

Meta-Evaluation:

a synthesis of standards and tools



A contribution to the 2030 Agenda and the



Dr. Olmedo J. Varela, Ph.D.
2017

Contents

Acronyms.....	ii
1. Introduction.....	1
Examining evaluation processes.....	1
Assessing and scoring the quality of evaluations	1
2. Framework to Examine Evaluation Processes	2
Examining the Quality of the Evaluation Process.....	2
Defining Purpose.....	2
Planning	2
Design.....	2
Implementation.....	2
Reporting	3
Learning and Utilization.....	4
3. Standards to examine evaluation processes.....	5
Organizations' internally established parameters for Evaluation and Research	5
Key quality dimensions for each phase of a typical evaluation process.....	6
4. Assessing the Quality of Evaluations.....	10
BOND Principles.....	10
The BOND Checklist and Scoring Tool	10

Acronyms

BOND	United Kingdom, Body for organisations working in international development
ECG	Evaluation Cooperation Group
IOCE	International Organization for Cooperation in Evaluation
NGO	Non-governmental Organisation
OECD	Organization for Economic Co-operation and Development
DAC	Development Assistance Committee
OECD/DAC	Organization for Economic Co-operation and Development's Development Assistance Committee
SMART	Specific, Measurable, Attainable, Realistic and Timely
ToR	Terms of Reference
UK	United Kingdom
UNDP	United Nations Development Program
UNEG	United Nations Evaluation Group

I. Introduction

The term “meta-evaluation” is used to mean the: “evaluation of an evaluation to judge its quality”. This is based on the definition of the “OECD/DAC Glossary of Key Terms” specifically related to the quality of evaluations. Using these guidelines, evaluation practitioners will be able to use the industry’s standards to help guide the examination of the quality of **evaluation processes**. The intention of presenting generally accepted standards is to make evaluators aware of the industry’s general consensus regarding key quality components of evaluation processes. In addition, the guidelines provide a specialized meta-evaluation tool to assess and score the quality of **evaluation practices**, including selected quality dimensions of the evaluation process. A meta-evaluation approach that focuses on the quality of evaluations processes and practice will progressively increase accountability of future examinations and assessments revealing performance of interventions from individual as well as across multiple evaluations.

Examining evaluation processes

The industry norms and standards related to the quality of evaluation processes have progressively evolved toward a basic consensus through global dialogues led by the Organization for Economic Co-operation and Development’s Development Assistance Committee (OECD/DAC) Evaluation Network, including, among other members, the World Bank, the Inter-American Development Bank and the United Nations Development Program (UNDP). These norms and standards are used as the basis for examining evaluation processes and to facilitate a consistent and systematic approach to meta-evaluation.

A detailed checklist is presented to provide meta-evaluation practitioners with the industry’s quality criteria and standards to help guide and examine evaluation processes. These are standards to be considered in meta-evaluation processes with a particular focus on reporting, learning and utilization (almost half of standards in the checklist belong to these three domains of the evaluation process). **These standards are more likely to be used as a reference** by evaluation teams to organize evaluation processes and by evaluation evaluators performing meta-evaluations. **As evaluation capacities and practices evolve in an organization overtime, program evaluations processes will be able to display a checkmark to an increasing number of items listed in the checklist.**

Assessing and scoring the quality of evaluations

The BOND tool is provided to set a recent and innovative evaluation resource for assessing and scoring the quality of evaluation practices, including selected quality dimensions of the evaluation process. The tool is based on five principles and twenty criteria. A checklist reflecting principles and criteria for assessing and improving the quality of evidence in evaluation reports is provided by the tool. The checklist includes four questions for each of the five principles and each question can be answered on scale of 1-4.

The conceptual framework and the methods and technics used to assess the quality of evaluation practices, including selected quality dimensions of the evaluation process are provided. The actual tool, including the checklist, the scoring scale definitions, and a “step by step” guide on how to interpret results are also provided.

