

Overview of Impact Evaluation

Patricia Rogers

UNICEF OFFICE OF RESEARCH

The Office of Research is UNICEF's dedicated research arm. Its prime objectives are to improve international understanding of issues relating to children's rights and to help facilitate full implementation of the Convention on the Rights of the Child across the world. The Office of Research aims to set out a comprehensive framework for research and knowledge within the organization, in support of UNICEF's global programmes and policies, and works with partners to make policies for children evidence-based. Publications produced by the Office are contributions to a global debate on children and child rights issues and include a wide range of opinions.

The views expressed are those of the authors and/or editors and are published in order to stimulate further dialogue on impact evaluation methods. They do not necessarily reflect the policies or views of UNICEF.

OFFICE OF RESEARCH METHODOLOGICAL BRIEFS

UNICEF Office of Research Methodological Briefs are intended to share contemporary research practice, methods, designs, and recommendations from renowned researchers and evaluators. The primary audience is UNICEF staff who conduct, commission or interpret research and evaluation findings to make decisions about programming, policy and advocacy.

This brief has undergone an internal peer review.

The text has not been edited to official publication standards and UNICEF accepts no responsibility for errors.

Extracts from this publication may be freely reproduced with due acknowledgement. Requests to utilize larger portions or the full publication should be addressed to the Communication Unit at florence@unicef.org

To consult and download the Methodological Briefs, please visit <http://www.unicef-irc.org/KM/IE/>

For readers wishing to cite this document we suggest the following form:

Rogers, P. (2014). Overview of Impact Evaluation, *Methodological Briefs: Impact Evaluation 1*, UNICEF Office of Research, Florence.

Acknowledgements: This brief benefited from the guidance of many individuals. The author and the Office of Research wish to thank everyone who contributed and in particular the following:

Contributors: Simon Hearn, Jessica Sinclair Taylor, Howard White

Reviewers: Nikola Balvin, Samuel Bickel, Christian Salazar, David Stewart

© 2014 United Nations Children's Fund (UNICEF)
September 2014

UNICEF Office of Research - Innocenti
Piazza SS. Annunziata, 12
50122 Florence, Italy
Tel: (+39) 055 20 330
Fax: (+39) 055 2033 220
florence@unicef.org
www.unicef-irc.org

1. WHAT IS IMPACT EVALUATION - AND WHY DOES IT MATTER?

Impact evaluations provide information about the impacts produced by an intervention. Impact evaluation can be undertaken of a programme or a policy, or upstream work – such as capacity building, policy advocacy and support for an enabling environment. This goes beyond looking only at goals and objectives to also examine unintended impacts. OECD-DAC defines impacts as “positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.”¹

An impact evaluation can be undertaken for formative purposes (to improve or reorient a programme or policy) or for summative purposes (to inform decisions about whether to continue, discontinue, replicate or scale up a programme or policy). It can be used by UNICEF and its partners at the global, regional or country level to inform decisions and for advocacy and advice.

This brief provides an overview of the different elements of impact evaluation and the different options for UNICEF programme managers for each of these elements, in terms of stages involved in planning and managing an impact evaluation. These stages are as follows, although these can sometimes vary in order or may be revisited:

- deciding to conduct an impact evaluation
- establishing governance and management arrangements
- preparing for the impact evaluation
- developing a [Terms of Reference](#) (ToR) for the evaluation
- engaging the evaluation team
- overseeing the evaluation, including the production of evaluation reports
- following up the evaluation.

Guidance is also provided on addressing ethical issues and ensuring quality during an impact evaluation.

This brief is an introduction to a UNICEF series of methodological briefs, which provides more detailed guidance on the core building blocks of impact evaluation (theory of change, evaluative criteria and evaluative reasoning), different evaluation designs and methods for data collection and analysis, and options in terms of participatory approaches.

2. DECIDING TO CONDUCT AN IMPACT EVALUATION

It is important that impact evaluation is addressed as part of an integrated monitoring, evaluation and research plan (IMERP) that generates and makes available evidence to inform decisions. This will ensure that data from other [monitoring](#) and evaluation components such as [performance monitoring](#) and [process evaluation](#) can be used as needed. It will also ensure that planning for an impact evaluation begins early, allowing for the collection of baseline data and, where appropriate, the creation of a [control group](#) or [comparison group](#) or the use of other strategies to investigate [causal attribution](#).

¹ OECD-DAC, *Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Paris, 2010. See <http://www.oecd.org/development/peer-reviews/2754804.pdf>.

An impact evaluation should only be undertaken when its intended use can be clearly identified and when it is likely to be able to produce useful findings, taking into account the availability of resources and the timing of decisions about the programme or policy under investigation. A formal [evaluability assessment](#) (EA)² might first need to be conducted to assess these aspects.

Formative impact evaluations are undertaken to inform decisions in regard to making changes to a programme or policy. While many formative evaluations focus on processes, impact evaluations can be used formatively if an intervention is ongoing. For example, the findings of an impact evaluation can be used to improve implementation of a programme for the next intake of participants.

Summative impact evaluations are undertaken to inform decisions about whether to continue, discontinue, replicate or scale up an intervention. Ideally, a summative impact evaluation not only produces findings about ‘what works’ but also provides information about what is needed to make the intervention work for different groups in different settings, which can then be used to inform decisions.

