

A synthesis of international standards for impact evaluation practices



A contribution to the 2030 Agenda and the



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

Contents

Acronyms.....	ii
1. Introduction.....	1
2. Definition and Basic Framework for Standards	2
General Consensus.....	2
Ongoing debates.....	3
3. Quality control Criteria and Standards for Impact Evaluation	5
4. Developing a Technical Concept.....	8
Criteria for selecting programmes for impact evaluation	8
The main components of the technical concept paper/proposal	8
5. Accessing the Industry's Standard Guidance	10

Acronyms

DFID	Department for International Development
DME	Design Monitoring and Evaluation
ECG	Evaluation Cooperation Group
IE	Impact Evaluation
IOCE	International Organization for Cooperation in Evaluation
NGO	Non-governmental Organisation
OECD/DAC	Organization for Economic Co-operation and Development's Development Assistance Committee
ToR	Terms of Reference
UNEG	United Nations Evaluation Group

I. Introduction

Within the realm of evaluation practices, impact evaluation is considered one of the most complex and demanding evaluation exercises. It involves extensive use of advanced social research methods and techniques aimed at establishing causal relationships. Industry standards on impact evaluation have been agreed through a global dialogue led by NONIE: a Network of Networks for Impact Evaluation includes the Organization for Economic Co-operation and OECD/DAC Evaluation Network; the United Nations Evaluation Group (UNEG); the Evaluation Cooperation Group (ECG); and the International Organization for Cooperation in Evaluation (IOCE). Guidelines and standards have been extensively used and gradually improved to facilitate better impact evaluation practices by the development community. These same guidelines and standards form the basis of this document.

These guidelines provide impact evaluation practitioners with a quality control tool (standards) to assess impact evaluation initiatives. It also provides access to valuable reference material guiding the development community's best practices for conducting impact evaluation, including recent innovations on methods and techniques.

Impact evaluation should be used when a programming approach and program theory is mature. It is a well-established programming approach, for which there already is extensive outcome evaluation data from programs that suggests that the approach has merit. Impact evaluations are costly and they should always be implemented with an external research partner.

2. Definition and Basic Framework for Standards

Impact evaluation involves a specialized exercise. It incorporates advanced social research methods and techniques to establish causal relationships in an attempt to assess, with high levels of scientific rigor, the attribution of impacts to interventions. This is usually done by using a control group to compare with the intervention group.

Impact evaluations are not undertaken for all programmes approaching a particular issue, but rather for a small sample of programmes to examine the specific impact of a particular type of intervention. In order to ensure that impact evaluations are impartial and meet international standards, such evaluations are usually undertaken by an academic partner or research institution.

General Consensus

Definitions of impact evaluation vary depending of the nature and scope of the evaluation inquiry. However, with minor modifications or additions since its conception, the DAC definition of impact continues to drive the conceptual foundation for defining impact evaluation within the development community. The DAC defines impact as: “Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended”.¹

- When applying the DAC definition for impact evaluations, there are two underlying concepts:
Attribution: the words “effects produced by” in the DAC definition imply an approach to impact evaluation that is about attributing impacts to interventions, rather than just assessing what happened.
- **Counterfactual:** It follows that in most contexts, knowledge about the impacts produced by an intervention requires an attempt to gauge what would have occurred in the absence of the intervention and a comparison with what has occurred with the intervention implemented.²

Complementary initiatives have confirmed this general consensus and provide additional insights that reveal common elements within the vibrant dialogue around theoretical and methodological dimensions of impact evaluation.

- **A Theory of Change** approach has become accepted as a basic foundation for most types of impact evaluations; there is a strong focus on tracing cause and effect to demonstrate whether or not an intervention has produced results; and
- **A Mixed Method** approach utilizing quantitative, qualitative, participatory and blended methods is now widely accepted as advisable to address the types of interventions that are now predominant in development initiatives³.

How does an impact evaluation differ?

