CEDIL Guidelines for good impact evaluation practice

Introduction

These guidelines are directed to policy practitioners, commissioners and researchers to inform the implementation, assessment, and understanding of impact evaluations. The largely draw on a CEDIL inception paper on conducting successful impact evaluations (Masset et al., 2018).

This note is one of a set of guidelines that CEDIL is piloting to improve the design, implementation and use of evaluations. They will be revised based on lessons identified during the design and implementation of CEDIL-financed research projects. We welcome your feedback, which should be sent to cedil@opml.co.uk

There are three key elements of good practice in evaluation:

- **Reliability:** answering evaluation questions in a credible and convincing way
- **Relevance:** providing evidence that is valuable for policy making
- **Use:** offering research findings that can be used by policymakers

These elements encompass both "internal" and "external" validity of evaluations. Reliability is a broader term for internal validity, while external validity is a component of relevance. Internal validity refers to unbiased estimation of causal effects of interventions. External validity refers to the ability to extrapolate results, by transferring them to other contexts or at a larger scale. The table below shows what action are needed to ensure that these three key elements are built in evaluation practice and their goals.

Key elements	What is required	Goal
Reliable evidence	Answering the evaluation questions in a credible and convincing way	• Provide causal estimation of the effects of programme
Relevance	Provide evidence to answer important, current policy questions	• Addressing priority development questions • Ensure results are scalable and transferable • Being aligned with timing of policy
Use	Research findings that can be used by policy-makers	• Leading to policy change

Impact evaluations are articulated in three stages: design, implementation, and analysis and presentation of results. Key factors specific to each stage are briefly discussed below, followed by some cross-cutting factors that do not belong to any specific stage. The same factors are also illustrated in Figure 1.

Design and planning

1. *Formulating the evaluation question*: evaluation questions should be formulated to solve a social problem or to inform the scalability and transferability of existing interventions
2. *Theories of change* explain how a project is expected to work and map out all pathways to impact that can be anticipated
3. *Systematic reviews* and other synthesis products, such as evidence gap maps, meta-analyses and critical reviews, should be used to inform research questions and evaluation design
4. *Fitness for purpose* identifies appropriate evaluation methods for specific evaluation questions
5. *Evaluability* assesses the technical, political and practical feasibility of an evaluation before it starts
6. *Piloting* an intervention provides an idea of project take-up and effect sizes
7. *Pre-analysis plans* pre-specify the analysis to be conducted once data are obtained and support the evaluation design as well as promoting transparency and accountability
8. *Team's composition* requires having the right set of technical skills and disciplinary approaches, including members with knowledge management skills to disseminate findings and for engaging with stakeholders
9. *Stakeholder engagement.* Project implementers, funders and users should be involved from the design stage of an evaluation. Effective engagement includes: understanding users' needs, facilitating access to evidence, building decision makers' skills, and fostering changes to decision-making processes (Oliver et al., 2018).

Implementation

1. *Project take-up* measures the scale of participation in the intervention
2. *Is there contamination of the control group* with subjects in the control areas accessing the activities from the intervention?
3. *Biases* at the implementation stage include attrition of study participants, non-compliance, and placebo effects
4. Check for *sample size and data quality* to ensure that studies are able to measure impacts of the desired size
5. *Ethical approvals* to protect the confidentiality of the information and prevent any harm on the subjects being studied
6. *Political Economy* of local governments and implementers must be verified to clear the risk that the evaluation or particular aspects of their design is opposed
7. *Ability to adapt* needs to planned and verified to prepare for changes in circumstances requiring adjustments in the design or implementation

Results

1. *Analysis.* An impact evaluation must provide the best possible estimates of project impact
2. *Outcomes.* The selection of meaningful outcome indicators, articulated in a pre-specified analysis plan to guarantee that the results are conclusive and credible
3. *Dissemination.* The results of a study need to be adequately disseminated in order to reach the desired audiences. Traditional dissemination methods include policy briefs, workshops, and publications. More recently, *knowledge brokers* have been charged with the role of disseminating evidence among policy-makers (Brown et al., 2018).
4. *Timeliness.* The findings of evaluations should be reported as soon as they are available in order to support decisions for modifying or discontinuing programmes. Since the results of evaluations are often available only after project completion, continuous engagement and intermediate products on emerging findings are needed.

