Food and Agriculture Organization of the United Nations

OED GUIDANCE NOTE ON EVALUATION SYNTHESES

FAO Office of Evaluation (OED)

Seda Kojoyan, Renate Roels, Paul Neuman, Carolina Turano
Under the leadership of: **Rachel Sauvinet Bedouin**

October 2021

1. Introduction

The purpose of this guidance note is to support colleagues at the Office of Evaluation (OED) of the Food and Agriculture Organization of the United Nations (FAO) with respect to the planning, formulation and conduct of evaluation syntheses. It could also be used as a reference document for FAO project personnel or partners, to better understand how an OED-led evaluation synthesis is conducted.

This note has been prepared by the OED evaluation synthesis working group and has benefitted from the OED lessons learned paper on regional evaluations syntheses (2020), as well as a literature review on the common practices in other international, bilateral, and multilateral development organizations.

This guidance note should be considered as a living document and could be amended and enriched as we enhance our experience on evaluation syntheses, or as new needs arise.

Following this introduction, section 2 briefly provides background, while section 3 discusses key principles for the conduct of an evaluation synthesis, including commonly used methods and suggestions on how to overcome certain limitations. Section 4 discusses how to implement evaluation syntheses within OED. An overview of useful resources is included in Appendix 1.

2. Background

2.1 Rationale and objective for conducting evaluation syntheses

In general, OED evaluations serve a twin purpose of accountability and learning:

i.   Accountability to FAO member countries, providing them with an assessment of the work of the Organization against its strategic objectives, as well as the organizational performance.

ii. Learning, to make informed decisions for enhanced FAO strategies, programmes and projects, and contribute to knowledge.

Through the conduct of evaluation syntheses, OED would like to efficiently use existing evaluation evidence from OED evaluations and build a body of knowledge for enhanced corporate learning. As this is an expanding area of work for OED, this guidance note aims to provide tips and insights from experience as well as to facilitate harmonized approaches.

2.2 Type of evaluation syntheses already conducted in OED

As part of the background research carried out for this guidance note, a review of past syntheses[1] conducted in OED has been done, in order to identify the main trends emerging from OED's past work on syntheses.

---

1   Going back to 2015.

The majority of evaluation syntheses conducted in OED are the result of specific demands coming from the Governing Bodies or meant to contribute towards thematic or larger evaluations. Additional synthesis work has also been conducted for other purposes, such as internal lessons sharing, e.g., the Synthesis of lessons from the Global Environment Facility (GEF) project evaluations. Within the timeframe examined, only one synthesis was jointly conducted with another agency, i.e., the International Fund for Agricultural Development (IFAD).

Since 2015, more than eight syntheses have been prepared to respond to specific requests from the Governing Bodies, such as the five Regional Syntheses (2020), requested by the Programme Committee (PC) to be presented during the Regional Conferences;[2] the Synthesis of evaluations of FAO's Regional and Sub-regional Offices (2015); and the Synthesis of lessons learned in the application of the Country Programming Framework (2018), both requested by the PC.

As mentioned, syntheses are also conducted as part of broader thematic evaluations, both as stand-alone products or as internal analysis carried out to feed into the main report. Examples of these are FAO's work on farmer fields schools, as part of the Sustainable Development Goal (SDG) 2 Evaluation, and the Synthesis of results from GEF project evaluations, in the context of the SDG 13 evaluation. An internal paper drawing lessons learned from the conduct of regional evaluation syntheses in 2020 was also developed and has greatly contributed towards this guidance note. In the table below, a comprehensive list of evaluation syntheses conducted in OED since 2015 can be found.

Table 1. List of evaluation syntheses conducted in OED since 2015	
Name of evaluation	**Year**
Peace Building Fund synthesis concept note	on-going
[Synthesis] FAO work on farmer field schools (part of the SDG 2 evaluation)	2021
Evidence summary on COVID-19 and food security (UNEG product, OED co-managed)	2021
Synthesis of results from GEF project evaluations (part of the SDG 13 evaluation)	2020
Regional evaluation syntheses for the Africa (RAF), Latin America and the Caribbean (RLC), Near East and North Africa (RNE), Europe and Central Asia (REU) and Asia and the Pacific (RAP) FAO regional offices	2020
Synthesis of FAO-GEF project evaluations - implementation issues	2020
Synthesis of FAO-GEF project evaluations - policy influence & governance support	2020
Synthesis of findings and lessons learnt from the Strategic Objective evaluations	2019
Synthesis of biodiversity-related project evaluations in Ecuador	2018
Synthesis of lessons learnt in the application of the Country Programming Framework	2018
FAO and IFAD's engagement in pastoral development -joint evaluation synthesis	2016
Synthesis of the evaluations of FAO's regional and sub-regional offices	2015

Source: Developed by the Evaluation Synthesis Working Group.