In World Vision, outcome evaluations are standard practice for all programmes. Such evaluations provide findings that reveal the contribution of programs or projects to changing societal conditions directly related to the well-being of children, including their families and communities by comparing change in indicators, pre and post intervention. Outcome evaluations mostly use mixed methods and also assess the theory of change or programme theory. However, outcome evaluations are not as rigorous as impact evaluations and cannot measure attribution, as they do not utilise a counterfactual.

¹ OECD, 'Glossary of Key Terms in Evaluation and Results Based Management,' (2010) in English, French & Spanish

² NONIE, 'Impact Evaluations and Development: NONIE Guidance on Impact Evaluation' (2009)

³ United Nations Evaluation Group-Impact Evaluation Task Force, 'Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management' (2013), Section 4 page 8

Impact evaluations differ from outcome evaluations by providing insightful learning derived from rigorous examination of the effects produced by a development intervention through a “cause and effect” approach that specifically addresses the issue of attribution, which incorporates the concept of contribution. Impact evaluations are able to measure attribution by using higher levels of scientific rigor, through the use of a counterfactual to provide a comparison.

The DAC defines attribution as the “ascription of a causal link between observed (or expected to be observed) changes and a specific intervention”. It further clarifies that: “attribution-based definitions of impact do not require that effects be produced solely or totally by the intervention. They anticipate the co-existence of other causes, so that the intervention will have contributed to the demonstrated effects.” It also highlights that DAC’s definition of impact specifically includes the possibility of partial attribution, or contribution, through its inclusion of secondary and indirect effects.

Furthermore, impact evaluations can be used to validate an organization’s theory of change and logic chain in a rigorous way. Do the assumptions about the project plans hold true? Have inputs led to outputs, and those outputs to outcomes? More rigorous evaluations can provide an evidence base to inform changes to policy and programme approaches.

Attribution
Causal Link between
observed changes and a
specific intervention

Ongoing debates

There are ongoing debates in the industry regarding the specific research methods used when performing impact evaluations.

For example, should impact evaluations use:

- Experimental or quasi-experimental designs
- Quantitative and/or qualitative methods
- Experimental design or statistical controls to formulate comparison between the “factual” and a counterfactual.

The debate is centred on whether or not a counterfactual based on qualitative or simple statistical analysis has the required scientific rigor claimed by quantitative social science methods.

The UN initiative, properly notices that both approaches are not mutually exclusive and suggest that impact evaluation could benefit from using mixed methods “at certain points”.

- “Causal pathways from outputs to impacts, which can be fairly straightforward or more complicated, and effects that become manifest relatively quickly or over longer timeframes;
- Different levels of analysis: national, institutional, community, household, etc.;
- Different types of interventions that require tailor-made approaches to assess impact (ranging from administrative reform, support to national legislation, to farmer subsidies and humanitarian aid).

Given the above, the focus of an impact evaluation can differ widely from one evaluation to another; correspondingly, there may be substantial variation in the mix of methods applied in the evaluation through which the ‘why’ and ‘how’ of an intervention can be explored, and that also may capture the form and extent of indirect and secondary effects”.

Mixed methods

Complementary initiatives have contributed to further substantiate the viability of such methodological integration. The mixed methods approach for impact evaluation integrates social science disciplines with predominantly quantitative and predominantly qualitative approaches to theory, data collection, data analysis and interpretation. Integration is expected to strengthen the reliability of data, validity of the findings and recommendations. It is also expected that by combining methods and techniques our understanding of the processes through which program outcomes and impacts are achieved, including how these are affected by the contextual factors is further improved⁴.

The following section provides a quality control and standards tool, reflecting the industry's conceptual and methodological framework presented and discussed in this section.

⁴ Bamberger, Michael, 'Introduction to Mixed Methods in Impact Evaluation', Inter-Action/Rockefeller Foundation, (2012)

3. Quality control Criteria and Standards for Impact Evaluation

Within the development industry, impact evaluations are expected to objectively reveal the attribution resulting from an intervention aimed at improving condition for sustainable development. Evaluation findings are, therefore, highly sensitive and it is expected to withstand high levels of scientific scrutiny. Such demanding expectations require observance of generally agreed standards and quality control criteria.