Cross-cutting factors

1. *Transparency:* An evaluation is more likely to follow best practices when activities are open to the scrutiny of the public and of other researchers and when the data and codes used in the analysis are publicly available
2. *Shocks and external events.* Project implementation and evaluations can be derailed by unexpected shocks, therefore a risk assessment at the planning stage and the formulation of contingency plans are helpful, particularly when researchers work in difficult contexts. Buffer time should be built in work-plan for unexpected events and communicated to funders and other stakeholders.
3. *Quality Assurance and peer-review groups.* Progress reviews during the evaluation, and peer-reviews ensure the quality of the design, implementation and analysis
4. *Independence.* Evaluators should be independent of programme managers and funders to ensure the credibility of the study.
5. *Equity.* Study design, data collection, analysis and reporting, should consider vulnerable groups and relevant disaggregations of the population. In particular, gender and disability are a primary concern.

Figure 1: Determinants of good evaluation practice
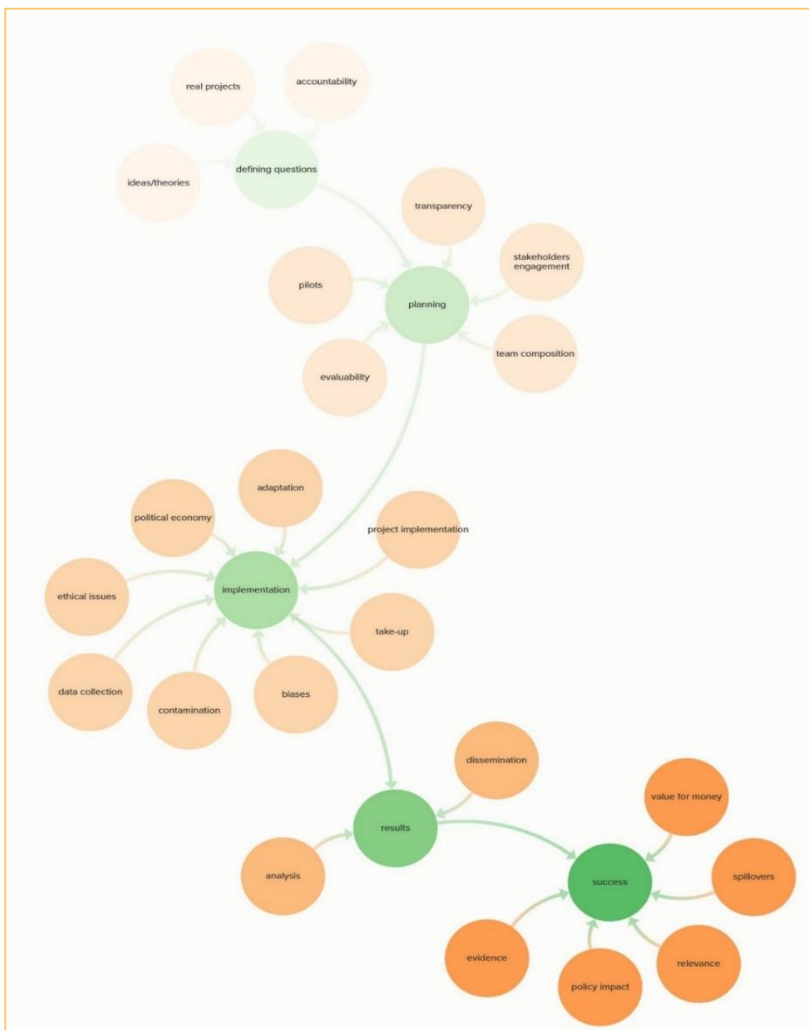
How to use these guidelines

These guidelines help identifying whether the required elements of a successful impact evaluation are present. It can be used as a reference tool by researchers to guide their work at different stages of their evaluations, or by reviewers and commissioners to monitor progress. Note that evidence use represents a fundamental component of the success of an evaluation study, but recommendations in this regards are presented in other CEDIL checklists on policy influence and stakeholder engagement.