The review shows that, although most evaluation syntheses consisted of desk reviews, as well as some additional interviews, overall, there is no harmonized approach, nor a shared understanding at OED of what constitutes an evaluation synthesis. Furthermore, methodological approaches are diverse, and there is a lack of a consistent approach used in sampling past evaluations. Finally, about half of evaluation syntheses included recommendations, and only three syntheses utilized qualitative software/computer-assisted approaches in the analysis.

---

2   Due to the COVID-19 pandemic, only the evaluation synthesis for Europe and Central Asia (REU) was presented and discussed during the Regional Conference for Europe.

2.3 Definitions and typology of evaluation syntheses

It is important to keep in mind that various forms of evidence can be synthesized in addition to evaluations. At the same time, it is likely that at OED most of the syntheses that will continue to be performed will be evaluation syntheses. There is not one specific definition for various types of evaluation synthesis. Each organization tailors them to their own needs and policy. An evaluation synthesis can be described as a whole in the following way:

Evaluation synthesis captures evaluative knowledge and lessons learned on a certain topic from a variety of existing evaluations through aggregated and distilled evidence in order to draw more informed conclusions (and sometimes recommendations) on a specific topic or question (The Presidency of the Republic of South Africa, 2014).

The following table clarifies the main differences in wider evidence synthesis, as well as provides links to further information. It distinguishes between the types of synthesis which can be appropriate to conduct, depending whether less or more time and resources are available. The former requires a more systematic approach, and often, an evaluation framework. Unsurprisingly, the degree of confidence grows with the more time- and resource-intensive syntheses.

It should be noted that Table 2 below is not exhaustive. It is prudent to keep in mind, as described by Barnett-Page and Thomas (2009), that "profusion of terms can mask some of the basic similarities in approach that the different methods share, and also lead to some confusion regarding which method is most appropriate in a given situation".[3]

Table 2. Typology of evidence syntheses				
	Name and definition/purpose		**Methodology**	**More information**
More labor and time-intensive; appropriate when in-depth analysis is needed, data is abundant (even if of varying quality), or mostly quantitative.	**Systematic reviews** – *look systematically across the evidence base. Explicit and transparent procedures. Clear inclusion and exclusion criteria.*			
	Realist synthesis	Synthesizing all relevant existing research in order to make evidence-based policy recommendations. Focus is on understanding the mechanisms by which an intervention works or does not.	Identifying review questions. Searching for primary studies. Quality assessment. Extracting the data. Synthesizing the data. Dissemination.	https://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-7-33
	Meta-evaluation	An evaluation of evaluations, aggregates findings from a series of evaluations.	Relies on an evaluation framework. Does not use complex statistical analysis but can include an evaluation of the quality of the evaluations and its adherence to established good practice in evaluation.	https://wmich.edu/sites/default/files/attachments/u58/2015/Quality_Context_and_Use.pdf https://www.betterevaluation.org/en/rainbow_framework/manage/review_evaluation_do_meta_evaluation https://www.alnap.org/help-library/evaluating-evaluations-a-meta-evaluation-checklist
	Meta-analysis	Using statistical tools to analyze data (quantitative), from different evaluations, i.e. combining studies.	Combines numeric evidence from experimental (and sometimes quasi-experimental studies) to produce a weighted average effect size.	To be used only where appropriate (i.e. impact evaluations). https://www.betterevaluation.org/evaluation-options/meta-analysis

---

3   We recommend this reading to those who wish to see a more comprehensive list of the different typologies for evaluation syntheses.

	Name and definition/purpose		Methodology	More information
Less labor and time-intensive; appropriate for a quick assessment of limited data.	**Lessons learned**	Lessons learned can develop out of the evaluation process as evaluators reflect on their experiences in undertaking the evaluation. Lessons can take the form of describing what should or should not be done or describing the outcome of different processes.	Combining lessons learned from a pool of evaluations, mindful of the context and further applicability. A greater number of supporting evidence from various evaluations for the lessons learned and their triangulation increases confidence of obtaining a significant and meaningful "lesson".	https://www.betterevaluation.org/evaluation-options/lessons_learnt https://archive.globalfrp.org/evaluation/the-evaluation-exchange/issue-archive/learning-organizations/lessons-learned https://journals.sagepub.com/doi/abs/10.1177/109821400102200307
	Rapid evidence assessment	Systematic search, but a quick overview of existing research on a topic.	Often a combination of key informant interviews and targeted literature searches.	https://www.betterevaluation.org/evaluation-options/rapid_evidence_assessment
	Textual narrative	Dividing the studies into relatively homogenous groups, reporting study characteristics within each group, and articulating broader similarities and differences among the groups. Useful in synthesizing evidence of different types (qualitative, quantitative, economic, etc.).	Typically, study characteristics, context, quality and findings are reported on according to a standard format and similarities and differences are compared across studies. Structured summaries can also be developed.	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3224695/ https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1783856/
	Vote counting	Gaining knowledge about a body of evidence simply by tallying the number of positive results to the negative.	Comparing the number of positive studies (i.e. studies showing benefits) with the number of negative studies (i.e. studies showing harm). Does not take into account quality of evidence, sample or effect sizes.	https://www.betterevaluation.org/evaluation-options/votecounting
Related, but with a different purpose	**Evidence gap mapping**	Maps out the existing evidence and gaps in order to set future research priorities.	While steps in the process are similar to the above, major difference is that these mappings do not synthesize what the evidence actually says.	https://www.campbellcollaboration.org/evidence-gap-maps.html https://www.3ieimpact.org/evidence-hub/evidence-gap-maps

