

Evaluation Questions Checklists

Description

[Tweet](#)



The purpose of this page

1. To provide information on existing checklists of this kind.
2. To collate and prompt ideas on how improved versions of a Evaluation Questions Checklist could be developed

Rationale: The selection of evaluation questions is central to the design of an evaluation. If these choices are not the best possible, the evaluation will be weakened accordingly

Feedback: Please feel free to use the Comment facility at the bottom of this page to make suggestions or comments.

Contents on this page:

1. [Existing resources](#)
 1. [Checklists](#)
 2. [References on the use of checklists](#)
2. [Suggestions for an improved Evaluation Questions checklist](#)
 1. [The purpose of a checklist](#)
 2. [A provisional checklist of types of evaluation questions](#)
 3. [Sourcing options for evaluation questions](#)
 4. [Quality criteria for evaluation questions](#)
 5. [Unresolved issues â?? your views please](#)
3. [Endnote: Where Evaluation questions fit in with other considerations](#)

1. Existing resources

1.1 Checklists

- Delahais, T. (2022) [Writing Better Evaluation Questions](#). Quadrant Conseil
- Wingate, L., & Schroeter, D. (2016). [Evaluation Questions Checklist for Program Evaluation](#). Western Michigan University. The Evaluation Center.
- Spark Policy Institute (2014) Developmental Evaluation Toolkit. [Developmental Evaluation Questions](#)
- USAID (2013) [Checklist for Defining Evaluation Questions](#).
- Davies, R (2012) [Evaluation questions: Managing agency, bias and scale](#)
- CDC (2013) [Good Evaluation Questions: A Checklist to Help Focus Your Evaluation](#)
- Preskill, H., & Jones, N. (2009). [A Practical Guide for Engaging Stakeholders in Developing Evaluation Questions](#). Robert Wood Johnson Foundation.

1.2 References on the use of checklists

- Gawande, A. (2011). [The checklist manifesto: How to get things right](#). Profile Books. **A great read.**
- Scriven, M. (2007). [The Logic and Methodology of Checklists](#). Western Michigan University. **Also recommended**
- Daniel Stufflebeam. (2000). [Guidelines for Developing Evaluation Checklists: The Checklists Development Checklist](#). Western Michigan University. The Evaluation Center.

2. Suggestions for an improved Evaluation Questions checklist

2.1 The purpose of a checklist

Suggested overall purpose: To help improve the *usefulness* and *quality* of an evaluation.

Why use a checklist? Checklists can serve the following purposes:

1. To remind people of the *range* of options that are available when choosing evaluation questions,
2. To be *transparent* about the choices that were made
 1. Optionally, to *explain* why other options were not taken up.
3. To ensure adherence to some *minimal standards*, if they can be found

Rationale: The three lists below are a mix of *options* to be chosen from (Lists 2.2 & 2.3) and some provisional minimal *standards* (2.4) despite the fact that I was initially wary of proposing any one-size-that-fits-all set of attributes of good evaluation questions.

The **Proposed use** column on the right side of the lists below is a nominal space which a checklist user might use to document if the evaluation questions developed to date are of this (row) type, along with any caveats or conditions. In its simplest form the response there could be a yes/no.

2.2 A provisional checklist of types of evaluation questions that can be asked

1: Question types	Description	Priority (binary/ranked/rated)
1. Descriptive	about what happened , when, where when, who. Without this information other questions below can be difficult to answer	
2. Valuative	about peoplesâ?? assessments of the value and significance of what happened. (aka Normative? see below), which will then inform choices about where to investigateâ?!	
3. Explanatory	about the causes of what happened or what happened as a result of a cause	
4. Predictive	about likely consequences of what happened	
5. Prescriptive	about what what could or should be done about what happened, or what is expected to happen	

2.2.1: Kinds of descriptions (relates to 1.1 above)

Priority
(binary/ranked/rated)

1. Comparable descriptions of multiple instances
2. Detailed and specific descriptions about individual cases/instances
3. Historical accounts â?? of events over an extended period of time
4. Stages on an expected process: activities, outputs, outcomes, impacts
5. The â??expectedâ?• versus â??unexpectedâ?• status of events described
6. The degree of agreement/dissent about the nature of the events described [â? others to be identified]

PS: Worth reading

â??[Descriptive statistics are essential to making complex analyses useful](#)â?• by Murphy, 2022