Table 1. When an impact evaluation might be appropriate

Issue	Impact evaluation might be appropriate when...	Impact evaluation might not be appropriate when...
Intended uses and timing	There is scope to use the findings to inform decisions about future programmes or policies.	There are no clear intended uses or intended users – for example, decisions have already been made on the basis of existing credible evidence, or need to be made before it will be possible to undertake a credible impact evaluation.
Focus	There is a need to understand the impacts that have been produced.	The priority at this stage is to understand and improve the quality of implementation.
Resources	There are adequate resources to undertake a sufficiently comprehensive and rigorous impact evaluation, including the availability of existing, good quality data and additional time and money to collect more.	Existing data are inadequate and there are insufficient resources to fill gaps.
Relevance	It is clearly linked to national and UNICEF strategies and priorities.	It is peripheral to national and UNICEF strategies and priorities.

The process of prioritizing interventions to undergo impact evaluation should involve a range of stakeholders, including government representatives, civil society, UN system managers and officers, and representatives of other partner organizations. The process should consider the relevance of the evaluation in terms of national and UN strategies, its potential usefulness and commitment to its use by senior managers, its potential use for advocacy for evidence-based policy, and accountability requirements.³

² For more information, see: BetterEvaluation, ‘Evaluability Assessment’, web page, BetterEvaluation, http://betterevaluation.org/themes/evaluability_assessment.

³ For further advice, see: United Nations Children’s Fund, *Revised Evaluation Policy*, E7ICEF72013/14, Executive Board Annual Session 18–21 June 2013, UNICEF, 2013.

It is also important to consider when it is appropriate to conduct an impact evaluation. Impact evaluations that are conducted belatedly will provide information too late to inform decisions. For this reason, some reports that are labelled 'impact evaluations' actually only report on intermediate outcomes that are evident during the life of the evaluation rather than on the long-term impacts of the intervention. For example, the evaluation of Mexico's Progresa/Oportunidades conditional cash transfer (CCT) programme⁴ looked at school attendance rates rather than learning outcomes.

But impact evaluations that are done too early will provide an inaccurate picture of the impacts. In some cases, impacts will be understated, as these will not have had sufficient time to develop, for example, children completing school after participating in an early intervention programme. In some cases, impacts will be overstated, for example, participants' knowledge and skills might decline over time if they don't have opportunities to practise and maintain them. In such cases, a later follow-up study will be needed to examine the durability of impacts.

3. ESTABLISHING GOVERNANCE AND MANAGEMENT ARRANGEMENTS

An impact evaluation should be formally planned and managed as a discrete project, with management arrangements clearly described from the beginning of the process.

Important structures and processes include:

- **Evaluation management team (EMT) or steering committee** – to select and supervise the work of external evaluators contracted to conduct the evaluation, including by reviewing the ToR, the inception report and a draft of the final report. The team provides technical guidance to the evaluators and also manages operational aspects of the evaluation such as the budget, field visits and contact with relevant UNICEF Country Offices and sections, and the quality assurance of the evaluation tools (e.g., the appropriateness of a particular survey). In a joint evaluation, this would include UNICEF staff and representatives of each partner organization.
- **Evaluation reference group (ERG)** – to provide technical guidance and advice on culturally appropriate interaction with the local communities (including culturally appropriate tools and evaluation methods) and on the development of the ToR, selection criteria for the evaluation team, review of the evaluation plan and draft evaluation report. The ERG may comprise both internal and external experts (including key stakeholders).

Acceptance of the final evaluation report is usually through a process of review by the EMT and ERG, with the UNICEF head of office responsible for the final sign-off/approval.

The primary intended users of the impact evaluation (whether organizations or individuals) should be identified and engaged as much as possible in the evaluation process, either on the EMT or as part of a stakeholder reference group. Primary intended users could include local, regional and global UNICEF managers and staff, and representatives of local partner organizations; local communities; local non-governmental organizations (NGOs); local funders; partner government departments; global offices of partner organizations and NGOs; global offices of funders; and relevant research institutes.

Appropriate processes must be developed to understand the evaluation requirements of the primary intended users and to decide how to prioritize these. In addition, there may be other intended users (e.g., agencies implementing similar programmes) who would be keen to learn from the impact evaluation but whose information needs are not a priority for the evaluation.

⁴ See Secretaría de Desarrollo Social (SEDESOL), Oportunidades, www.oportunidades.gob.mx.

4. PREPARING TO CONDUCT THE IMPACT EVALUATION

Reviewing the theory of change

A theory of change explains how activities are understood to produce a series of results that contribute to achieving the intended or observed impacts. Some version of a theory of change is likely to have been developed as part of the planning for a programme or policy. When preparing for an impact evaluation, the theory of change should be reviewed and revised as necessary. For example, the existing theory of change may have gaps or unrealistic assumptions that should be revised; it can also be out of date if the programme has evolved since the theory of change was developed.

