sources (to include a broad range of rights holders and project stakeholders) is not always clear.

The reviewers have found limited information on the rationale for the choice of a specific evaluation design, for example, how evaluability and context have informed the choice of evaluation tools, and how the selection of case studies has followed from the analysis of the overall subject of the evaluation. A **data analysis plan** is often missing.

Conversely, the reviewers have noted excessive standardisation in CSPEs and DEs, including in the selection of data collection methods. In DEs, allocation of resources (funding, team size) appears to be quite uniform despite vast differences in scope.

While the evaluations have worked with mixed methods and a range of respondents, evaluation reports frequently display significant gaps in describing (i) sampling/case selection, (ii) data analysis processes, and (iii) related bias risk and their mitigation.

Clarity and transparency: All evaluation reports include descriptions of data collection processes and annexes with data collection tools. Some IR and ER describe data collection methods only, omitting any discussion of data analysis processes. As noted above (Approaches and methods), it is often unclear whether the evaluation has been adapted to the context of the evaluation.

The discussion of challenges and limitations is often limited. Risks and potential bias linked to remote data collection (common across the evaluations reviewed), and related mitigation strategies, generally receive little attention. The robustness of quantitative analyses is often unclear: surveys were frequently used in CSPEs and

DEs, but statistical significance is unclear, and it appears they are rarely designed to high professional standards (although higher quality surveys are often found in other types of evaluation).

In terms of data quality, evaluation teams frequently find WFP monitoring data to be lacking or deficient. The voices of intended beneficiaries are rarely represented. Across evaluations, most stakeholders mapped as part of the evaluations are within the WFP system. The frequent lack of gender disaggregated data and intersectional gender analysis suggests that intended beneficiaries tend to be implicitly viewed as a homogenous category rather than people with varying needs, interests and power.

## GOOD PRACTICES TO BE LEVERAGED

Due to restrictions linked to the COVID-19 pandemic, the Gambia CSPE carried out extensive data collection and initial data analysis when preparing the ToRs. This practice could be used in future evaluations, so that evaluation teams can focus on collecting data from intended beneficiaries and external stakeholders, and on deeper analysis.

The Bangladesh CSPE and the PE CSP Policy include a simple table to show how the standard evaluation matrix contained in the inception report template was changed during the inception phase. Importantly, the tables include an explanation of why the changes were made.

## AVENUES FOR FUTURE PRACTICE

Evaluation teams should be discouraged from adopting a one size fits all approach to data collection.

Ideally, inception and evaluation reports should describe data collection and analysis methods

in sufficient detail so that readers can obtain an accurate picture of the robustness of data and the quality of evidence produced. A data collection and analysis plan could make it clear why and how certain methods are used in the evaluation. Evaluation teams should be encouraged to include a transparent reflection on the quality of evidence and its implications on the evaluation in their reports.

The description of a sample chosen for a survey should characterise and contextualise the sample, so that evaluation users understand the extent to which it matches the purpose of the evaluation (acknowledging that representativeness can seldom be achieved). Likewise, the inclusion and exclusion criteria used when selecting case studies should be made explicit. For example, an example might identify a typical case for illustration, or document particularly successful practice, or investigate examples which don't fit the overall pattern and try to understand why.<sup>v</sup> Moreover, to facilitate effective management of the evaluation process, the evaluation team should set out an auditable trail of its intentions in terms of data collection, analysis and reporting.



WFP/Nick Sells

# 5. WFP evaluation guidance

To what extent should WFP evaluation guidance be strengthened and enhanced to help improve evaluation design, while encouraging a flexible and adaptive approach to increase innovation/creativity, rather than compliance?

## CURRENT PRACTICE AND CHALLENGES

WFP EQAS evaluation guidance ensures all evaluations cover a comprehensive range of evaluation criteria and questions. A degree of standardization can support synthesis and, in recent years, OEV has undertaken synthesis studies in a number of areas across CSPEs, DEs and PEs. However, for such standardisation to contribute to good evaluation quality, the central elements of ToR need to be tailored to the objective, subject, and context of each evaluation. EMAP reviewers advocate for more focus and flexibility while maintaining a degree of standardisation and strengthening guidance on key principles, such as transparency about the quality of evidence.

