

| Key Quality Issues and Concerns in Qualitative Evaluation <i>(Based on Spencer et al, 2003 pages 71-72)</i> | |
|---|---|
| <i>Underlying concern</i> | <i>Ways of ensuring quality (Examples)</i> |
| <i>The defensibility of the approach</i> | <ul style="list-style-type: none"> A clear logic of enquiry Clarity of questions posed Rationale for questions Responsiveness to real life context Fitness for purpose |
| <i>The rigour of conduct</i> | <ul style="list-style-type: none"> Collection of in-depth data Careful recording of data Contextual documentation Systematic and thorough analysis Explication of conceptual and analytic process Auditable documentation |
| <i>The relationship of the researcher to the researched</i> | <ul style="list-style-type: none"> Ethical behaviour (e.g. gaining consent) Involvement of participants in study Reflexive awareness of investigators' role Open and empathetic fieldwork skills Recognition of different subjective perspectives |
| <i>The credibility of claims</i> | <ul style="list-style-type: none"> Triangulation Validation by informants/respondents Peer review Consideration of alternative explanations & negative cases Balanced presentation of evidence Demonstrating links between data and conclusions |
| <i>The broader contribution of the study</i> | <ul style="list-style-type: none"> Relevance and utility to policy Ongoing involvement of potential users in planning and discussing recommendations Timeliness Clear reporting and active dissemination Linking findings to broader research and theory |

An assessment checklist at the proposal stage

- Have impacts been identified and understood?
- Are stakeholders going to be involved in validating these impacts?
- Has existing knowledge about this kind of programme, including ToCs, been taken into account?
- Are programmes purposes understood and evaluation questions clearly stated?
- Has the proposal shown how IE design is able to link cause and effect and answer evaluation questions?
- Is the proposed design consistent with programme attributes and the simplicity or complexity of the programme?
- Is the timing of the IE consistent with the likely trajectory of intended change?
- If the programme is complex are the proposed methods able to disentangle more than one cause?
- Are proposals putting forward measurement of impacts consistent with the kind of programme data available and collectable; and the designs and methods to be used?
- Have protocols and methodological guidance, where these exist, been cited and used
- Are examples of work by members of the proposal team that uses similar methods and designs been provided.

| Drawing up Terms of Reference for Impact Evaluations | | | |
|---|--|---|---|
| Design issues | Specific questions | Rationale | Implications for ToR |
| Identifying impacts | How should programme impacts and effects be identified? | Conceptualising & identifying impacts is difficult; and sometimes data is unavailable. When to assess impacts and which impacts affect whom, are also design issues. Stakeholders' participation helps identify valid impacts | Proposers should indicate how they understand and will identify impacts – including impacts for different groups. Commissioners should indicate data availability problems. |
| Building on what is known | Is there already substantial knowledge about how these kinds of programmes work, perhaps a credible Theory of Change? | If much is already known there might both be risks of duplication and waste; and advantages building on existing knowledge | Proposers should demonstrate familiarity with current state of evaluation/research knowledge and indicate how this will shape their use of Theories of Change |
| The overall purpose of the evaluation | What kind of use for whom is envisaged – demonstrating past effectiveness; scaling-up and replication; improvement; learning for future policy and practice? | Purposes of IE may differ. It is important to identify main purposes as this determines evaluation questions and choice of methods able to answer these questions. | Proposers should be expected to discuss how overall purpose connects with evaluation questions – and show an awareness of design and method implications |
| Programme attributes, scale and complexity | Is the programme made up of a single intervention or several? What is the programme 'architecture'? | Programme attributes constrain the choice of IE designs and methods. Multi-level or decentralised programmes offer opportunities for nested designs | Proposers should be asked to demonstrate understandings of programme attributes and the implications for designs and combinations of designs. |
| Context and contribution | How important is context and how far are different causal and contextual factors likely to influence impacts? | Programmes that are open to multiple influences – complex, embedded rather than simple and self-contained -will need to focus on the contribution of programme interventions rather than attribution | Proposers should be asked to discuss the programme context including the importance of multiple causal factors; and how this relates to a contribution or attribution focus |
| Measurement and extent | Does the IE set out to measure how much of an impact a programme has had – and is this feasible? | Sometimes possible to assess contribution but not extent (how much). Numbers will determine the possible statistical designs and methods. | If appropriate, proposers should be asked to discuss their approach to measurement and extent |

| How an IE is conducted: Standards and Criteria over the Evaluation Life-cycle (Reliability, Robustness and Transparency) | |
|---|--|
| <i>Choice of designs & methods</i> | Are designs and associated methods put forward that are established, well documented and able to be defended? Do the chosen designs take into account Evaluation Questions and intervention attributes? |
| <i>Reliability</i> | Are they able to explain how an intervention contributes to intended effects for final beneficiaries? Do the EQs allow for success and failure (positive and negative effects) to be distinguished? |
| <i>Proper application of designs and method</i> | Are the ways that designs and methods are applied clearly described and documented? Does the application of designs and methods and subsequent analysis follow any protocols or good practice guidelines? Is the evaluation team knowledgeable about the methods used? |
| <i>Robustness</i> | |
| <i>Drawing legitimate conclusions</i> | Do conclusions clearly follow from the findings? Has the evaluation explained the effects of the programme? How are evaluative judgements justified? |
| <i>Transparency</i> | Have stakeholder judgements been taken into account when reaching conclusions? Are the limitation of the evaluation and its conclusions described? |