Not all versions of a theory of change actually explain *how* an intervention is understood to work (e.g., results chains and logframes). For example, a programme may have a stated theory of change that it will undertake hygiene education activities that explain the importance of handwashing for health, to encourage better handwashing and improve health. A good theory of change will explain the actual causal mechanism that is understood to work, that is, does hygiene education work by providing new information (e.g., that soap is needed to improve the effectiveness of handwashing, or that ash can be used as a substitute for soap), by reminding people of information they already know, by reinforcing or changing social norms, or by creating opportunities for people to discuss and reduce barriers to accessing water and soap? Clarifying this causal mechanism, or identifying it as an important information gap, can be useful in preparing for an impact evaluation. The impact evaluation can be improved by using the theory of change to:

- identify relevant variables that should be included in data collection
- identify intermediate outcomes that can be used as markers of success, in situations where the impacts of interest will occur after the evaluation time frame
- identify aspects of implementation that should be examined to see if the failure to achieve intended impacts is due to a failure to implement the intervention successfully
- identify potentially relevant contextual factors that should be included in data collection
- guide data analysis
- provide a framework for reporting findings.

More information about developing and using a theory of change is available in Brief No. 2, Theory of Change.

Identifying and mobilizing resources

It is important to create an 'evaluation envelope' (i.e., money and resources set aside for the evaluation) early on so that sufficient resources will be available when the evaluation is conducted. If decisions about the broad evaluation design and costs can be made during the planning for the intervention, then sufficient resources can be set aside for the evaluation. An estimate of the scale of resources required can be developed by referring to the budgets of previous similar evaluations, by using templates or by developing an initial design and doing a costing for it.⁵

As well as funding that may be provided to an external evaluation team, another important resource that must be ensured is the available time of key contributors, including the governance group, review bodies (such as a technical advisory group or peer reviewers), [key informants](#) and facilitators of site visits.

⁵ For a checklist of items to consider, see Horn, Jerry, 'A Checklist for Developing and Evaluating Evaluation Budgets', Western Michigan University Checklist Project, Kalamazoo, 2001. See http://www.wmich.edu/evalctr/archive_checklists/evaluationbudgets.pdf.

Access to data systems and previous documentation is also needed. There should be a good understanding of the data that are already available, so that the evaluation design can focus on verifying information and filling gaps rather than gathering data that are redundant.

Deciding the process for developing the evaluation methodology

A good evaluation methodology requires sufficient understanding of the specific evaluation needs and resources as well as general technical expertise. Therefore it is best to develop the evaluation design through a process that brings together these elements. The diagram below shows four possible process options for developing the evaluation methodology.

Figure 1. Process options for developing the evaluation methodology

Methodology by commissioning agency, up front	Initial methodology by commissioning agency, then revised	Initial methodology by evaluators, then revised	Methodology by evaluators, up front as separate project
<ul style="list-style-type: none"> Set out in ToR 	<ul style="list-style-type: none"> Initial methodology revised in inception report 	<ul style="list-style-type: none"> Initial methodology revised in inception report 	<ul style="list-style-type: none"> Set out in ToR The actual evaluation is commissioned as a separate project

Methodology developed by commissioning agency: This is where the ToR sets out the design, either by naming this explicitly or in the way that the budget is constructed (e.g., a certain number of days each for in-office document review and for in-field interviews). It can be an appropriate option when the commissioning agency has sufficient technical expertise, either because of the skills of its staff or because the evaluation uses a pre-existing design. In other circumstances, there is a risk that this option will fail to make the best use of opportunities because of the limited repertoire of evaluation designs and methods familiar to the commissioners.

Initial methodology developed by commissioning agency, then revised: This is where the ToR sets out an indicative design that is then reviewed by the evaluators as a first step following their appointment. They set out in an inception report the final design – including any revisions – based on discussions with the commissioning agency and a review of the available data. This option provides an opportunity for additional technical and logistical input into the methodology and is therefore likely to deliver a stronger design than the indicative design originally proposed.

Initial methodology developed by evaluators, then revised: This is where the ToR asks potential evaluators to propose and cost an evaluation design. This is then reviewed and revised as a first step following their appointment. They set out in an inception report the final design – including any revisions – based on discussions with the commissioning agency and a review of the available data. As with the previous option, this scenario is likely to produce a stronger design as it is based on a better understanding of the realities on the ground.

Methodology developed as a separate project: This is where a separate project is undertaken to develop the evaluation methodology. This can involve considerable consultation with the evaluation's

intended users and a review of the available data in the form of an evaluability assessment.⁶ The conduct of the actual evaluation is then advertised as a separate contract; the team that has developed the methodology may not be chosen to undertake the evaluation if they lack the necessary expertise, coverage or availability. This option is most appropriate in situations where a complicated methodology is needed to address the particular features of the evaluation.

5. DEVELOPING A TERMS OF REFERENCE (TOR) FOR THE EVALUATION

Developing an appropriate ToR document is a critically important step in an impact evaluation. While the specific format of a ToR can vary, especially if it is for a joint evaluation, it must set out the essential elements of the impact evaluation, including its purpose, objectives and intended uses, the high-level key evaluation questions (KEQs) it aims to answer, the specific deliverables and timelines involved, and the essential competencies of the evaluation team. The ToR often includes the proposed methodology for the impact evaluation – options for developing and refining this have been discussed above.

Background

The ToR should provide a summary of the development of the intervention being evaluated and what has led to the decision to undertake an impact evaluation. Key concepts should be defined, for example, the ToR for an impact evaluation of UNICEF's emergency preparedness systems included an overview of 'preparedness.'⁷

Evaluation purpose, objectives and scope (including KEQs)

The evaluation purpose refers to the rationale for conducting an impact evaluation. Evaluations that are being undertaken to support learning should be clear about who is intended to learn from it, how they will be engaged in the evaluation process to ensure it is seen as relevant and credible, and whether there are specific decision points around where this learning is expected to be applied. Evaluations that are being undertaken to support accountability should be clear about who is being held accountable, to whom and for what.