All evaluations have guidance through the Centralized and Decentralized Evaluation quality assurance systems as appropriate (CEQAS and DEQAS). Guiding documents come in the form of a step-by-step Process Guide, templates, checklists and Technical Notes (TNs), and in the interaction between evaluation managers (EMs) and evaluation teams.

The guidance documents define WFP's formal expectations from evaluations in some detail. This has resulted in a relatively uniform, comprehensive report structure, which touches on a predetermined catalogue of items. As a result, evaluation teams may need to make a special effort to discuss unanticipated questions and wider, systemic issues. The practice of covering all OECD-DAC criteria might distract attention away from other important issues, such as WFP's strategic positioning.

Broader written guidance on principles that need to be observed for appropriate evidence quality and utility appears to be missing from the comprehensive set of guidance documents or is difficult to extract.

While the written guidance has a similar set of components across the range of centralized evaluations, it may not reflect the diversity of the types of evaluation. In any one year, there will be considerably more CSPEs being undertaken (30 completed or ongoing in 2021 versus 3 CEEs, 3 PEs and 3 SEs).<sup>vi</sup> This production line of CSPEs may require a different approach to guidance compared to other types of evaluations, such as SEs where there is much more flexibility to choose the scope, objectives and approach.

The EMAP reviewers do not advocate for throwing out the proverbial baby with the bathwater. Templates and checklists can be helpful. But it is unclear to what extent they are used and understood by their users. Also, it is important to define the key users of WFP guidance. Written guidance may be extremely helpful for evaluation managers, but it cannot be expected to make up for major gaps in an evaluation team's abilities.<sup>8</sup>

Clearly OEV guidance goes beyond the documents that form part of the EQAS but also include the engagement of the OEV evaluation managers with the evaluation teams. Teams relying only on written guidance may comply through using buzzwords rather than exploring the different elements of the evaluation design and the interactions between them. The evaluation manager should set the tone for the



evaluation from the start and there should be important points in the evaluation process where **evaluation managers can work closely with the teams in a partnership** and guide them as necessary. For instance, the end of data collection and the start of the analysis phase is a point where evaluation teams and EMs can come together to determine whether the data collected are likely to meet both methodological and WFP's information needs. Any process should also take into account the fact that managers of decentralized evaluations in country offices are often extremely busy with other tasks.

### **GOOD PRACTICES TO BE LEVERAGED**

The Lebanon CSPE includes a reflection of experiences and lessons for future evaluations. The checklist for evaluation reports suggests that "Where appropriate, it should provide reflection on experience and lessons for future evaluation". OEV would benefit from all

evaluations giving such feedback although it may be more critical, and therefore useful, if not included in the report itself.

### **AVENUES FOR EXPLORATION**

It would be interesting to assess to what extent and how evaluation teams have used WFP guidance, including both written guidance and engagement with WFP evaluation managers and other staff. That would help determining where more resolute guidance is needed – for example, for clarity on the quality of evidence –, and where existing guidance documents should be simplified, adapted, or removed from the guidance package.vii In addition to discussing this topic in OEV's meeting with long term agreement (LTA) companies, OEV should continue to carry out regular focus group discussions or workshops with evaluators who have hands-on experience in applying WFP guidance.viii



WFP/Annabel Symington

# 6. Lessons to strengthen WFP's evaluation function

This section presents recommendations that have emerged across the EMAP reviews and in discussions with EMAP members, OEV and Regional Evaluation Units.