The following checklist helps determine whether or not an impact evaluation fulfils such expectations from the design stage to the use of results stage. Sound impact evaluation initiatives are expected to be able to display a checkmark to all of the items listed in the checklist, including specific page numbers where the standard is applied.

Done 
To do 

Checklist ⁵				
Institutional strategies, priorities, guidelines, and principles				
Organization's internally established parameters for Evaluation and Research practices		✓	—	If ✓, page #
1.	In line with the organization's evaluation and research strategy and priorities			
2.	In line with the organization's research guidelines and definitions			
3.	Research standards observed and signed off by appropriate technical lead			
4.	In line with the organization's ethical principles for conducting evaluation and research			
5.	Organization's Ethics Committee Review signed off			
6.	If plan to publish, in line with the organization's standards for publishing research findings			
7.	Partnerships with Academic or Research Institution should comply with the organization's principles and guidelines			

⁵ This checklist reflects the minimum impact evaluation standards required for impact evaluation initiatives. The technical lead's criteria to substantiate the selection of specific methods and techniques and the thematic content of the impact evaluation, 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 among institutional strategies, priorities, and principles, including design, implementation and utilization; and conclusions should also be assessed by the respective technical teams.

Participation, contextualization and ethical considerations ⁶		✓	–	If ✓, page #
Consultation				
8.	Stakeholders ⁷ have been actively involved and consulted during the process of formulating IE questions			
9.	If available, governmental or community-based administration and information systems have been consulted and extensively used			
10.	Arrangements involving local partners in the evaluation process have been considered			
Contextualization				
11.	IE has been contextualized and previously conducted IE has been consulted and discussed according to such context			
12.	The IE is contributing to evaluation capacity building at the local level			
13.	The IE is contributing to inform the organization's policy making, donors, and partner agencies			
Ethical Considerations				
14.	Evaluators have made explicit their interests and values as they relate to the intervention under evaluation			
15.	Arrangements have been established to monitor and remedy bias or lack of impartiality			
16.	Measures have been established to ensure confidentiality and to avoid risks to informants			
17.	Feedback and validation procedures that involve stakeholders have been established and used			
Validity and Rigour of Design				
Contribution		✓	–	If ✓, page #
18.	The design is able to identify multiple causal factors			
19.	The design takes into account whether causal factors are independent or interdependent			
20.	The design analyses the effects of contingent, adjacent and cross-cutting interventions			
21.	Issues of 'necessity', 'sufficiency' and 'probability' are discussed			
Explanation				
22.	The evaluation make it clear how causal claims will be arrived at			
23.	The chosen design is able to support explanatory analysis (e.g. answer how and why questions)			
24.	The organization's Theory of Change and sectors' specific assumptions are used to support explanation			
25.	Alternative explanations are considered and systematically analysed			

⁶ Some of these parameters are covered by many organizations' 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 impact evaluation. For general agreement on standards and specific parameters regarding quality control criteria by the international development community see:

- Department of International Development, DFID, 'Broadening the Range of Designs and Methods for Impact Evaluations' (2012), Chapter 6
- United Nations Evaluation Group, 'Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management' (2013), Section 4

⁷ Some examples of stakeholders: key staff related to the intervention, target group members, policy makers, evaluators, commissioners, NGOs, and grass roots/civil society organizations.

Effects				
26.	Long term effects are identified			
27.	Effects are related to intermediate effects and implementation trajectories			
28.	The question 'impact for whom' is addressed in the design			
Technical Implementation and Utilization				
Choice of design and methods		✓	—	If ✓, page #
29.	Designs and associated methods are established, well documented and able to be defended			
30.	The chosen designs take into account the Evaluation Questions and intervention attributes			
Reliability				
31.	Answers to questions and attributes are able to explain how an intervention contributes to intended effects for final beneficiaries			
32.	The Evaluation Questions allow for success and failure (positive and negative effects) to be distinguished			
Proper application of designs and methods (Robustness)				
33.	The ways that designs and methods are applied are clearly described and documented			
34.	The application of designs and methods and subsequent analysis follow any protocols or good practice guidelines			
35.	The evaluation team is knowledgeable about the methods used			
Drawing legitimate conclusions (Transparency)				
36.	Conclusions clearly follow from the findings			
37.	The evaluation explained the effects of the programme			
38.	Evaluative judgements are justified			
39.	Stakeholder judgements have been taken into account when reaching conclusions			
40.	The limitations of the evaluation and its conclusions are described			