The objectives of the evaluation are usually expressed in terms of the OECD-DAC [evaluation criteria](#)⁸ (see table 2).

⁶ For more information, see: BetterEvaluation, 'Evaluability Assessment', web page, BetterEvaluation, http://betterevaluation.org/themes/evaluability_assessment.

⁷ UNICEF Evaluation Office, *Evaluation of UNICEF's Emergency Preparedness Systems, Terms of Reference – Final*, UNICEF, 2012. See http://www.unicef.org/evaluation/files/Evaluation_of_UNICEF_Emergency_Preparedness_Systems_-_FINAL_ToR.pdf.

⁸ Organisation for Economic Co-operation and Development – Development Assistance Committee, 'Evaluation of development programmes, DAC Criteria for Evaluating Development Assistance', web page, OECD. See <http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>.

Table 2. OECD-DAC evaluation criteria

Relevance	The extent to which the objectives of the intervention are consistent with recipients' requirements, country needs, global priorities and partners' policies.
Effectiveness	The extent to which the intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance.
Efficiency	A measure of how economically resources/inputs (funds, expertise, time, equipment, etc.) are converted into results.
Impact	Positive and negative primary and secondary long-term effects produced by the intervention, whether directly or indirectly, intended or unintended.
Sustainability	The continuation of benefits from the intervention after major development assistance has ceased. Sustainability looks to the probability of continued long-term benefits.

Impact evaluations can also be intended to address equity, gender equality and human rights.

An evaluation may focus on some of these criteria rather than all of them. In an impact evaluation, more specific descriptions need to be developed for each criterion in terms of 'what success looks like'.

More information about defining the evaluative criteria is available in Brief No. 3, Evaluative Criteria.

Impact evaluations should be focused around answering a small number of high-level KEQs. These are not interview questions, rather questions that will be answered through a combination of evidence (see table 3 for some examples).

These questions should be clearly linked to the intended uses of the impact evaluation. For example, if an evaluation is intended to inform the scaling up of a pilot programme, then it is not enough to ask 'Did it work?' or 'What were the impacts?'. A good understanding is needed of how these impacts were achieved in terms of activities and supportive contextual factors to replicate the achievements of a successful pilot. Equity concerns also require that impact evaluations for all UNICEF programmes go beyond simple average impact to identify for whom and in what ways the programmes have been successful.

Table 3. Examples of key evaluation questions for impact evaluation linked to the OECD-DAC evaluation criteria

Relevance
To what extent did the intended impacts match the needs of the intended participants?
To what extent did the intended impacts match the stated priorities of partner governments and the UN system?

Effectiveness

Did the intervention produce [the intended impacts] in the short, medium and long term?

For whom, in what ways and in what circumstances did the intervention work?

Did [the intended impacts] reach all intended participants?

How did the intervention contribute to [the intended impacts]?

What were the particular features of [the intervention] that made a difference?

What variations were there in implementation? What has been the quality of implementation in different sites? To what extent and in what ways did implementation change over time as the intervention evolved?

To what extent are differences in impact explained by variations in implementation?

How did the intervention work in conjunction with other interventions to achieve outcomes?

What helped or hindered the intervention to achieve these impacts?

Efficiency

What resources have been used to produce these results? How does this compare to similar interventions?

What strategies have been used to ensure the efficiency of the intervention?

Impact

What unintended impacts (positive and negative) did the intervention produce?

Sustainability

Are impacts likely to be sustainable? (To be studied at the conclusion of the intervention.)

Have impacts been sustained? (Follow-up study required.)

The ToR must make the scope of the evaluation clear. For example, the ToR for the evaluation of the UNFPA-UNICEF Joint Programme on Female Genital Mutilation/Cutting: Accelerating Change⁹ explicitly referred to the different levels of the programme and the extent to which the evaluation was expected to address each of these.

Evaluation design and methods

The evaluation methodology sets out how the KEQs will be answered. It specifies designs for causal attribution, including whether and how comparison groups will be constructed, and methods for data collection and analysis. It is helpful to create a matrix showing which data collection/collation and analysis methods (including using existing data) will be used to answer each KEQ.

Within the KEQs, it is useful to identify the different types of questions involved – descriptive, causal and evaluative. Impact evaluations must have credible answers to all of these: how have things

⁹ United Nations Population Fund and United Nations Children's Fund, *Joint Evaluation, UNFPA-UNICEF Joint Programme on Female Genital Mutilation/Cutting: Accelerating Change, 2008–2012*, Volume I, UNFPA/UNICEF, New York, September 2013. See http://www.unfpa.org/webdav/site/global/shared/documents/Evaluation_branch/Joint_Evaluation_-_Sept_2013/Main_Report/FGM-report_12_4_2013.pdf.

changed, to what extent is this change due to the intervention being evaluated and, overall, has the intervention been a success? Most impact evaluations address both descriptive and causal questions, but many fall short of addressing evaluative questions (i.e., questions about quality and value) even though doing so is precisely what makes evaluation so much more useful and relevant than the mere measurement of [indicators](#) or summaries of observations and stories (see Brief No. 4, Evaluative Reasoning).