Technical standards for IE designs and methods

The second part of a framework is concerned with ‘Technical Standards’ for methods and designs that mainly relate to what is usually labelled validity and rigour. These standards specifically concern IE. They ask how the designs and methods chosen address the three main characteristics of IE following the definitions and discussions contained in this report. We have argued that IE should:

- Address the contribution made by an intervention
- Provide a clear causal link to effects
- Offer an explanation of how a programme worked

For each of these categories a set of criteria are suggested as generic questions that could be asked of any IE design.

| Technical standards and criteria to judge the quality of IE designs and methods (Validity and Rigour) | | |
|---|--|--|
| <i>Contribution</i> | <i>Explanation</i> | <i>Effects</i> |
| <p>Is the design able to identify multiple causal factors?</p> <p>Does the design take into account whether causal factors are independent or interdependent?</p> <p>Can the design analyse the effects of contingent, adjacent and cross-cutting interventions?</p> <p>Are issues of ‘necessity’, ‘sufficiency’ and probability discussed?</p> | <p>Does the evaluation make it clear how causal claims will be arrived at?</p> <p>Is the chosen design able to support explanatory analysis (e.g. answer how and why questions)?</p> <p>Is theory used to support explanation? (e.g. research-based theory, Theory of Change) If so how has theory been derived?</p> <p>Are alternative explanations considered and systematically eliminated?</p> | <p>Are long term effects identified?</p> <p>Are these effects related to intermediate effects and implementation trajectories?</p> <p>Is the question ‘impact for whom’ addressed in the design?</p> |
| <p>Please attach any protocols, guidelines or quality assurance systems used in connection with this design</p> <p>Please also provide previous reports or publications that illustrate how this design has been used previously for IE purpose.</p> | | |

JUDGING CONCLUSIONS AND RECOMMENDATIONS

The strength of conclusions depends on various factors, for example:

- The soundness of the IE design
- The way that design and associated methods were implemented – which is why transparency and ‘auditability’ of methods and data are important
- Consistency between the conclusions drawn and the evidence base and designs on which these conclusions are based – for example a theory-based design cannot on its own quantify impact; nor can a counterfactual based design on its own predict what might happen in a different setting
- The scope of evidence: what kinds of judgements does evidence support – e.g. the IE of a specific programme cannot be used to judge an entire class of similar programmes across different settings
- The judgement of the evaluators – conclusions rely on judgement as there is rarely an automatic link between evidence and conclusions; hence the importance of evaluators making their criteria and often their values explicit.
- Evaluators acknowledging the limitations of all designs and methods; and the innate difficulties of going beyond probability and plausibility.

Different commissioners have different expectations of how far evaluators should go in making recommendations. Policy commissioners often take the view that evaluators are not sufficiently knowledgeable about policy contexts to make sensible recommendations.

A sensible middle-ground is to expect evaluators to put forward recommendations, based on ‘sensemaking’ discussions and workshops with key users of findings, even though commissioners may subsequently need to situate these recommendations into a wider body of organisational or policy knowledge.

Consideration of the validity of recommendations should take account of:

- The connection between recommendations and conclusions
- The strength of evidence that fed into conclusions
- The criteria and values used to justify conclusions
- The input of stakeholders into a validation process
- The extent to which conclusions are supported by a more extensive evidence base, e.g. from previous evaluations, syntheses and research

Reviewing Reports: Checklist

When substantive reports that include findings are produced it is worth revisiting the TOR framework and proposal assessment checklists in the first instance. However a checklist at this stage can be more focussed. For example commissioners and managers should ask:

- Does the report make it clear how causal claims have been arrived at?
- How have different types of theory been used - testing programme assumptions or building on wider research? Has new theory been developed?
- Is the report clear about when and where impacts can be observed?
- Does the report convincingly identify contextual and causal factors and take them into account?
- Is the chosen design able to support explanatory analysis (answering how and why questions) if this was required?
- Is there a consistent link between evaluation questions asked, overall design, data collection and analytic methods used?
- Have alternative explanations that do not depend on programme effects been considered and systematically eliminated or accounted for?
- Have beneficiaries and other stakeholders been involved in scoping the evaluation and validating and interpreting results?
- Are the ways methods were applied and data collected clearly described and well documented?

A positive answer to these questions makes an IE reliable and defensible. They incorporate some of the key elements that a researcher would call validity, robustness, rigour and transparency. A final judgement on these qualities cannot only be up to commissioners; they also require third party peer reviews.