Source: Adapted from Better Evaluation's "Synthesise Data Across Evaluations", the paper "Methods for the synthesis of qualitative research: a critical review" as well combined and counter-validated definitions from the 3ie; OECD; and the government of the Republic of South Africa.

2.4 Advantages of evaluation syntheses

The following represent some of the major advantages of conducting evaluation syntheses. The limitations of evaluation syntheses, with corresponding mitigation measures, are discussed in the Section 3 of this document.

i.   further strengthen the use of evaluation findings;

ii.   facilitate wider use and accessibility of evaluations, by compressing findings of many evaluations into one report;

iii. promote learning and a deeper reflection/understanding of a particular topic and the strategic implications of certain findings;

iv.   contribute to decision-making processes in an effective and efficient way (e.g., when there is no time or resources to do a new evaluation);

v.   cost-saving, since evaluation syntheses mostly uses existing knowledge opposed to starting from scratch;

vi.   methodological: evidence could become more reliable, with a higher degree of confidence in the evaluation results, if it is subjected to a critical appraisal during the synthesis process;

vii. realist synthesis: since the focus is on program theory, it may be more useful for high-level decision-making; additionally, complex programmes consist of a number of linkages, each with a varied evidence base; realist synthesis can help with such knowledge; and

viii. focus on structural and common factors, rather than specific interventions and their unique characteristics.

3. Principles

3.1 When to conduct evaluation syntheses

Evaluation syntheses essentially support the same objectives of accountability and learning associated with all evaluations as well as serve to promote enhanced FAO strategies, programmes and projects. There are contexts in which evaluation syntheses may be preferable to the conduct of evaluations drawing upon primary data collection and analysis. While any decision about whether to conduct an evaluation synthesis depends upon the evaluand and other context-specific factors, there are certain times in which an evaluation synthesis may be called for. These include:

i.   when there are a number of evaluations that provide evidence on a given topic;

ii. to contribute effectively and efficiently to decision-making processes and frame discussions, particularly when there is neither adequate time (i.e. to respond to urgent requests), nor resources to undertake a full-fledged evaluation;

iii. to feed into a thematic evaluation, e.g., on the SGDs – in other words, as part of the evaluation process;

iv. for selected events or processes, e.g., regional conferences (for Governing Bodies or Member States), to comply with donor requests or commitments and on selected operational issues (Operational Partners Implementation Modality, for example);

v.   when there are a number of evaluations that provide evidence on a given topic, particularly in cases when evidence would lead to uncertain conclusions or when some of the information is conflicting;[4]

---

4   The value-added and utility of conducting an evaluation synthesis should be considered, even when ample evidence is available for review. Such a decision should take into account, among others, longevity of its use, impact on visibility, accountability or learning from evaluations. For more guidance see section 4.1.

vi.   to save on direct and indirect costs (both for the evaluator and for the groups being evaluated);[5]

vii. to focus on structural/common factors associated with positive outcomes or challenges, and not only on issues specific to a given intervention or context;

viii. to improve the robustness of evidence by building on work that has been done, instead of starting from scratch; and

ix.   to enhance the applicability of existing findings and developing new knowledge through the integration process.

3.2 Commonly used approaches and methods

The approach for conducting a systematized evaluation synthesis can be divided in the following eight phases:

i.   involvement of key stakeholders from day one: determine audience needs, interests and product preferences (see section 2.2 and 3.1);

i.   selection of a topic and the conduct of an evaluability assessment (see section 4.2 for more information);

ii. preparation of overall terms of reference (TORs) and where necessary complemented by additional concept notes;[6] TORs should contain:

• rationale for conducting the evaluation synthesis and selected topic (see also section 3.1);

• scope, purpose, audience and use of the evaluation synthesis;

• well-defined evaluation questions, criteria for theme selection and detailed methods/tools; and

• risks, possible limitations and timeline.

iii. identification of the sample size and review protocols;

iv. screening, codifying and extracting information/data for analysis;[7]

v.   analysis, validation and triangulation of collected information/data;

vi. preparation and review of the report; and

vii. dissemination on the OED webpage and, if requested, presentation of the findings to the PC and other internal/external stakeholders.

An evaluation synthesis usually relies on multiple sources for data collection and mixed methods for the analysis, validation and triangulation of evidence against the key questions. Generally, a qualitative approach is used to synthesize findings from completed evaluation reports. Where relevant, and where time and resources allow, the scope could be expanded (e.g. with other FAO policies or topic-related external sources). Interviews or quantitative approaches may also be applied.