Different evaluation designs and data collection and analysis methods are needed to answer these different types of evaluation questions, as outlined below.

[Descriptive questions](#)

Descriptive questions ask about how things are and what has happened, including describing the initial situation and how it has changed, the activities of the intervention and other related programmes or policies, the context in terms of participant characteristics, and the implementation environment.

There is a range of data collection/collation methods for (but not necessarily exclusive to) answering descriptive questions. This includes:

- Collecting data from individuals or groups: through interviews, questionnaires or specialized methods such as [dotmocracy](#), [hierarchical card sorting](#), seasonal calendars, projective techniques and stories.
- Observation: structured or non-structured; participant or non-participant; recorded through notes, photos or video; participatory or non-participatory.
- Physical measurement: biophysical measurements or geographical information.
- Existing documents and data: including existing data sets, official statistics, project records and social media data.

In most cases, an effective combination of quantitative and qualitative data will provide a more comprehensive picture of what has happened. Many impact evaluations fail to gather adequate [baseline data](#) about conditions or adequate data about the implementation of the intervention.

More information about different options for data collection and analysis, and about combining qualitative and quantitative data, is available in Brief No. 10, Overview: Data Collection and Analysis Methods in Impact Evaluation. Specific advice on using interviews is available in Brief No. 12, Interviewing.

[Causal questions](#)

Causal questions ask whether or not, and to what extent, observed changes are due to the intervention being evaluated rather than to other factors, including other programmes and/or policies.

There are three design options that address causal attribution:

- Experimental designs – which construct a control group through random assignment.
- Quasi-experimental designs – which construct a comparison group through matching, regression discontinuity, propensity scores or another means.
- Non-experimental designs – which look systematically at whether the evidence is consistent with what would be expected if the intervention was producing the impacts, and also whether other factors could provide an alternative explanation.

All of these options require significant investment in preparation and early data collection, and cannot be done if an impact evaluation is limited to a short exercise conducted towards the end of intervention implementation – an area of weakness common to many impact evaluations.

More information about different design options is available in Brief No. 6, Overview: Strategies for Causal Attribution; Brief No. 7, Randomized Controlled Trials; Brief No. 8, Quasi-experimental Designs and Methods; and Brief No. 9, Comparative Case Studies.

Evaluative questions

Evaluative questions ask about the overall conclusion as to whether a programme or policy can be considered a success, an improvement or the best option.

Since development interventions often have multiple impacts, which are distributed unevenly (equity is a particular concern in the UN system), this is an essential element of an impact evaluation. For example, should an economic development programme be considered a success if it produces increases in household income but also produces hazardous environment impacts? Should it be considered a success if the average household income increases but the income of the poorest households is reduced?

In any impact evaluation, it is important to define first what is meant by ‘success’ (quality, value). One way of doing so is to use a specific [rubric](#) that defines different levels of performance (or standards) for each evaluative criterion, deciding what evidence will be gathered and how it will be synthesized to reach defensible conclusions about the worth of the intervention. At the very least, it should be clear what trade-offs would be appropriate in balancing multiple impacts or distributional effects. This is an area of weakness: many impact evaluations fail to address this issue systematically or at all.

More information about the importance of addressing evaluative questions and about specific options for how to go about it (in terms of a recommended process as well as ways for achieving defensible conclusions) is available in Brief No. 4, Evaluative Reasoning.

Evaluation team profile

The ToR should identify the selection criteria for the evaluation team, including mandatory and desired competencies. These may include: experience working in the relevant country or region; language skills; content knowledge (in a sector or issue such as gender); technical skills in evaluation design, data collection, data analysis and reporting; communication skills; and project management skills. It may be necessary to ensure that the team as a whole has adequate coverage, for example, current knowledge of local networks and ability to access key informants, and knowledge of evaluations of comparable interventions in other regions. It is important to ensure that the team has the technical skills to collect and analyse data, the communication skills to effectively negotiate the design and to report findings, and the understanding that evaluations must focus on producing useful and accessible findings, not just academic reports.

If participatory approaches – involving members of stakeholder groups, especially community members – to conducting evaluations are seen as appropriate, then the evaluation team must have the appropriate skills and experience to facilitate this.

More information about strategies to engage different stakeholders (including community members) in evaluations is available in Brief No. 5, Participatory Approaches.

More information about processes to develop a ToR for an evaluation is available in the UNICEF Evaluation Technical Note No. 2, What goes into a Terms of Reference (TOR).¹⁰

6. ENGAGING THE EVALUATION TEAM

The procurement of an external evaluation team should follow standard UNICEF procedures designed to ensure a wide pool of appropriate potential evaluators from which to select the most appropriate option.

The ToR should provide sufficient information to enable potential evaluators to provide an informed technical and financial bid. For this reason, scoping studies or evaluability assessments are best completed prior to deciding to undertake an impact evaluation and advertising for external evaluators. To encourage a wide pool of potential evaluators, the ToR should be advertised widely, including on evaluation discussion lists as well as by direct mail to potential evaluators. Signalling opportunities for academic publication can encourage university evaluators and researchers to submit proposals. Effective strategies to select the best evaluation team begin with having appropriate selection criteria in the ToR. Reviewing examples of previous evaluation reports, where possible, and discussing these with the prospective teams should give a good insight into how well matched their skills and values are to the requirements of the evaluation.