## WHERE SHOULD WFP GUIDANCE BE STRENGTHENED?

- Discussion of equity as a key aspect of effectiveness – which will also mean giving more room to the voices of intended beneficiaries in evaluation. Current WFP work on defining intersectionality and intersectional analysis, and on looking at inclusion more broadly is likely to support this.
- Evaluators to be encouraged to tailor evaluations to their context, i.e., to propose reduction, focusing, and prioritisation of questions – less is more!
- Evaluators to involve more external sources in data collection during inception and evaluation phases, including critical voices, to strengthen triangulation and context sensitivity
- Written guidance to be streamlined and complemented with regular experience exchange between evaluation managers to enhance consistency of approaches in evaluation management
- The evaluation inception report to include clear linkages between the different elements and a tentative data collection and analysis plan so that there is an overall logic to a design that draws on all of them.

Furthermore, it could also be helpful to conduct an annual review with the users of evaluations at various levels to find out how they have used evaluations, what they have appreciated in evaluations (and why), and what they found

lacking, useless or unhelpful (and why). This would be an excellent commitment to learning from evaluation; it does not have to be onerous.<sup>ix</sup>

## WHAT NEEDS TO (CONTINUE TO) BE STANDARDISED?

There is a need to find a balance between a degree of standardisation and tailoring evaluations to their context (which, in its broader definition, includes the evaluand, its broader context, as well as intended use of the evaluation).

- Evaluability assessments more focused on supporting evaluation design including stakeholder analysis; ToC review should be undertaken with the ToR, if possible, and supplemented in the IR as necessary
- More specific guidance on transparency in describing (i) data collection and analysis processes, (ii) the rationale behind the choice and mix of methods, and (iii) the quality of evidence
- Definition of WFP expectations from theory-based evaluation and ToC (re-) development

## AREAS TO DO LESS OF

EMAP advisors have found some areas where less detailed guidance and less standardisation could contribute to enhancing evaluation quality, in that evaluation teams would be encouraged to tailor evaluations to their purpose and context.

- Coverage of all OECD criteria in all evaluations: for greater learning benefits, it may be useful to focus evaluations on a sub-set of evaluation criteria depending on the context-specific evaluation questions – especially where some of the OECD-DAC criteria have been addressed in other relevant evaluations.
- Fine-grained standardisation across evaluations (e.g., detailed report templates), as no amount of standardisation can replace an evaluation team with the appropriate capacities. If certain sets of information are needed for syntheses of evaluations, a separate table – with definitions of all key concepts – could be used to compile such information.
- The volume of guidance documentation would benefit from some shrinking and focusing. Shrinking could consist in making it more accessible by restructuring it into a single, short mandatory document with expandable sections. Focusing can also imply slightly expanding or reorienting key sections, for example on evaluability assessment and on using a theory of change. This may need to include shortening and sharpening technical notes (TN). Focus could be facilitated by targeting guidance to specific users and by recognizing that some guidance is essential in an evaluation process and some is for reference only.

WFP carries out large numbers of evaluations, some of which are routine requirements (e.g., CSPEs). For the sake of efficiency and learning, it may be useful to reduce the number of routine evaluations, and/or focus on specific learning questions. Alternatively, some evaluations could be framed as some different type of evaluation which is faster, more reflective and more action oriented