[The gorilla experiment, or why simple scatter plots can be so useful](#), by Yanai and Lercher, 2020

2.2.3: Kinds of value criteria (relates to 1.2 above)

Priority
(binary/ranked/rated)

1. Preferences [aka Relevance? [See below](#)]
2. Equity â?? of process, of outcomes
3. Inclusion â?? partners/beneficiaries in design / steering / management / implementation
4. Sustainability â?? of an intervention and/or of its outcomes
5. Effectiveness
6. Economy, efficiency, cost-effectiveness
7. Resilience, robustness
8. Coherence â?? with the efforts of others
9. Consistency â?? with other activities of the same programme/agent

10. Fidelity – consistency of actual implementation with design intentions

For more options please also see:

[Evaluation Values and Criteria Checklist](#) by Daniel L. Stufflebeam, 2001
[Better Criteria for Better Evaluation. Revised Evaluation Criteria Definitions and Principles for Use](#) OECD/DAC Network on Development Evaluation, 2019

2.2.4: Types of explanatory questions about (relates to 1.3 above)

Priority
(binary/ranked/rated)

1. Causes where the intervention and expected outcome were both present (True Positives)
 2. Causes where the intervention was present but the expected outcome was absent present (False Positives)
 3. Causes where the intervention was absent but the expected outcome was present (False Negatives)
 4. Causes where the intervention and the expected outcome were both absent (True Negatives)
 5. The effects of a cause (equifinality) [i.e. both situations 1 and 2 above]
 6. The causes of an effect (multifinality) [i.e. both situations 1 and 3 above]
 7. Necessary conditions for an outcome to be present (=no false negatives), sufficient or probable status of a known or expected cause
 8. Sufficient conditions for an outcome to be present (=no false positives),
 9. Probable conditions for an outcome to be present (false negatives and false positives are present, but they are in the minority)
- [Others to be identified]

2.2.5: Types of prediction questions (relates to 1.4 above)

Priority
(binary/ranked/rated)

1. Ownership: Who owns/supports particular predictions?
 2. Evidence: What are the evidence requirements for particular predictions to adequately tested?
 3. Accuracy 1: How often is the outcome likely to be present (versus absent) when the intervention is present? (aka Consistency/Precision)?
 4. Accuracy 2: How often is the outcome likely to be present when the intervention is present (versus absent)? (aka Coverage/Sensitivity/Recall)?
- [Others to be identified]

2.2.6: Types of prescriptions/ recommendations that can be sought (relates to 1.5 above)

Priority
(binary/ranked/rated)

1. Customised versus widely applicable recommendations
 2. Prioritised recommendations
- [Others to be identified]

[Other content yet to be provided here]

2.3 Sourcing options for evaluation questions

1: Documents		Priority (binary/ranked/rated)
1. Previous evaluations [or monitoring or audits] of the same or similar activities, especially those needing clarification or not yet answered		
2. Strategy documents referring to the purposes and ambitions of this and other similar activities		
3. Theory(ies) of Change (ToC) for this specific activity.		
4. Published studies and reviews of the same kind of activity [â?others to be identified]		
2: People		Priority (binary/ranked/rated)
1. Funders, who are financing the activity, and others similar		
2. Implementers: Persons and organisations involved in the delivery of the activity		
3. Beneficiaries: Those intended to benefit from the activity		
4. Other stakeholders, who are expected to have an interest in the activity (e.g. independent media, civil society organisations, researchers, members of parliament, government bodies, corporate interestsâ?l) [â?others to be identified]		
3: Process		Priority (binary/ranked/rated)
1. Time is allowed for the iterative development of evaluation questions I.e questions are proposed, responded to, refined, and then agreed on. At least one cycle. [â?others to be identified]		
[Other content yet to be provided here]		

2.4 Quality criteria for evaluation questions

Criteria	Description	Priority (binary/ranked/rated)
1. Ownership	1. The evaluation questions have some identifiable owners, who want to know the answers to those questions	

2. Usefulness

1. The question owners, and/or others, can envision how they would use the evidence in response to each question (see USAID ref. above)

3. Feasibility

1. It is feasible to answer an evaluation question given what is known about:
 1. likely data availability and quality,
 2. the time and resources available to an evaluation team
2. The questions are feasible to answer given the current stage of the program/policy cycle
3. It is possible to obtain an answer to the question ethically and respectfully
4. The question can be answered in a timely manner, i.e., before any decisions potentially influenced by the information will be made

4. Relevance

1. The questions clearly relate to the intentions of an intervention, or its counterfactual (its absence).
2. Evaluation is the best way to answer the question, rather than some other (non-evaluative) process.