2. Framework to Examine Evaluation Processes

In this document, meta-evaluation is defined as a process of evaluating an evaluation or set of evaluations to “judge its quality”¹. The components of the process are examined during a meta-evaluation exercise, based on the industry’s key quality dimensions for each phase of a typical evaluation process: defining purpose, planning, designing, implementing, reporting, and learning from and using evaluation results². Using these guidelines, the quality of programme evaluation processes and practices in a given organization are examined according to internationally agreed standards.

Examining the Quality of the Evaluation Process

Examination of the quality of evaluation processes, within the context of a meta-evaluation, involves a participatory reflexion exercise on each phase of a typical evaluation process against internationally agreed standards related to key quality dimensions: defining purpose, planning, designing, implementing, reporting, and learning from and using evaluation results³. Consistent with the industry’s standards, expectations of evaluation processes regarding each phase are summarized next.

Defining Purpose

Defining purpose involves building a clear rationale for the purpose and intended use of the evaluation. The rationale should address the reasons why and for whom the evaluation is being undertaken and how will it contribute to learning and accountability. In addition, clear statements regarding the specific objectives, objects involved in the evaluation, and the scope of the evaluation are expected.

Planning

Planning involves assessing the feasibility of the evaluation as well as the mechanisms to ensure participation of relevant stakeholders. The feasibility of the evaluation is expected to be examined by determining whether or not the development intervention is adequately defined and its results verifiable. Relevant stakeholders are expected to be involved early on in the evaluation process and they should have the opportunity to contribute to the evaluation design. The potential for a joint evaluation conducted collaboratively with other agencies is expected to be considered during the planning phase.

Design

The *design* phase involves formulating specific evaluation questions derived from the evaluation objectives as well as the selection and application of evaluation criteria based on relevance, efficiency, effectiveness, impact and sustainability⁴. A key component of the *design* is the selection of approaches and methods chosen to answer the evaluation questions, including techniques for data collection and analysis. This component is crucial for the meta-evaluation component related to the quality of evidence used to report evaluation findings. Terms of Reference (ToR) detailing the various components of the evaluation process are expected in the design phase, including the purpose, scope, and objectives of the evaluation; the methods used; the resources and time allocated; and the reporting requirements.

Implementation

Implementation involves the selection of the evaluation team, including assessment of the team’s skills and evaluation credentials. Ideally, evaluators are expected to be independent from the development

¹ OECD/DAC, ‘Glossary of Key Terms in Evaluation and Results Based Management,’ (2010), Page 27

² OECD/DAC, ‘Quality Standards for Development Evaluations’, (2010), Page 5

³ Ibid

⁴ For evaluation criteria, see DAC criteria for evaluating development assistance in OECD/DAC, ‘Evaluating Development Cooperation: summary of key norms and standards’, (2010), Pages 13-14

intervention. When constraints regarding resources for this particular expectation and other reasons are substantiated, this situation should be reported and the evaluation process should be sensitive to potential conflicts of interest. The implementation phase includes finalising and testing data collection tools, training of data collectors and supervision / ensuring data quality. On-going consultation with stakeholders is expected in the implementation phase as well as strict confidentiality of individual informants. Implementation is expected to take place according to planned schedule and allocated budget, but needs to remain flexible to the realities of the field.

Reporting

Assessing the quality of *reporting evaluation findings* constitutes a core piece of any meta-evaluation exercise. A significant proportion of standards are actually related to this particular phase.

Audience

The evaluation report is expected to clearly communicate its purpose and findings should be reported in accordance with the characteristics of the intended audience. The report should include an overview of the report, highlighting the main findings, conclusions, recommendations and the overall lessons learnt.

Context

The context of the development intervention should be described and the evaluation should identify and assess the influence of the context on the performance of the development intervention. Description should include policy context; the specific policies established by the development agency and partners; objectives and strategies; and development context, including socio-economic, political and cultural factors. The institutional context and stakeholder involvement should be also incorporated in the description.

Assumptions

A key component of the report is the description and assessment of the logic of the intervention or the theory supporting the underlying assumptions and factors affecting the performance of the intervention. The validity and reliability of information sources should be described in sufficient detail to allow assessment of the adequacy of the information. Detailed explanations regarding the selection of case studies or sampling exercises are expected to be provided in the report. Any limitations regarding the representativeness of the samples are expected to be identified.