It is important to allow sufficient time between advertising the ToR and the deadline for the submission of proposals, and between the acceptance of a proposal and the start of the work. There should also be reasonable time frames for the completion of the work. Time frames that are too tight can severely limit the number of evaluators available to work on the project and may compromise the quality of the work.

7. OVERSEEING THE EVALUATION

Evaluation work plan

In many cases, an indicative work plan will have been developed as part of the ToR. This sets out the various stages of data collection and analysis, and the deadlines for various deliverables. Where possible, the work plan should build in opportunities for interim reporting and ongoing analysis rather than reserve all reporting for the end of the evaluation. The work plan should be reviewed and revised as part of the [inception report](#).

Once accepted, the management should monitor the work plan and governance structures be put in place – and revised, if necessary – to address emerging issues or unanticipated delays.

Time should be scheduled in for the review and revision of key deliverables, including the revised methodology and work plan, data collection and analysis tools (if these must be approved) and draft reports.

¹⁰ UNICEF Evaluation Office, 'What goes into a Terms of Reference (TOR)', *Evaluation Technical Note No. 2*, UNICEF, 2002. A checklist of questions for UNICEF evaluation managers to work through when developing Terms of Reference. See http://www.unicef.org/evaluation/files/TechNote2_TOR.pdf.

Reporting findings

Reports should be developed in accordance with the UNICEF-developed GEROS¹¹ (Global Evaluation Reports Oversight System) and evaluators provided with a template to facilitate the production of a compliant document.

Reports should address the UNICEF-Adapted UNEG Evaluation Reports Standards.¹² In summary, these aim to ensure:

1. The report is well structured, logical, clear and complete.
2. The report presents a clear and full description of the 'object' of the evaluation.
3. The evaluation's purpose, objectives and scope are fully explained.
4. The report presents a transparent description of the methodology applied to the evaluation, which clearly explains how the evaluation was specifically designed to address the evaluation criteria, yield answers to the evaluation questions and achieve the evaluation purposes.
5. Findings respond directly to the evaluation criteria and questions detailed in the scope and objectives section of the report, and are based on evidence derived from data collection and analysis methods described in the methodology section of the report.
6. Conclusions present reasonable judgements based on the findings and substantiated by evidence, and provide insights pertinent to the object and purpose of the evaluation.
7. Recommendations are relevant to the object and purpose of the evaluation, are supported by evidence and conclusions, and were developed with the involvement of relevant stakeholders.
8. The report illustrates the extent to which the design and implementation of the object, the assessment of results and the evaluation process incorporate a gender equality perspective and human rights-based approach (HRBA), including child rights.

The evaluation report should be structured around the key messages in the executive summary and provide easy access to more detailed explanation and evidence. Using summary statements as headings and chart and table titles makes it easier for readers to focus on these key messages.

Management response

An evaluation requires a formal management response¹³ in the form of a written response to the recommendations made in the evaluation report. Some agencies include this in the final report; in UNICEF, the management response is made available as a separate document. The response might agree, partially agree or disagree with a recommendation and should provide an explanation for any partial acceptance or rejection of a recommendation. For recommendations that have been accepted, or partially accepted, key follow-up actions should be identified, with a time frame specified and the responsible unit named.

¹¹ For further information about reporting requirements, see: United Nations Children's Fund, Global Evaluation Reports Oversight System (GEROS), UNICEF, 2013. See http://www.unicef.org/evaluation/files/GEROS_Methodology_v7.pdf.

¹² UNICEF Evaluation Office, UNICEF-Adapted UNEG Evaluation Reports Standards, UNICEF, 2010. See http://www.unicef.org/evaldatabase/files/UNEG_UNICEF_Eval_Report_Standards.pdf.

¹³ United Nations Children's Fund, *Guidance for Management Response to Evaluation: Enhancing Critical Engagement for the Strategic Use of Evaluations*, UNICEF, 2012. See http://www.unicef.org/evaldatabase/files/2011_South_Africa_-_management_response_SCCFS_evaluation.pdf, and <https://intranet.unicef.org/epp/evalsite.nsf/8e1ddc662803020785256ede00706595/6a1a69ce471c376585257c40006677c1?OpenDocument>.

It is important to identify an individual to coordinate the management response and an agreed deadline by which comments must be provided. The management should respond within two months of receiving the final evaluation report.¹⁴

8. FOLLOWING UP THE EVALUATION

Disseminating findings

In addition to the formal evaluation report, other knowledge products and processes can be used to make findings more readily available. These include other written materials such as an evaluation brief, an e-newsletter or an entry on a wiki, all of which would contain a summary of the evaluation.

The findings can also be shared through activities such as presentations at meetings, at workshops and in existing communities of practice, and for webinars. Such presentations can even be recorded and shared more widely, for example, through YouTube or Vimeo.

All UNICEF evaluations should also be uploaded to the UNICEF evaluation database as part of the Country Office Annual Reporting (COAR) system.

Tracking follow-up

Tracking progress of the implementation of recommendations should be done as part of annual planning and review processes. Each previously identified action should be reported on in terms of its progress – whether ‘initiated’, ‘not initiated’, ‘completed’ or ‘no longer applicable’.