Organization of a systematic desk-review. Once the topic is selected, the organization of a systematic desk review starts. One of the first steps includes the identification of the scope (representative sample size) by screening a large number of evaluation reports and by filtering, coding and organizing the information/data. The evaluation synthesis should use purposive

---

5   This alone is not a sufficient condition to conduct evaluation syntheses, however.
6   For example, while the TORs of the OED Regional Synthesis provide generic guidance, specific concept notes were prepared for each of the five regions highlighting the specific approach, scope and objective, etc. For this exercise each region had a certain degree of flexibility in terms of methodological approach and tools to use, and the report for MENA region, for example, decided to consult the Country Programming Frameworks as well, beyond evaluation reports.
7   This may involve some training of the evaluation team, especially if (new) software is used.

sampling: selecting reports that are most likely to provide the most useful information to answer the evaluation questions. For example, what reports (and individuals for the interviews later) will we learn from the most?

An evaluation of evaluations should be a pre-requisite for any synthesis (assessing the actual quality of the universe of evaluations to be synthesized). For this exercise, it is key to develop well-defined and demarcated inclusion/exclusion quality criteria. Many evaluation offices, among those of the World Food Programme (WFP) and IFAD, use a structured analytical framework, where evidence is rated for quality on a scale, and only evaluations above a certain threshold are included in the synthesis. Currently, OED is also working on a quality assurance system.

The evaluation guidelines of the department for performance monitoring and evaluation (DPME) of the Government of South Africa and the second evaluation manual of the independent office of evaluation of IFAD (IOE) both state that it could be useful to develop a theory of change before starting the screening. This should help in the identification of the precise scope of the evaluation synthesis and assist with the searching for themes/codes that meet this scope. A theory of change will also help in structuring the analysis of the evidence and the presentation of findings.

Screening (and finding) a large number of relevant reports can be overwhelming. During a first round, the sample size of eligible reports and initial themes/coding is defined. The identification of themes should be conducted according to the established criteria in the TORs (e.g. food security, gender, monitoring and evaluation, etc.). The majority of the coding takes place once the sample size is defined. It often happens that after reading/coding a couple of reports, a more refined list of codes emerges.

To facilitate data extraction based on codes and keywords of the analytical framework, qualitative analysis software could be used, such as MAXQDA or NVIVO. Another option, used for the OED regional synthesis reports, is the add-in for Word, DocTools ExtractData.[8] This is a free, simple tool that can extract (isolate) Word's Review/New Comment entries from a document. Whenever software tools are not available, a simple-word search or strategic reading (i.e. looking at the frequency of agreed words and close alternative words and/or reading an agreed number of sections such as inter alia the executive summary, conclusions and recommendations) can be applied. With or without software, both approaches need time: while the simple-word search is per definition time-consuming, for the use of software one needs to calculate extra training time to prepare and refine the codes.

To synthesize the results, a detailed recording grid (e.g. in Excel) should be used. To increase reliability, it is good practice to have at least two persons coding the same evaluation report (to control for bias), before summarizing the consolidated evidence base. In case of software, the codes need to be reviewed before using the software programme. It is crucial that all team members follow the same review protocols and that the process is systematic. The rigour of the process is crucial for the credibility and quality of the evaluation synthesis. Full-time supervision of the evaluation team that implement the review protocols is advised. This role often involves quality checking, ensuring consistency using regular spot-checks, distributing evaluations to team members, and overseeing the time taken to complete protocol tasks per report. The larger the scope, sample and the more complex the screening process, the longer this exercise will take.

A lesson learned from the OED Regional Synthesis is to include reference data in the recording grid (e.g. evaluation code, page number), which allows for quick searches when extra information is required or when concerns are raised on the evidence. Once the recording grid is ready, it is time to clean, screen, organize and qualify the data, taking out any evidence that is not well substantiated or not useful.

---

8   Available at https://www.thedoctools.com/word-macros-tips/word-macros/extract-comments-to-new-document/.

The final step is analysing the evidence collected. Once commonalities between extracts have been established, these can be synthesized into findings. Depending on the scope and utility of the evaluation synthesis, examples of extracts can be included under each evaluation question for illustrative purposes.

Interviews and briefings. The desk-review exercise is supplemented and triangulated by selected face-to-face or virtual interviews and/or focus-group discussions with internal and external informants at different levels to better contextualize, enrich and verify certain findings and to collect updates on findings from completed evaluation reports (ensuring that the evaluation synthesis findings are not outdated). Naturally, findings are also further validated through one or more (de)briefing sessions.