Sources of Information

The report should reveal how the evaluation cross-validates the information sources and how it critically assesses the validity and reliability of the data. Detailed information, including complete lists of interviewees and other information sources consulted is expected to be included in the report but, it should be presented in ways that do not conflict with the privacy and confidentiality of participants.

Methods

Another key component of the report is the description and explanations related to the evaluation methods used. Depending on the scope and type of evaluation, the assessment of outcomes and/or impacts, attribution and/or contribution to results are expected to be explained. The report is expected to acknowledge any constraints encountered during the evaluation process and how these could have affected the evaluation, including the independence and impartiality of the evaluation. It is expected to detail the techniques used for data collection and analysis. The criteria to select specific techniques should be substantiated and limitations and shortcomings are expected to be explained.

Analysis

In a typical evaluation process, the analysis of evaluation results should be presented in a clear manner, providing a clear logical distinction between findings, conclusions, recommendations and lessons learnt. It is expected that the analysis of evaluation findings develops according to the intervention's underlying

assumptions. Analysis should reveal a clear line of evidence to support the conclusions and recommendations and lessons learnt follow logically from the conclusions.

Evaluation Questions

In the evaluation report, all the questions detailed in the ToR for the evaluation are expected to be answered. The evaluation questions as well as any revisions should be thoroughly documented to allow assessment of the extent to which the evaluation team has satisfactorily met the evaluation objectives and addressed the questions, including those related to cross-cutting issues.

Changes and Disagreements

Acknowledgement of changes and limitations encountered during the evaluation process are expected to be clearly explained. These include issues related to the methods, sources of information and its validity and reliability. Explanations regarding discrepancies between the planned and actual implementation and products are expected. Also expected is reporting on disagreements within the evaluation team regarding particular judgments and recommendations and any unresolved differences of opinion. Similarly, relevant stakeholders are expected to have the opportunity to comment on the report's drafts. The final report should reflect stakeholders' comments and acknowledges any substantive disagreements.

Learning and Utilization

The evaluation process should facilitate effective learning and it should provide the necessary information to meet the needs of the intended users. The evaluation's main conclusions, recommendations and lessons should be clearly stated and these should be relevant, targeted and actionable. The evaluation should be also be delivered according to the planned schedule to ensure optimal use of the evaluation findings.

Evaluation findings should be stored and managed in ways that make them available to the target audience and all development partners, including identification of interested parties and the wider development community. Dissemination should be systematic making sure utilization is maximized and transparency of the evaluation process is honored. Utilization is expected to be reflected in a systematic use of recommendations through actions taken by the targeted audience. A management response plan should be formally established with institutional follow-up mechanisms. Institutional agreements regarding action plans should be tracked to ensure accountability of its implementation.

3. Standards to examine evaluation processes

Evaluations assess what happened as a result of program implementation providing a systematic learning from evaluation processes to inform decision making and planning.

The following checklist can be used as a reference to help examine evaluation processes within the context of a meta-evaluation exercise: from the definition of purpose phase to the utilization of evaluation findings phase. As evaluation capacities and practices evolve overtime in a given organization, program evaluations will be able to display a checkmark to an increasing number of items listed in the meta-evaluation checklist, including specific page numbers where the standard is applied by the examined evaluation.

Examination of the quality of evaluation processes, within the context of a meta-evaluation, involves a participatory reflexion exercise on each phase of a typical evaluation process against internationally agreed standards related to key quality dimensions: defining purpose, planning, designing, implementing, reporting, and learning from and using evaluation results. Consistent with the industry's standards, a detailed list regarding each key quality dimension is presented next.

The checklist isn't used here to assess the quality of a conducted evaluation and it doesn't result in a score. The next section provides an institutionally agreed tool to conduct these two tasks.