Documenting lessons learned about evaluation

Ideally, the lessons learned about evaluation should also be documented and shared, at least internally. Specifically, the Revised Evaluation Policy of UNICEF¹⁵ states:

“Division directors are responsible for planning, resourcing and commissioning evaluations of the global policies and initiatives for which they are accountable, and for responding to relevant evaluation lessons and recommendations. Directors implement and support actions to strengthen evaluation capacity and engagement in the headquarters units they oversee and the global networks they manage.”

This requirement ties in directly with the formal management response referred to above, including specific actions to be taken based on the evaluation findings.

¹⁴ United Nations Evaluations Group, *UNEG Guidance on Preparing Management Responses to UNDAF Evaluations*, Guidance Document, UNEG, 2014. See http://www.unevaluation.org/UNDAF_management_response_guidance.

¹⁵ United Nations Children’s Fund, *Revised Evaluation Policy of UNICEF*, UNICEF, New York, 2013, paragraph 37. See http://www.unicef.org/about/employ/files/2013-14--Final_approved_Revised_Evaluation_Policy.pdf.

9. ANTICIPATING AND ADDRESSING ETHICAL ISSUES

Evaluations should follow the *Ethical Guidelines for Evaluation*¹⁶ set out by the United Nations Evaluation Group (UNEG) and to which UNICEF adheres. These guidelines establish a code of conduct for evaluators as well as ethical obligations to participants in the evaluation.

These obligations include:

- respect for dignity and diversity
- rights
- confidentiality
- avoidance of harm.

These obligations must be explicitly addressed in the design of the evaluation and throughout the evaluation process. Their intention is to ensure respect for and protection of the rights and welfare of human subjects and the communities to which they belong, in accordance with the Universal Declaration of Human Rights and other human rights conventions.

UNICEF requires that all evaluations it supports must adhere to the laws of the country in which the relevant UNICEF office is located. This includes an ethical review by “an objective and qualified body”,¹⁷ where this is deemed necessary.

Evaluations that involve children as informants, as data collectors or in other ways require particular attention to be paid to ethical issues. The UNICEF Office of Research - Innocenti released in October 2013 *Ethical Research Involving Children*, an online resource¹⁸ that brings together expert thinking about key ethical issues involving children and how these might be addressed in different research (and evaluation) contexts.

More information about ethical issues is available in Brief No. 5, Participatory Approaches and Brief No. 10, Overview: Data Collection and Analysis Methods in Impact Evaluation.

10. ENSURING A QUALITY IMPACT EVALUATION

Quality impact evaluation involves striking an appropriate balance between competing criteria: validity¹⁹, utility (especially timeliness and relevance), feasibility and ethics. These can be in tension, for example, a manager might need to choose between an evaluation design that can provide very accurate and comprehensive findings – and one that will produce adequate information in time to inform a particular decision. Some evaluation designs focus on improving internal validity (the ability to correctly draw a causal attribution) but at the cost of external validity (the ability to apply the findings to other settings).

¹⁶ United Nations Evaluation Group, *Ethical Guidelines for Evaluation*, UNEG, New York, 2007. See https://www.iom.int/jahia/webdav/site/myjahiasite/shared/shared/mainsite/about_iom/eva_techref/UNEG_ethical_guidelines.pdf.

¹⁷ United Nations Children's Fund, *Revised Evaluation Policy of UNICEF*, UNICEF, New York, 2013, paragraph 54(c). See http://www.unicef.org/about/employ/files/2013-14--Final_approved_Revised_Evaluation_Policy.pdf.

¹⁸ The new resources include: the International Charter for Ethical Research Involving Children; a website, www.childethics.com, specifically designed to provide a rich repository of evidence-based information, resources and links to journal articles to guide and improve research involving children and to provide a platform for further critical reflection and dialogue; and a compendium of ethical issues and challenges, including a collection of more than 20 case studies as well as structured questions to guide ethical research involving children (the web page called 'Getting Started').

¹⁹ Different types of validity are important: construct validity refers to the ability of measures and other tools to accurately represent reality; internal validity refers to the ability to correctly draw a causal attribution, i.e., to identify if observed changes are due to the intervention; and external validity refers to the ability to generalize the results to other populations or sites.

Strategies for ensuring a quality impact evaluation include:

- being aware of the trade-offs being made and explicitly discussing the available options
- engaging a range of people in the governance of the evaluation to ensure that its focus and the decisions made around trade-offs are appropriate
- involving a technical advisory group or peer reviewers to review the evaluation design and draft reports.

11. OTHER BRIEFS IN THIS SERIES

This is the first in a series of methodological briefs on impact evaluation produced by UNICEF. The briefs in this series are as follows:

Building Blocks

1. Overview of Impact Evaluation
2. Theory of Change
3. Evaluative Criteria
4. Evaluative Reasoning
5. Participatory Approaches

Research Design

6. Overview: Strategies for Causal Attribution
7. Randomized Controlled Trials
8. Quasi-experimental Design and Methods
9. Comparative Case Studies

Data Collection and Analysis

10. Overview: Data Collection and Analysis Methods in Impact Evaluation
11. Developing and Selecting Measures of Child Well-Being
12. Interviewing
13. Modelling

12. KEY READINGS AND LINKS

Readings and links relating to specific aspects of impact evaluation are provided in the other briefs in this series. Some general introductions to impact evaluation are provided below:

IE4D Group, 'Impact Evaluation for Development: Principles for Action', IE4D Group, 2011. This discusses strategies to manage and undertake development evaluation in ways that support development. See

<http://www.scalingimpact.net/files/Impact%20Evaluation%20for%20Development%20-%20Principles%20for%20Action.pdf>.