5. Prioritisation

1. The relative priority of different questions or groups of questions is clear.

6. Uncertainty

1. See [2.5 Unresolved issues](#) below

7. Writing style

1. Only one question is asked per sentence
2. Indented questions are used where follow-on questions can then be appropriate., once previous question is answered
3. A mix of open and closed ended questions is used.
 1. Closed ended questions describe specific claims or hypotheses that need to be tested.
 1. All interventions have some testable claims
 2. Open ended questions are used where no specific expectations about what will be found, or where it is inadvisable to disclose those expectations before the beginning of an evaluation
 1. But where so, it can still be useful for those to be documented beforehand

[Others to be identified]

[Other content yet to be provided here]

2.5 Unresolved issues – your views please

1. **Uncertainty:** For any given activity, people's views may vary as to whether it has been effective or not (or how it rates on any other criteria of interest). If an evaluation does not have the time and resources to examine and test every claim they have to make choices as to which to focus on. Should they focus on claims that have a wide degree of support e.g. 95% of x people think it is true, or should they focus on claims that are more evenly disputed e.g. where 50% think the claim is true and 50% who do not?
 1. These differences of opinion may be evenly scattered across different stakeholder groups, or they may be associated with particular stakeholder groups. Should more attention be paid to differences of opinion that are of the second kind, associated with identifiable stakeholder groups?
 2. PS: The whole idea of focusing on testing claims, versus just making open ended inquiries, is argued in detail here: Davies, R (2012) [Evaluation questions: Managing agency, bias and scale](#)
2. **Relevance.** This is the first of the six DAC criteria. But I have my doubts about whether this is the best word to use. The first para that explains the concept says *“The extent to which the intervention objectives and design **respond to beneficiaries**, global, country, and partner/institution **needs, policies, and priorities**, and continue to do so if circumstances change.”*• [my emphasis added]. In short, I think it is how an activity fits with peoples/organisations **preferences** (expressed one way or another)

3. **Valuative/Normative, or?:** Not sure what the best summary term is to describe questions about peoples' assessments of the value and significance of what happened? Any advice?
4. **EQ workflow:** Does it make sense to think about using different types of evaluation questions at different stages of a workflow? For example, as suggested in Figure 1 below [bearing in mind that in reality there are likely to be various feedback loops between these stages]:

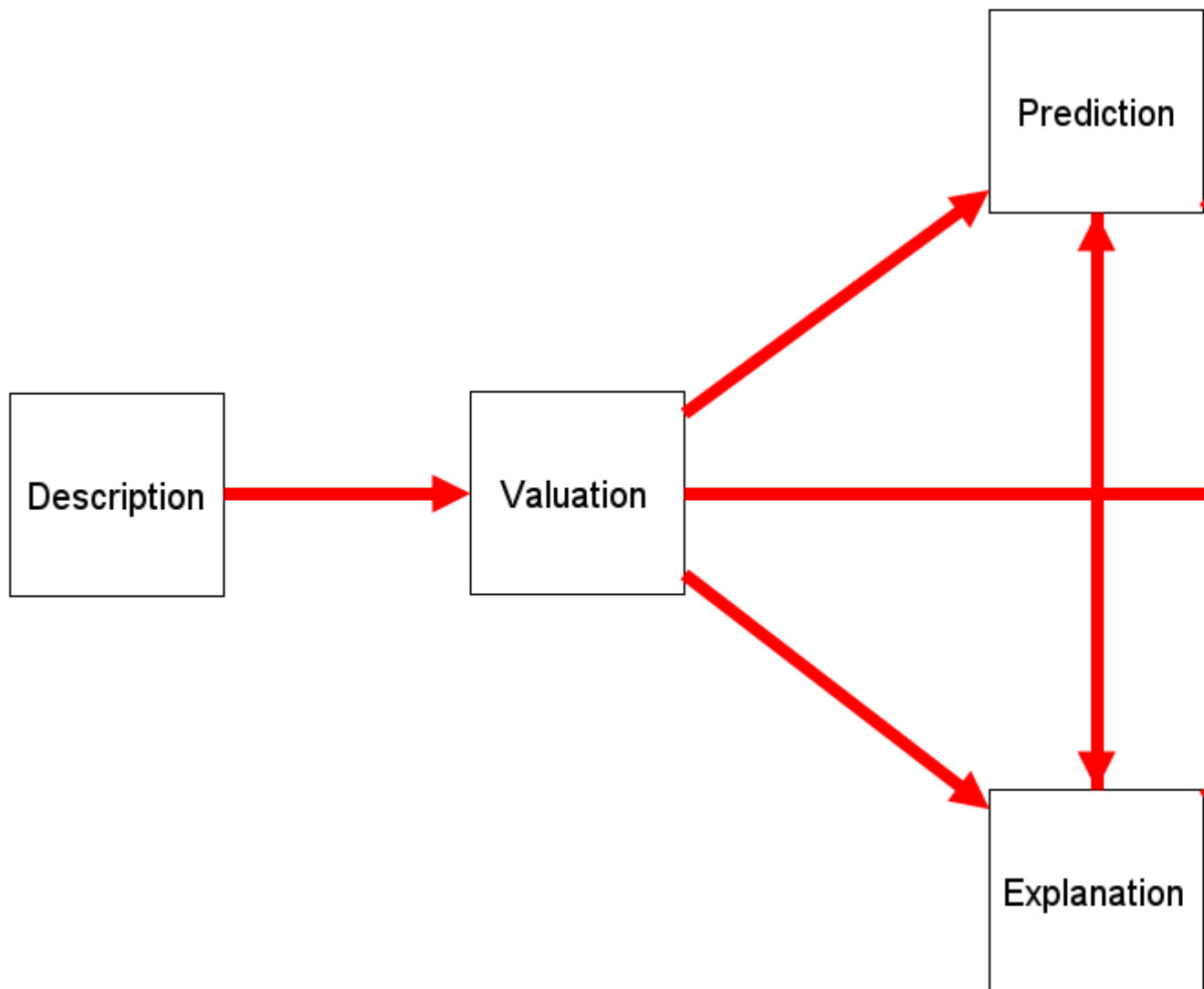


Figure 1: A hypothetical workflow involving different kinds of evaluation questions

Post script: Courtesy of [Tom Aston](#), I found this other diagram that suggests that valuation might better come first, before description (but I am not persuaded)

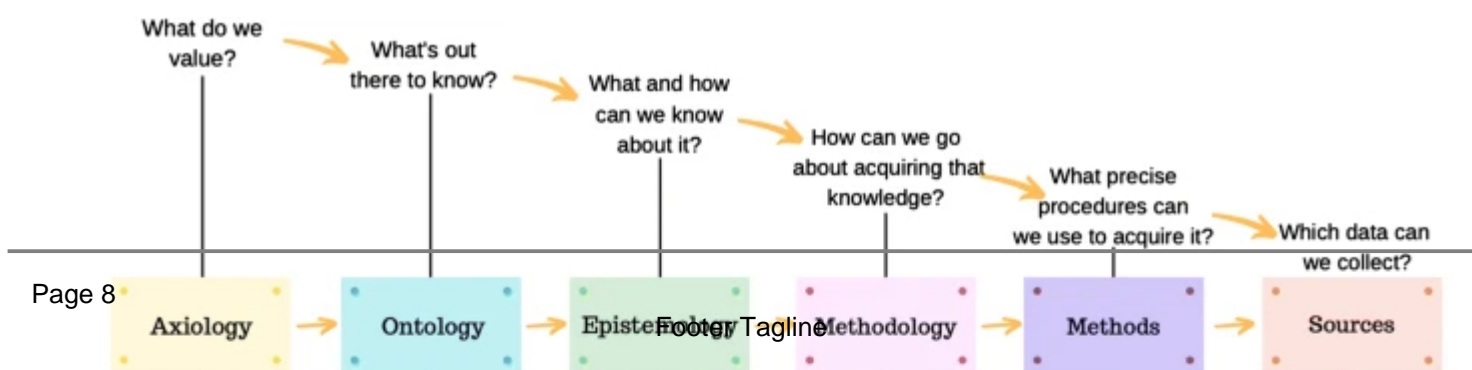


Figure2: From Brown, M. E. L., & Dueñas, A. N. (2020). A Medical Science Educator's Guide to Selecting a Research Paradigm: Building a Basis for Better Research. *Medical Science Educator*, 30(1), 545-553. <https://doi.org/10.1007/s40670-019-00898-9>

3. Endnote: Where Evaluation Questions fit in with other considerations

Evaluation questions should not be thought about in isolation. This diagram below is taken from the Austrian government's [Guidance Document on Evaluability Assessments](#). What stakeholders want to know needs to be answered, or can be answered. They need to fit, to some extent, with the Theory of Change (about what was intended), and with current and potential availability of data (about what can be known).