Checklist ⁵			
<i>Organizations' internally established parameters for Evaluation and Research</i>		✓	page #
1.	In line with the organization's evaluation and research strategy and priorities		
2.	In line with the organization's evaluation competencies		
3.	In line with the organization's research guidelines and definitions		
4.	Organization's Research Review completed and signed off by appropriate group		
5.	In line with the organization's ethical principles for conducting evaluation and research		
6.	Organization's Ethics Committee Review parameters completed and signed off		
7.	If partnering with Academic Institutions, in line with the organization's policies		

⁵The technical team's criteria to substantiate the selection of specific methods and techniques and the thematic content of evaluations, including its conceptual and strategic alignment with the organization's philosophical framework and specific institutional agenda, should be assessed by the respective technical teams. Consistency between institutional strategies, priorities, and principles, including design, implementation and utilization should also be assessed by the respective technical teams.

<i>Key quality dimensions for each phase of a typical evaluation process</i> ⁶		✓	page #
8.	Explanations have been provided regarding the reasons for conducting the evaluation		
9.	Explanations have been provided regarding selection of the targeted audience		
10.	Explanations have been provided on how the evaluation is to be used for learning and/or accountability functions		
11.	The specific objectives of the evaluation clarify what the evaluation aims to find out		
12.	The development intervention being evaluated (the evaluation object) is clearly defined, including a description of the intervention's logic or theory		
13.	The evaluation scope defines the time period, funds spent, geographical area, target groups, organisational set-up, implementation arrangements, policy and institutional context and other dimensions covered by the evaluation		
14.	Discrepancies between the planned and actual implementation of the development intervention are identified		
15.	The feasibility of the evaluation has been examined by determining whether or not the development intervention is adequately defined and its results verifiable		
16.	Relevant stakeholders were involved early on in the evaluation process and given the opportunity to contribute to evaluation design		
17.	Join evaluation to be conducted collaboratively with other agencies were explored in the planning process		
18.	The evaluation objectives were translated into relevant and specific evaluation questions		
19.	Evaluation questions were decided early on in the process and these inform the selection and/or development of evaluation methods		
20.	The evaluation questions address cross-cutting issues, such as gender, the environment and human rights		
21.	The evaluation did apply the agreed DAC criteria for evaluating development assistance: relevance, efficiency, effectiveness, impact and sustainability ⁷		
22.	The selection of the evaluation approach and derived evaluation methods were substantiated in accordance with the purpose, scope, and evaluation questions		
23.	The evaluation methods included specification of the techniques for data collection and analysis		
24.	The selected methods answer the evaluation questions using credible evidence		
Cont. Design			
25.	A clear distinction is made between the different result levels (intervention logic containing an objective-means hierarchy stating input, output and outcome)		
26.	Indicators for measuring achievement of the objectives are validated according to institutionally agreed standard indicators		

⁶ Sometimes these parameters are covered by the organization's guidelines and standards related to evaluation and research listed above. However, the following parameters should be in place to ensure observance of generally agreed standards specifically related to the quality of evaluation practices. For general agreement on standards and specific parameters included in this checklist, see:

- OECD/DAC, 'Quality Standards for Development Evaluations', (2010)
- OECD/DAC, 'Evaluating Development Cooperation: summary of key norms and standards', (2010)
- OECD/DAC, 'Evaluating Development Activity: 12 lessons from the OECD/DAC' (2013)

⁷ Note: The application of these and any additional criteria depends on the evaluation questions and the objectives of the evaluation. If a particular criterion is not applied and/or any additional criteria added, this is explained in the evaluation report. All criteria applied are expected to be defined in unambiguous terms.

See

- DAC criteria for evaluating development assistance in OECD/DAC, 'Evaluating Development Cooperation: summary of key norms and standards', (2010), Pages 13-14