4. Developing a Technical Concept

Criteria for selecting programmes for impact evaluation

Valuable contributions have been provided by international development agencies, including the OECD and the World Bank to define criteria that can be applied to development organizations' types of interventions. Criteria will help facilitate selecting programs or projects for impact evaluation and to substantiate investment in such specialized and demanding evaluation practice⁸. The following selection criteria derive from a tailoring effort to synthesize the industry's general considerations that can be used to substantiate selection of interventions by development organizations for impact evaluation:

Exploring new approaches

- Innovative schemes testing a new and promising approach

Testing existing approaches

- Interventions for which there is scant solid evidence of impact in the given context
- Pilot or replicable programs. The program can be substantially scaled up or can be applied in a different setting.

Scope and strategic relevance

- Strategically relevant. The program is a flagship initiative; requires substantial resources; covers, or could be expanded to cover, a large number of people; or could generate substantial savings.
- Influential. The results will be used to inform key policy decisions.

These criteria are consistent with generally established considerations used when submitting development interventions to the demanding scrutiny of impact evaluations.

The main components of the technical concept paper/proposal

The most important contribution to be made in terms of preliminary quality control mechanisms is likely to be through the development of an impact evaluation technical concept document. This will be done by the organization's technical teams and supporting specialists leading the impact evaluation initiative. The technical concept should include

- The nature of the evaluation (key questions);
- A draft theory of change (or a change model);
- A description of the methods to be used and how they link to the key evaluation questions;
- Whether (and if so how) to treat counterfactuals and/or how to use such methods as causal contribution analysis;

⁸ See

- World Bank, 'Impact evaluation in practice' (2011)
- OECD, 'Outline of Principles of Impact Evaluation'

- Which partners should be associated with the impact evaluation and in what roles;
- How impact evaluation results will be used;
- The Quality Assurance/Control plan for the impact evaluation.

A Technical Concept document, outlining these components forms the basis for the Terms of Reference defining the role of each impact evaluation team member; establishing criteria for selecting impact evaluation proposals and to follow up implementation through quality control measures; and providing main stakeholders with a clear understanding of the processes involved.⁹

⁹ United Nations Evaluation Group, 'Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management' (2013), Section 4 page 8

5. Accessing Key Resources from the Industry's Literature

The DAC/OECD definition of impact continues to drive the industry's conceptual reference for defining impact evaluation, including operationalization processes and impact evaluation practices. The industry's standards and guidelines, including step by step guidance have been progressively experiencing a process toward basic consensus through a global dialogue led by NONIE.

As a result of this successful dialogue, Step by Step Guidance to perform impact evaluation are being extensively used by the development community and gradually improved to facilitate better impact evaluation practices. Access to key resources, including recent innovations on methods and techniques is offered in this section. The industry's entity responsible for each resource is presented as well as its title and a basic classification based on the type of resources.

Key Resources for Performing Impact Evaluation			
Industry's Entity	Title	Type	Link
NONIE	Impact Evaluations and Development: NONIE Guidance on Impact Evaluation	Industry's Main Guidance	 Guidance on Impact Evaluation.pdf
World Bank	Impact Evaluation in Practice	Guidance and Case Studies	 Impact Evaluation in Practice.pdf
United Nations	Impact Evaluation: Guidance on Selection, Planning and Management	Guidance & Standards	 Impact Evaluation Guidance.pdf
DFID	Broadening the Range of Designs and Methods for Impact Evaluations	Design, Methods & Standards	 IE Design and Methods.pdf
Inter-Action/ Rockefeller Foundation	Mixed Methods in Impact Evaluation	Mixed Methods Approach	 IE mixed-methods.pdf