Rogers, Patricia J., RMIT University and BetterEvaluation, 'Introduction to Impact Evaluation', *Impact Evaluation Notes* No. 1, InterAction, Washington, D.C., 2012. This provides an overview of issues for programme managers to address. See <http://www.interaction.org/sites/default/files/1%20-%20Introduction%20to%20Impact%20Evaluation.pdf>.

White, Howard, 'Some Reflections on Current Debates in Impact Evaluation', *International Initiative for Impact Evaluation Working Paper* No. 1, 2009. A summary of key debates about measuring and attributing impacts. See http://www.3ieimpact.org/media/filer/2012/05/07/Working_Paper_1.pdf.

UNICEF Evaluation Office, 'What goes into a Terms of Reference (TOR)', *Evaluation Technical Note* No. 2, UNICEF, 2002. A checklist of questions for UNICEF evaluation managers to work through when developing a Terms of Reference. See http://www.unicef.org/evaluation/files/TechNote2_TOR.pdf.

UNICEF Evaluation Office, 'Guidance on prioritization of major evaluations at the decentralized level', UNICEF, 2010, [https://intranet.unicef.org/epp/evalsite.nsf/0/089a45b4f5292588852577e00056f69c/\\$file/guidanceprioritizationofevaluations.pdf](https://intranet.unicef.org/epp/evalsite.nsf/0/089a45b4f5292588852577e00056f69c/$file/guidanceprioritizationofevaluations.pdf).

UNICEF Office of Research - Innocenti, Childwatch International Research Network, Centre for Children and Young People at Southern Cross University, Australia, and Children's Issues Centre at the University of Otago, New Zealand, 'Ethical Research Involving Children', website, 2013, www.childethics.com.

GLOSSARY

<u>Baseline data</u>	<i>Data collected before an intervention for the purposes of being compared to data collected after the intervention is implemented to measure change over time.</i>
<u>Causal attribution</u>	<i>Ascription of a causal link between observed (or expected to be observed) changes and a specific intervention.</i>
<u>Comparison group</u>	<i>In a quasi-experimental research design, this is the group of research participants that, for the sake of comparison, does not receive the treatment or intervention given to the intervention group. Comparison group subjects are typically not randomly assigned to their condition, as would be true of control group subjects in an experimental design study. See: control group, experimental group.</i>
<u>Control group</u>	<i>Participants in a research study/evaluation who do not receive the experimental treatment/intervention.</i>
<u>Dotmocracy</u>	<i>Data collection method for collecting data from individuals or groups. It collects levels of agreement on written statements among a large number of people.</i>
<u>Evaluability assessment</u>	<i>A feasibility check to determine whether a programme or policy is able to be evaluated. It determines whether an intervention is adequately defined and its results verifiable, and whether an evaluation is the best way to answer questions posed by key stakeholders. It also checks whether adequate resources – in terms of data, budget, capacity and interest in using the results – to conduct an evaluation are available.</i>
<u>Evaluation criteria</u>	<i>The values (i.e. principles, attributes or qualities held to be intrinsically good, desirable, important and of general worth) which will be used in an evaluation to judge the merit of an intervention. Examples include OECD-DAC criteria,²⁰ HRBAP and criteria used for humanitarian assistance.</i>
<u>Hierarchical card sorting</u>	<i>Data collection method that provides insight into how people categorize and rank different phenomena.</i>
<u>Inception report</u>	<i>Report produced at the end of the inception period outlining work already undertaken, preliminary results, references and a workplan for the remainder of the project, for example.</i>
<u>Indicator</u>	<i>A verifiable measure that has been selected by programme or policy management to make decisions about the programme/policy. For example, the proportion of students achieving a passing grade on a standardized test.</i>
<u>Key informant</u>	<i>Research/evaluation participants who are likely to be well informed about an issue, and willing to answer without bias.</i>
<u>Monitoring</u>	<i>The collection and analysis of data as the project progresses to assure the appropriateness of the evaluation, its design and participant protections.</i>
<u>Performance Monitoring</u>	<i>A continuous process of collecting and analyzing data to compare how well a project, programme, or policy is being implemented against expected results.</i>

²⁰ Organisation for Economic Co-operation and Development – Development Assistance Committee, ‘Evaluation of development programmes, DAC Criteria for Evaluating Development Assistance’, web page, OECD. See <http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>.

<u>Process evaluation</u>	<i>An evaluation of the development and implementation of a particular programme/intervention. It analyses the effectiveness of programme operations, implementation, service delivery, etc. Ongoing process evaluation is called programme monitoring. See: monitoring.</i>
<u>Rubric</u>	<i>An evaluation tool which uses a set of criteria and a rating/scoring guide predetermined by the evaluator(s). Rubrics can be used for evaluating presentations, projects, portfolios, and so on.</i>
<u>Terms of Reference (ToR)</u>	<i>A statement of the background, objectives, intended users, key evaluation questions, methodology, roles and responsibilities, timelines, deliverables, quality standards, evaluation team qualifications and other relevant issues which specify the basis of a UNICEF contract with the evaluators.</i>