27.	Disaggregated data is presented to clarify any differences between sexes and between different groups of poor people	✓	page #
28.	The resources provided for the evaluation were adequate in terms of funds, staff and skills, to ensure that the objectives of the evaluation were fulfilled effectively		
29.	The evaluation process was driven by governance and management structures designed to fit the evaluation's context, purpose, scope and objectives		
30.	Terms of Reference (ToR) were established specifying the purpose, scope, and objectives of the evaluation; the methodology to be used; the resources and time allocated; reporting requirements; and any other expectations regarding the evaluation process and products		
31.	A transparent and open procurement procedure was used for selecting the evaluation team		
32.	The members of the evaluation team had a mix of evaluative skills and thematic knowledge		
33.	A gender balance was considered in the formation of the evaluation team		
34.	Evaluators were independent from the development intervention, including its policy, operations and management functions, as well as intended beneficiaries ⁸		
35.	Potential conflicts of interest were addressed openly and honestly		
36.	The evaluation team was able to work freely and without interference		
37.	Access to all relevant information and cooperation was assured		
38.	The full range of stakeholders, including both partners and donors, were consulted during the evaluation process and given the opportunity to contribute		
39.	The criteria for identifying and selecting stakeholders was specified		
40.	The rights and welfare of participants in the evaluation were protected		
41.	Anonymity and confidentiality of individual informants were protected when requested or as needed		
42.	The evaluation was conducted and results were made available to the organization's officials and to key stakeholders in a timely manner to achieve the objectives of the evaluation		
43.	The evaluation was carried out efficiently and within budget, and/or changes in conditions and circumstances were reported and un-envisaged changes to timeframe and budget were explained,		
44.	Changes in conditions and circumstances were discussed and agreed between the relevant parties		
45.	The evaluation report contains an executive summary providing an overview of the report and highlighting the main findings, conclusions, recommendations and any overall lessons		
Cont. Reporting			
46.	The context of the development intervention is described, including policy context, the organization's policies, objectives and strategies; development context (socio-economic, political and cultural factors); and institutional context and stakeholder involvement		
47.	The evaluation identifies and assesses the influence of the context on the performance of the development intervention.		

⁸ Evaluators are expected to be independent from the development intervention. When constraints regarding resources for this particular expectation or other reasons are substantiated, this situation should be reported and the evaluation process should be sensitive to potential conflicts of interest.

48.	The report describes and assesses the intervention logic or theory, including underlying assumptions and factors affecting the success of the intervention	✓	page #
49.	The report describes the sources of information used in sufficient detail so that the adequacy of the information and the quality of evidence can be assessed		
50.	The report explains the selection of case studies or any samples, including identification of limitations regarding the representativeness of the samples		
51.	The evaluation cross-validates the information sources and critically assesses the validity and reliability of the data		
52.	Complete lists of interviewees and other information sources are included with rigorous consideration of the privacy and confidentiality of participants		
53.	The report describes and explains the evaluation methodology and its application		
54.	In assessing outcomes and impacts, attribution and/or contribution to results are explained		
55.	The report acknowledges any constraints encountered and how these have affected the evaluation, including the independence and impartiality of the evaluation		
56.	The report details the techniques used for data collection and analysis, including explanation and justification of the choices and limitations and shortcomings		
57.	The report presents findings, conclusions, recommendations and lessons separately and with a clear logical distinction between them		
58.	Findings flow logically from the analysis of the data, showing a clear line of evidence to support the conclusions		
59.	Conclusions are substantiated by findings and analysis		
60.	Recommendations and any lessons follow logically from the conclusions		
61.	Any assumptions underlying the analysis are made explicit		
62.	The report answers all the questions detailed in the ToR for the evaluation and where this is not possible, explanations are provided		
63.	The original questions, as well as any revisions to these questions, are documented in the report for readers to be able to assess whether the evaluation team has sufficiently addressed the questions, including those related to cross-cutting issues, and met the evaluation objectives		
64.	The report explains any limitations in process, methodology or data, and discusses validity and reliability		
65.	The report indicates any obstruction of a free and open evaluation process which may have influenced the findings		
66.	Any discrepancies between the planned and actual implementation and products of the evaluation are explained		
67.	Evaluation team members had the opportunity to dissociate themselves from particular judgements and recommendations on which they disagree		
68.	Any unresolved differences of opinion within the team are acknowledged in the report		
Cont. Reporting			
69.	Relevant stakeholders are given the opportunity to comment on the draft report and the final evaluation report reflects these comments and acknowledges any substantive disagreements		
70.	In disputes about facts that can be verified, the evaluators investigated and changed the draft where necessary		