Additional methods and analyses. Desk research of evaluation synthesis practices in a selection of other evaluation offices gathered a few examples of additional methods that could be used (in addition to the desk-review of completed in-house evaluation reports and key informant interviews). For example, the independent evaluation offices of the Asian Development Bank (ADB), the African Development Bank (AfDB) and IFAD tend to include also secondary data, and the United States Agency for International Development (USAID) and AfDB also occasionally used an online survey. The independent evaluation offices of IFAD and ADB also use, in some larger evaluation syntheses exercises, a section with best practices or lessons from other development organizations. A quick overview of additional methods is summed up below:

i.   **additional desk reviews:** to expand evidence with, for example, related corporate documents, external sources on the topic, or project documentation (e.g. project document, progress reports, etc.);

ii. **electronic surveys**: before or during the data–collection, to assess priorities or early lessons learnt;

iii. **country visits[9]/field-level key informant interviews and focus group discussions:** to ensure new/updated info is included;

iv. **meta-analysis**: a statistical method for combining numeric evidence to produce a weighted average effect size; it could be used to compare alternative interventions or compare an intervention to a control group; it is also used to illustrate correlations and trends;

v.   the inclusion of annexes with **case-studies**: e.g. to illustrate performance of specific projects;

vi. **benchmarking exercises**: e.g. to illustrate lessons and best practises of similar multilateral/bilateral development organizations; and

vii. **outlier-analysis**: e.g. to illustrate what the assessed excelling or underperforming projects have in common.

3.3 How to address common limitations

In previous sections, we discussed the advantages of conducting an evaluation synthesis, as well the common methodology steps for a systematized approach. In reviewing a number of syntheses evaluations, certain common limitations emerged as well. In Table 3 below, the Working Group made to match these with potential remedies for each.

---

9   If no country visits are possible, the team working on an evaluation synthesis could ask other OED colleagues going on a mission to ask some evaluation questions on their behalf.

Table 3. Limitations and remedy measures in evaluation syntheses	
Limitation	**Remedy**
Synthesis for the "sake of it", questionable utility in synthesizing recent evidence.	Engage the target audience in the design of the synthesis, to obtain their views on the relevance and potential use of the synthesis.
Evaluation evidence may have become outdated by the time of the synthesis evaluation.	Identify key evidence which is no longer valid. Validate and/or complement data through the conduct of additional interviews or gathering new data. Triangulate data through secondary sources, consulting documents beyond evaluation reports.
Variability in the quality of evaluation reports.	Conduct a quality assessment of the evaluation data to decide which data can be used in the synthesis. Additionally, and more broadly, have clear inclusion/exclusion criteria.
Variability of the level of detail in evaluation reports. When using different type of evaluations (I.e., Project, Thematic, Country Programme Evaluations), the depth in which certain topics are covered in the reports varies, which makes it difficult to compare and present quantitative data on progress and achievements on certain topics. Lack of systematic coverage of organization-wide issues. OED's evaluations do not cover all FAO's work; therefore, synthesis based only on evaluation reports do not provide a comprehensive analysis of FAO's work on a certain area or topic.	Consult secondary sources, such as FAO internal studies, and gather new data. Build upon existing OED evaluation syntheses on the same or related topics.[10] Consider tapping into existing systematic and realist syntheses, from outside sources, to capture evidence on the topic of an up-and-coming evaluation synthesis. For example, evaluation syntheses already exist on the topic of cash transfers.
Lack of a systematized evaluation reports database makes it difficult to identify and access evidence.	Build a robust database of evaluation reports, with tagging. This can add greatly in obtaining, relatively quickly, a picture of the available evidence for a given evaluation synthesis.
Recommendations provided by the evaluations may have already been taken into account or become "invalid" for other reasons.	Engage the target audience in the design of the synthesis, to obtain their views on the relevance and potential use of the synthesis.

Source: Analysis by the Evaluation Synthesis Working Group. [10]

4. Implementation of OED-led evaluation syntheses

4.1 Criteria for conducting evaluation syntheses

Evaluability assessment. Before deciding on whether to prepare an evaluation synthesis report on a given topic, an evaluability assessment should be carried out, which entails "the extent to which an activity or project can be evaluated in a reliable and credible fashion" (OECD-DAC 2010; p.21). As a minimum, one should check:

i. the desired **utility** (audience needs, interests and product preferences; longevity of its use; impact on visibility);

ii. related, the probable **usefulness** of an evaluation (relevance and timeliness);

iii. the **objective/purpose** of the evaluation synthesis (compiling facts and/or lessons learned vs. accountability);

iv. the **accessibility and quality of the evaluation reports** (validity, reliability and geographic distribution of the findings) before including them in the study sample (as explained in section C.2) and the extent the reports will allow for systematic data-collection and the identification of causal relations/trends;

---

10   An Excel sheet capturing OED evaluation syntheses from the recent years has already been developed.

v. if there is a sufficient amount of timely, up-to-date evaluative evidence to prepare the synthesis;

vi. **accessibility and availability of stakeholders** (e.g. for interviews, briefings) to ensure a consultative exercise from the start till the dissemination of the evaluation synthesis;

vii. if there is any expected **complexity**; and

viii. **trade-offs**: Time and cost effectiveness vs. utility (e.g. rapid assessments may cost less and provide more timely information, while systematic evaluation syntheses may be more expensive and provide more relevant findings).