WFP/Sayed Asif Mahmud

## Annex 1: Short biographies of members of the EMAP

<p><b>Khalil Bitar</b></p>	<ul style="list-style-type: none"> <li>• +13 years of experience</li> <li>• Specialized in evaluation in countries affected by fragility, conflict, and violence; equity and social justice issues in evaluation; youth empowerment evaluation capacity strengthening, and transformative evaluation practices</li> </ul>
<p><b>Paul Knox Clarke</b></p>	<ul style="list-style-type: none"> <li>• +25 years of experience</li> <li>• Extensive experience in strategy development, organizational structures and the international humanitarian sector</li> </ul>
<p><b>Rick Davies</b></p>	<ul style="list-style-type: none"> <li>• +30 years of experience</li> <li>• Specific expertise on evaluation participatory approaches, social network analysis, theories of change, qualitative comparative analysis, evaluability assessments, most significant change</li> </ul>
<p><b>Rajeev Pillay</b></p>	<ul style="list-style-type: none"> <li>• +30 years' experience</li> <li>• Specific expertise on economic transformation, institutional change and capacity development, preventive development, peacebuilding, stabilisation and early warning</li> </ul>
<p><b>Michaela Raab</b></p>	<ul style="list-style-type: none"> <li>• +30 years of experience</li> <li>• Specific expertise on gender-responsive evaluation, theory-based approaches in evaluation of complex interventions (e.g., on strengthening civil society), qualitative comparative analysis and strategy development</li> </ul>
<p><b>Michael Reynolds</b></p>	<ul style="list-style-type: none"> <li>• +25 years of experience</li> <li>• Strong experience with managing and conducting country programme evaluations and strategic evaluations</li> </ul>
<p><b>Patricia Rogers</b></p>	<ul style="list-style-type: none"> <li>• +30 years of experience</li> <li>• Expert in both quantitative and qualitative data for evaluation (e.g., cost benefit analysis, appreciate enquiry), and theory-based evaluation</li> </ul>

## Annex 2: Evaluation Documents Reviewed by the EMAP

Name of evaluation	Report	Reviewer
Myanmar DE	Evaluation report	Khalil Bitar
Libya DE	Evaluation report	
Palestine DE	Evaluation report	
Ghana DE	Evaluation report	
South Sudan DE	Evaluation report	
HQ-led Gender DE	Evaluation report	
HQ-led School Feeding DE in Lebanon	Evaluation report	
Burkina Faso DE	Evaluation report	
Madagascar DE	Evaluation report	
RBJ thematic DE on market development	Evaluation report	
RBN thematic DE on partnerships	Evaluation report	
Dominican Republic DE	Evaluation report	
El Salvador DE	Evaluation report	
Sri Lanka DE	Evaluation report	
CSPE China	Evaluation report	Rajeev Pillay
CSPE El Salvador	Evaluation report	
CSPE Honduras	Evaluation report	
CSPE El Salvador	Evaluation report	
CSPE Bangladesh	Evaluation report	Mike Reynolds
CSPE Lao PDR	Evaluation report	
CSPE Lebanon	Evaluation report	
CSPE Zimbabwe	Evaluation report	
PE Climate Change	Terms of reference	Rick Davies
	Inception report	
PE Resilience	Inception report	
PE CSPs	Inception report	Mike Reynolds
CEE Myanmar	Concept note	Paul Knox Clarke
	Terms of reference	
CEE Sahel Region	Concept note	Patricia Rogers
SE Nutrition and HIV/AIDS	Inception report	Patricia Rogers
	Evaluation report	
SE on PSEA	Concept Note	Michaela Raab

## Annex 3: Selection of evaluations for review by the EMAP

Type of evaluation	Number of evaluations reviewed	Selection criteria
Decentralized evaluations	<b>14</b>	a sample of 1/3 of the reports, ensuring equal regional representation and based on the rationale for selection provided by the regional evaluation unit to focus on those with specific methodological aspects
Country strategic plan evaluations	<b>8</b>	all the 2020/2021 CSPEs completed by January 2022
Policy evaluations	<b>3</b>	all evaluations started in 2022 (not those started in 2021 and ongoing in 2022)
Corporate emergency evaluations	<b>2</b>	all evaluations started in 2022 (not those started in 2021 and ongoing in 2022)
Strategic evaluations	<b>1</b>	all evaluations started in 2022 (not those started in 2021 and ongoing in 2022)

## Annex 4: Guidance on development of evaluation questions by type of evaluation

(taken from ToR templates)