The checklist

Design document	
What is the aim of the evaluation? *(continuing/ discontinuing an intervention, improving an existing intervention, scaling up or implementing an intervention elsewhere, informing general policy debates, contributing to scientific knowledge)*	
In what way are the evaluation questions policy relevant? *(the results can be used to improve living standards of a sufficiently large population or group)*	
Was an evaluability assessment conducted?	
Was a pilot or other formative assessment conducted?	
Is the theory of change well defined? *(including a diagram or a set of equations explaining how the intervention will affect the outcomes, a narrative of the causal links between factors, and a discussion of assumptions and supporting factors)*	
Does the evaluation draw on systematic reviews and other evidence synthesis work on the same subject?	
Does the research team include members with different backgrounds and skill sets?	
Are the evaluation methods proposed appropriate to answer the evaluation questions?	
Is the sample size sufficiently large to detect small project effects?	
Does the evaluation include an assessment of the main risks it will face?	
Is there a credible stakeholder engagement and policy influence plan?	
Has the research team engaged with potential users about their needs from the evaluations as part of the evaluation design?	
CEDIL supports studies developing and testing innovative approaches to impact evaluation. What is the innovation in methods or measurement being developed or tested in your study?	
Is this innovation clearly described so it may be adopted and tested by others?	
Is the study set out to identify impacts for specific groups? *(The study should identify **vulnerable groups** and the relevant equity disaggregations. Design, data*	

Design document	
collection, analysis and reporting should address evaluation of results by **gender and disability** *– guidelines on how to address disability can be found on the website)*	

Transparency of research	
Was the study approved by an institutional review board (IRB)?	
Is the evaluation registered?	
Is a pre-analysis plan available?	
Will data and codes for the replication of results be available to the public?	
Will the public be able to openly access the results of the evaluation?	

Baseline report	
Are the outcomes balanced between the project and the control groups? *(this refers more generally to any potential bias in the data including for qualitative studies)*	
Are the outcome data of sufficient quality? *(representativeness, missing values, variances, biases, and measurement error)*	

Midterm report	
Was the intervention taken up by a sufficiently large number of beneficiaries?	
What is the rate of attrition?	
Are spill-over effects or contamination of the control group likely?	
Has the evaluation the support of project implementers and project funders?	
Are any changes to design required by changes in implementation circumstances?	

Final report	
Does the method of analysis provide the most accurate and unbiased estimates of project effects?	
Are results correctly reported? *(effect sizes as well as statistical significance, full reporting of all proposed analysis)*	
Are impacts on intermediate outcomes reported?	
Are results disaggregated by gender and other relevant characteristic?	
Are project costs reported at a sufficient level of detail?	
Are the conclusions sufficiently clear? *(include findings on specific project implementation, project modification, project scaling up, implementation in other context, overall significance)*	

Final report	
How do the results transfer to other settings or how can they be scaled up within the same context?	
Was the study reviewed by an independent advisory group?	

References

Brown, J., Cartridge, R., Davies, S., ul Hassan, E., Hsia, C., Lal, T., Mackison, D., Menon, R., Moratti, M., Murunga, V., Le Nestour, A., Whitehead, L., and H. White (2018), Research Uptake in Policy-Making: from Papers to Policy, *CEDIL Inception paper No. 14*: London

Oliver S, Roche C, Stewart R, Bangpan M, Dickson K, Pells K, Cartwright N, Hargreaves J, Gough D (2018), Stakeholder Engagement for Development Impact Evaluation and Evidence Synthesis, *CEDIL Inception paper No. 3*: London

Masset E, Rathinam F, Nath M, Vigneri M, Wood B (2019), Successful Impact Evaluations: Lessons from DFID and 3ie. *CEDIL Inception Paper No. 6*: London