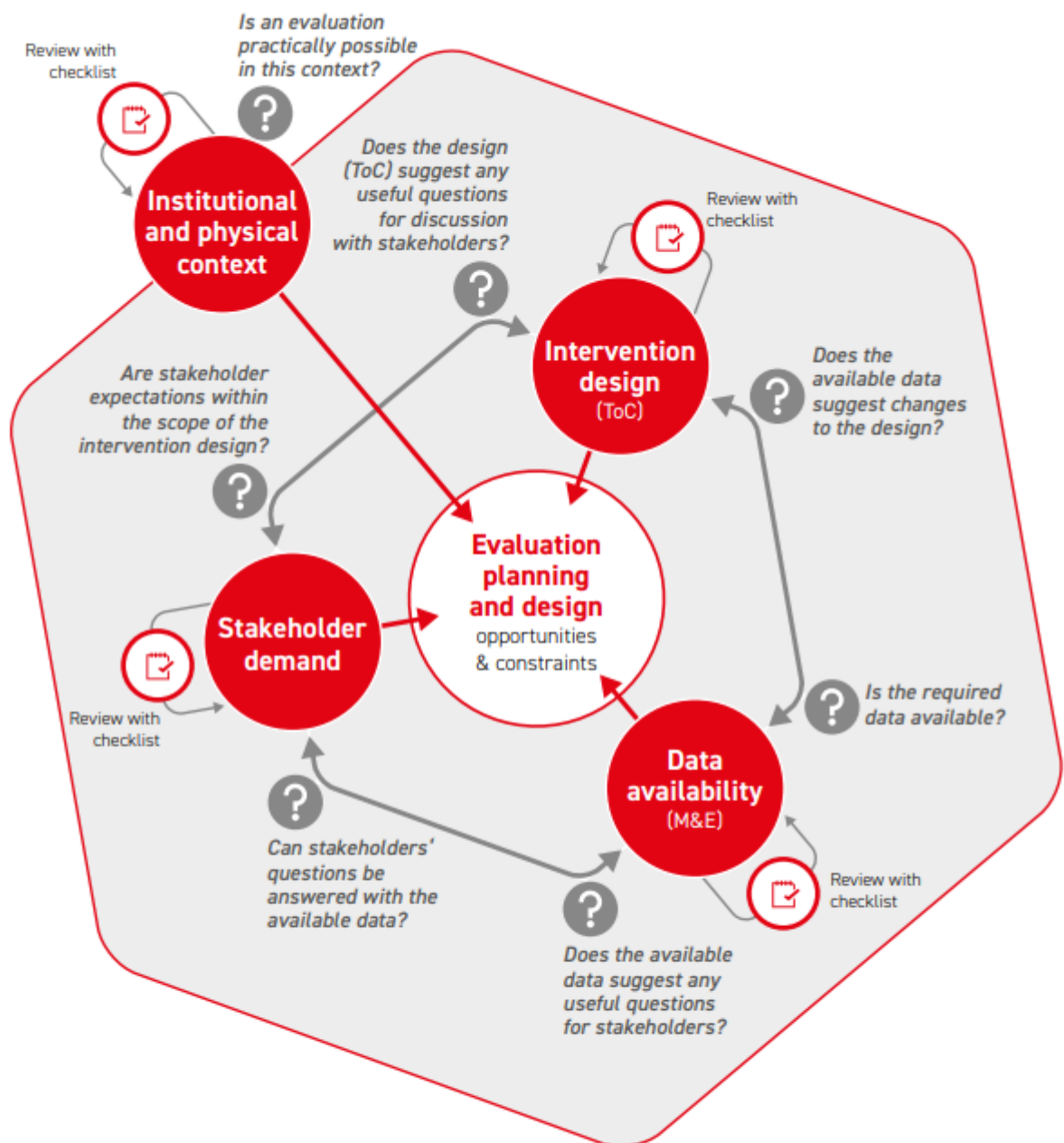


FIGURE 1: The four facets of evaluability and their associated checklists. Adapted illustration from Davies, 2013.

Date
16/09/2025
Date Created
21/01/2022