DE	CSPE	PE	CEE	SE
<b>Fully flexible</b>	<b>Common EQs</b>	<b>Common EQs</b>	<b>Flexible</b>	<b>Fully flexible</b>
<p>List the main evaluation questions, plus any other questions, as appropriate.</p> <p>Questions should be clear, sufficiently address the selected evaluation criteria and be relevant to the subject, purpose and intended use of the evaluation. They should be adequate and realistic in view of the time and resource available.</p> <p>Gender equality and women's empowerment should be mainstreamed throughout the evaluation questions and subquestions with consideration of how the perspectives of men, women, boys and girls will be sought in the evaluation process. Data requires disaggregation by gender. Similarly, equity and wider inclusion dimensions should be considered.</p>	<p>The evaluation will address four main questions common to all WFP CSPEs.</p> <p>EQ 1: To what extent is the CSP evidence based and strategically focused to address the needs of the most vulnerable?</p> <p>EQ 2: What is the extent and quality of WFP's specific contribution to CSP strategic outcomes and the UNSDCF in country name?</p> <p>EQ 3: To what extent has WFP used its resources efficiently in contributing to CSP outputs and strategic outcomes?</p> <p>EQ 4: What are the factors that explain WFP performance and the extent to which it has made the strategic shift expected by the CSP?</p>	<p>The following questions are to be used in all policy evaluations:</p> <p>EQ 1: How good was the policy (relevance)?</p> <p>EQ 2: What results has the policy achieved (effectiveness)?</p> <p>EQ 3: What were the enabling or hindering factors for results achievement (coherence, efficiency, sustainability)?</p>	<p>The evaluation will develop specific evaluation questions that are appropriate to the emergency response and context being evaluated. As a guide, four main questions may be considered (see below). The evaluation team may further develop and tailor sub questions as relevant and appropriate to the WFP Emergency Response and country context, including as relates to assessing the response to the COVID crisis. Sub-questions indicated below are just samples, which needs customisation.</p> <p>EQ1 – To what extent is WFP emergency response strategically designed to meet national and people's humanitarian needs?</p> <p>EQ2 – What is the extent and quality of WFP's performance in emergency response in Country X (or region or name of the emergency response)?</p> <p>EQ3 – To what extent has WFPs used its resources efficiently in the emergency response?</p> <p>EQ4 – What are the factors that explain WFP emergency response performance?</p>	<p>The evaluation questions and sub-questions should:</p> <ul style="list-style-type: none"> <li>• Be presented with clear links to the purpose and objectives, as well as the intended use of the evaluation.</li> <li>• Sufficiently address OECD-DAC evaluation criteria</li> <li>• Be adequate and realistic in line with the time and resources available</li> <li>• Be developed while applying a gender, equity and inclusion lens, by considering it within each of the criteria and in the evaluation questions.</li> </ul>

# Endnotes

---

<sup>i</sup> Syntheses of evaluations and system-wide evaluations are also not covered

<sup>ii</sup> <https://www.betterevaluation.org/en/plan/approach/QUIP>

<sup>iii</sup> <https://mande.co.uk/special-issues/evaluation-questions-checklists/>

<sup>iv</sup> For example, the OECD/DAC publication [Applying Evaluation Criteria Thoughtfully](#) looks at how to include equity in evaluation criteria.

<sup>v</sup> A Michael Patton video describes useful strategies for purposeful sampling:  
[https://www.betterevaluation.org/en/rainbow\\_framework/describe/sample](https://www.betterevaluation.org/en/rainbow_framework/describe/sample) and  
<https://www.youtube.com/watch?v=lnGvg1vk7IM>

<sup>vi</sup> WFP Annual Evaluation Report for 2021

<sup>vii</sup> The EMAP reviews have been based chiefly on documents – evaluation concepts, ToRs, inception reports, evaluation

reports, WFP evaluation guidance documents, plus some discussions with members of WFP's evaluation function. There has been no data collection with evaluation teams or participants in any specific evaluation.

<sup>viii</sup> This was done with evaluation team leaders who worked on the first batch of CSPEs and it was reportedly very helpful, especially informing the thinking on evaluation of efficiency in CSPEs.

<sup>ix</sup> OEV has planned to pilot this approach in November 2022.



**Office of Evaluation**

**World Food Programme**

Via Cesare Giulio Viola 68/70

00148 Rome, Italy

T +39 06 65131 [wfp.org](http://wfp.org)