A working paper from the UK Department for International Development (DFID) synthesised the checklists used by 11 different agencies into a set of three checklists that cover the dimensions of evaluability. These can provide a useful "starter pack" which can be adapted according to circumstances. The lists can be consulted here: an Evaluability Assessment checklist[11]

Relevance of the evaluation synthesis scope. It is essential that the scope of an evaluation synthesis matches with the scope of that which the evaluation synthesis needs to inform (e.g. new policy, decision making on new interventions, etc.). Another important point of attention is the timeframe of the evaluation reports in the sample. While this will depend on the topic and utility of the evaluation synthesis, we should avoid synthesizing outdated information, and therefore try to avoid using evaluation reports that are older than five years.

Validity and credibility. Firstly, it is important that the whole process of including/excluding quality criteria, screening and coding is transparent and systematic. Secondly, it is important that the evaluation synthesis addresses the implications of variations by understanding how the various interventions in the sample were actually implemented. This means that the evaluation synthesis needs to explore implementation aspects as well as contextual variation, next to the results.

Lessons learned or/and with recommendations. The evaluation team should decide upfront if the evaluation synthesis should contain a final chapter with lessons learned and/or recommendations. In case of the latter, the evaluation synthesis should choose to include either i) newly formulated strategic recommendations building on the recommendations previously made by OED evaluation reports that informed the synthesis; or ii) reiterating the most important existing recommendations. If the evaluation synthesis has recommendations, there should also be a follow-up report after one year, to capture evidence of utilization.

Dissemination. To increase the utility of an evaluation synthesis, wide and strategic dissemination is desired. To make an evaluation synthesis more digestible for the Programme Committee, FAO divisions and external stakeholders, several evaluation offices (e.g. WFP, IFAD and ADB) prepare summary reports, in addition to the full evaluation syntheses report. Moreover, at IFAD, annual in-house learning events are organized on evaluation synthesis topics to enhance visibility, learning and stimulate collective action among various divisions.

4.2 OED working arrangements for evaluation syntheses

Evaluation syntheses cut across the work of all OED teams. As mentioned in the beginning of this note, it is an opportune moment for OED to increase the volume of the evaluation syntheses it produces and disseminates to the FAO decision-making bodies; internal and external.

Depending on the complexity and scope, evaluation syntheses could be prepared by OED colleagues, or OED could recruit one or more external consultants that work under an OED evaluation manager.

Furthermore, depending on time, resources and interest in shared learning, the evaluation could also be conducted jointly with other evaluation offices on topics of shared interest. While a joint evaluation

---

11 This checklist has been extracted from pages 19-23 of the following report: Davies, R., 2013. *Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations.* Report of a Study Commissioned by the Department for International Development (see bibliography for link).

synthesis could provide a wider perspective on a topic, it should be taken into account that joint-evaluation exercises tend to take much longer as protocols and work styles differ among evaluation offices.

With a view of strengthening OED's work on evaluation synthesis, the following is proposed:

i. Methodological training for evaluation syntheses (could be part of the OED-wide learning initiatives).

ii. "Manual"-type guidance material with step-by-step advice to OED colleagues on the conduct of evaluation syntheses. For example, it would be important to dedicate enough time for the start-up stage, which is the most labour-intensive stage for evaluation syntheses managers.

iii. Repository of a selection of tools and guidelines, which OED colleagues find most useful and relevant to evaluation syntheses at the OED. A starting point could be to select these from the Appendix 1 of this document.

iv. A SharePoint repository of past OED syntheses and best examples from other development agencies.

v. The procurement of software, for use across OED on as-needed basis, to aid in the qualitative assessment of a large volume of text [e.g. from evaluations included in a given synthesis study].

vi. Finally, beyond an encouragement to work across teams within OED for evaluation syntheses, either i) an establishment of a peer-review system for evaluation syntheses, or ii) imbedding an evaluation syntheses quality review in the forthcoming OED quality support mechanism. As needed, part-time assistance across OED teams can also be discussed (i.e. to make use of persons with expertise such as software, subject-area, or utilization-focused packaging of the syntheses' results).

Bibliography

Barnett-Page, E., & Thomas, J. 2009. Methods for the synthesis of qualitative research: a critical review. *BMC medical research methodology,* 9(59). (available at https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-9-59)

Better Evaluation. n.d. Synthesise Data Across Evaluations. In: *Better Evaluation* [online]. (available at: https://www.betterevaluation.org/en/rainbow_framework/synthesise/synthesise_across_evaluations)

Davies, R. 2013. Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations. Report of a Study Commissioned by the Department for International Development. (available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/248656/wp40-planning-eval-assessments.pdf)

FAO. 2020. *Regional Syntheses of OED evaluations 2014-2019 Lessons Learnt.* Rome. (internal document)