71.	In the case of opinion or interpretation, stakeholders' comments were reproduced verbatim, in an annex or footnote, to the extent that this does not conflict with the rights and welfare of participants	✓	page #
72.	Conclusions, recommendations and lessons are clear, relevant, targeted and actionable so that the evaluation can be used to achieve its intended learning and accountability objectives		
73.	The evaluation is delivered in time to ensure optimal use of the results		
74.	Systematic dissemination, storage and management of the evaluation report is ensured to provide easy access to all development partners, to reach target audiences, and to maximise the learning benefits of the evaluation		
75.	Recommendations are systematically responded to and action taken by the person(s)/body targeted in each recommendation, including a formal management response and follow-up		
76.	All agreed follow-up actions are tracked to ensure accountability for their implementation		
77.	The evaluation results are presented in an accessible format and are systematically distributed internally and externally for learning and follow-up actions and to ensure transparency		
78.	In light of lessons emerging from the evaluation, additional interested parties in the wider development community are identified and targeted to maximise the use of relevant findings		



4. Assessing the Quality of Evaluations

A critical quality dimension of the evaluation process to be assessed by a meta-evaluation exercise is the assessment of the quality of the evidence used to report evaluation findings. For assessing the quality of the evidence used by organizations' programmes to report evaluation findings, including selected quality dimensions of the evaluation process, the BOND tool⁹ is presented and it could be incorporated to assess and score organizations' conducted evaluations.

The tool has been designed specifically for NGOs and it can be used when commissioning, designing and reviewing evidence-based work. The principles provide a valuable framework that promotes high quality evaluations. Well known organizations have been supportive of the BOND's evidence principles initiative.

BOND Principles

The tool covers the following dimensions of evidence and these constitute the five principles of quality:

1. Voice and inclusion: beneficiaries voices were listened to on the effect of the intervention and the data identified who was affected and how
2. Appropriateness: the methods used were appropriate for the intervention and purpose of the baseline or evaluation
3. Triangulation: mixed methods and a variety of perspectives were used to make conclusions about the evidence.
4. Contribution: the change that occurred is demonstrated by the data and the evidence explains how it we contributed to it
5. Transparency: data sources, methods used, results achieved and the strengths and limitations of the evidence are clearly and openly documented.

The BOND Checklist and Scoring Tool

The tool is based on five principles and twenty criteria¹⁰. A checklist reflecting principles and criteria for assessing and improving the quality of evidence in evaluation reports, including selected quality dimensions of the evaluation process, is provided by the tool. The checklist reflecting principles and criteria includes four questions for each of the five principles and each question can be answered on scale of 1-4.

As a mechanism to facilitate the scoring and to ensure consistency, the BOND tool has developed scales for each question. Scores for each question are then added up providing an overall score for the principles. As shown on the table below, depending on the score, the principle is then assigned a color.

⁹ BOND UK, NGO Evidence Principles <http://www.bond.org.uk/effectiveness/principles#the-checklist>

¹⁰ Bond for International Development, 'Principles and checklist for assessing the quality of evidence v.1.0', (2013)

An overall score for the principle of 4 – 6	Evidence is weak in this area
An overall score for the principle of 7 – 10	Evidence meets a minimum standard in this area
An overall score for the principle of 11 - 13	Evidence meets a good standard in this area
An overall score for the principle of 14 – 16	Evidence meets gold standard in this area

According to the guide, by default, each principle is weighted 20% (See table below). Based on the quality of evidence criteria, evaluators, using the tool within a MS Excel environment, change the weight on each principle and provide a narrative to explain the reasoning behind the weightings. The total weighting must add up to 100%.

Principle	Total score	Weighting	Weighted Score	Quality Assessment
1) Voice and Inclusion	0 / 16	20%	0 / 16	0%
2) Appropriateness	0 / 16	20%	0 / 16	0%
3) Triangulation	0 / 16	20%	0 / 16	0%
4) Contribution	0 / 16	20%	0 / 16	0%
5) Transparency	0 / 16	20%	0 / 16	0%

Very detailed information about the tool, including its conceptual framework and the methods and techniques used to assess the quality of evidence is provided in a guide. The actual tool, including the checklist, the scoring scale definitions, and a “step by step” guide on how to interpret results can be accessed in the MS Excel file displayed below.



The_Evidence_Principles.xlsx