IFAD. 2015. *Evaluation Manual. Second edition.* Independent Office of Evaluation, IFAD. Rome. (available at: https://www.ifad.org/documents/38714182/39748829/manual.pdf/bfec198c-62fd-46ff-abae-285d0e0709d6)

International Initiative for Impact Evaluation (3ie). 2019. *Does incorporating participation and accountability improve development outcomes? Meta-analysis and framework synthesis.* 3ie Systematic Review 43. (available at: https://www.3ieimpact.org/evidence-hub/publications/systematic-reviews/does-incorporating-participation-and-accountability)

OECD-DAC. 2010. *Glossary of key terms in evaluation and results based management*. Paris. (available at: https://www.oecd.org/development/peer-reviews/2754804.pdf)

Miranda, R. 2020. Methods and tools used for the RLC evaluation synthesis report. Case study. (internal document)

The Presidency of the Republic of South Africa. 2014. *DPME Evaluation Guideline No 2.2.15. Guideline on Evaluation Synthesis.* Page 1. Department Performance Monitoring and Evaluation (DPME). Pretoria. (available at: https://evaluations.dpme.gov.za/images/gallery/Guideline%202.2.15%20Evaluation%20Synthesis%20accepted%2014%2003%2020.pdf)

USAID. 2019. *Learning from Evaluation Syntheses. Technical report on making evidence from evaluations more accessible to decision-makers.* Washington, DC. (available at: https://pdf.usaid.gov/pdf_docs/PA00TP3V.pdf)

USAID. 2019. *Discussion Note: Making Evidence Accessible Through Evaluation Synthesis.* Washington, DC. (available at: https://usaidlearninglab.org/sites/default/files/resource/files/evaluation_synthesis_discussion_note_v1_final_april.24.2019.pdf)

Appendix 1. Useful resources

American Journal of Evaluation. "Quality, Context, and Use Issues in Achieving the Goals of Metaevaluation."
https://wmich.edu/sites/default/files/attachments/u58/2015/Quality_Context_and_Use.pdf

This article discusses the multiple ways in which quality can be defined, the political and cultural contexts of metaevaluation, and issues surrounding use and misuse. A metaevaluation of evaluations of international agricultural research centers illustrates these topics.

ALNAP. "Learning from What We Know: How to improve evaluation synthesis for humanitarian organisations."
https://www.alnap.org/help-library/learning-from-what-we-know-how-to-improve-evaluation-synthesis-for-humanitarian

Although the volume of evaluations published by humanitarian actors has increased tenfold between 2007 and 2017, the evaluation system still struggles to provide evidence about collective performance of the humanitarian sector. Evaluation synthesis provides one option for building information about response-wide performance. This paper provides guidance on when to do an evaluation synthesis and how to do it well.

ALNAP. "Evaluating Evaluations: A Meta-evaluation Checklist."
https://www.alnap.org/help-library/evaluating-evaluations-a-meta-evaluation-checklist

What are the criteria of merit for an evaluation in any field, including program evaluation? That is probably the most important question to which someone conducting a meta-evaluation. This article provides a list of criteria for how to conduct a meta-evaluation and what sources to include.

Better Evaluation. "Introducing systemic reviews."
https://www.betterevaluation.org/pt/node/1555

In this chapter, the authors outline the purpose of systematic reviews and explain their value in making decisions. Furthermore, they "consider what 'systematic' means when applied to reviewing literature and explains how review methods may vary while being systematic." It also provides an overview of debates regarding synthetic reviews.

Better Evaluation. "Rapid evidence assessment."
https://www.betterevaluation.org/evaluation-options/rapid_evidence_assessment

The article provides an overview of Rapid Evidence Assessment, a process that is faster and less rigorous than a full systematic review but more rigorous than ad hoc searching, it uses a combination of key informant interviews and targeted literature searches to produce a report in a few days or a few weeks.

Better Evaluation. "Synthesize data across evaluations."
https://www.betterevaluation.org/pt/node/249

This overview covers strategies to apply when locating the evidence (often involving bibliographic searches of databases, with particular emphasis on finding unpublished studies), assessing its quality and relevance in order to decide whether or not to include it, extracting the relevant information, and synthesizing it as well as different strategies and definitions of what constitutes credible evidence.

Better Evaluation. "Synthesize data from one or more evaluations."
https://www.betterevaluation.org/sites/default/files/Synthesize%20-%20Compact.pdf

The overview provides guidelines on how to combine data to form an overall assessment of the merit or worth of the intervention, or to summarize evidence across several evaluations.

DFID. Assessing Strength of Evidence
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/291982/HTN-strength-evidence-march2014.pdf

This note offers recommendations on how to use evidence judiciously for the benefit of designing and implementing effective policy and programmes, including the appraisal of the quality of individual studies and the assessment of the strength of bodies of evidence.

Eval Forward. "Use of synthesis and meta-analysis in development evaluation."
https://www.evalforward.org/resources/synthesis-and-meta-analysis

This webinar was held on 30 October 2019 and presented the experience of different UN agencies and the International Initiative for Impact Evaluation (3ie) in using synthesis and meta-analysis in evaluation. Participants discussed how to improve the quality and use of synthesis in development evaluations and learn about their potential application in different development interventions.

Regional syntheses of OED evaluations 2014–2019: Lessons Learnt
This document has the purpose of sharing lessons learned from conducting the OED Regional Syntheses as to inform OED decisions on conducting similar syntheses in the future. It is an internal document.

The Feinstein International Centre. "The Humanitarian Evidence Programme."
https://fic.tufts.edu/research-item/the-humanitarian-evidence-program/

The site features systematic reviews of evidence types, each of which is accompanied by a protocol detailing the methodology and an evidence brief synthesizing key insights. You can also find the program's guidance documents for how to map, synthesize, and critically appraise humanitarian evidence.

The Feinstein International Centre. "Synthesizing practices of evidence appraisal in the humanitarian field."
https://www.alnap.org/help-library/synthesizing-practices-of-evidence-appraisal-in-the-humanitarian-field

This paper synthesizes evidence appraisal practices to inform evidence syntheses in the Humanitarian Evidence Program. The paper helps answer the question, 'How confident are we in the quality of evidence supporting a finding? Rather than prescribing a particular evidence appraisal approach, the Humanitarian Evidence Programme summarizes evidence appraisal practices relevant to the field and offers some suggestions in critically applying them to the realities of humanitarian data analysis, synthesis, and interpretation.

International Initiative for Impact Evaluation. Evidence Hub
https://developmentevidence.3ieimpact.org/

3ie's Development Evidence Portal is a repository of rigorous evidence on what works in international development. This portal includes evaluations and synthesis of studies conducted in low-and middle-income countries. It combines records from 3ie's Impact Evaluation and Systematic Review repositories, as well as, evidence gap maps.

UNEG. "Compendium of Evaluation Methods Reviewed"
http://www.unevaluation.org/document/detail/2939

The Compendium of Evaluation Methods Reviewed is a guidance document resulting from discussions among members of the UNEG Working Group on Evaluation Methods on seven evaluation methods. i) methods supporting evaluation design: evaluability assessment, theory of change and storyline approaches; ii) synthesis and meta-analysis; iii) contribution analysis; iv) qualitative comparative analysis; v) randomized controlled trials and quasi-experimental designs; vi) outcome harvesting/outcome evidencing; and vii) culturally responsive evaluation. The Compendium presents examples of interesting UN evaluation practices to a wider audience, with an additional focus on the underlying principles that led evaluation managers to choose the method they did.

The Presidency Republic of South Africa. "Guideline on Evaluation Synthesis"
https://evaluations.dpme.gov.za/images/gallery/Guideline%202.2.15%20Evaluation%20Synthesis%20accepted%2014%20 03%2020.pdf

This guideline is designed to assist government departments to effectively plan and manage evaluation synthesis. The Guideline provides a definition and description of evaluation synthesis, followed by key issues to be considered, key questions that should be addressed, and typical evaluation methods. This is a broad guideline that can be applied in different contexts.

USAID. "Learning from Evaluation Syntheses."
https://pdf.usaid.gov/pdf_docs/PA00TP3V.pdf

This report focuses on evaluation syntheses as aid for learning and improving the utilization of evidence from evaluations. It focuses on how decision-makers use such evidence to inform foreign assistance policies, strategies, projects, and activities, and for monitoring and improving evaluation quality.

USAID. "Making Evidence Accessible through Evaluation Synthesis."
https://usaidlearninglab.org/sites/default/files/resource/files/evaluation_synthesis_discussion_note_v1_final_april.24.2019.pdf

Section I of this Discussion Note describes the types of evaluation syntheses. Section II examines how syntheses can enhance evidence-based decision-making within USAID's Program Cycle. Section III discusses stages in the evaluation synthesis process, including design, implementation, dissemination, and utilization tracking. Section IV provides links to additional resources for those interested in pursuing one of the main types of evaluation syntheses or for understanding how to identify evidence gaps that learning agendas can help fill.

IOD PARC "Evaluation Methodologies"
https://www.iodparc.com/resource/evaluation-methodologies/

This review provides a brief overview of three synthesis methodologies used to deliver wider learning (what works where and why) around evaluation processes and specified topics: meta-evaluation, synthesis evaluations and systematic reviews.

Practical tools

Meta-Analysis Effect Size Calculator

David B. Wilson, Ph.D., George Mason University. Available at:
https://www.campbellcollaboration.org/research-resources/effect-size-calculator.html

Recorded trainings

Campbell Collaboration, available on YouTube

i.   Question formulation

ii. Searching, coding and quality

iii. Meta-analysis methods

iv. Advanced methods

v.   Policy engagement

Office of Evaluation
E-mail: evaluation@fao.org
Web address: www.fao.org/evaluation

Food and Agriculture Organization of the United Nations
Rome, Italy