

Evaluative Criteria

Greet Peersman

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:

Peersman, G. (2014). Evaluative Criteria, *Methodological Briefs: Impact Evaluation 3*, 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: Dugan Fraser, Patricia Rogers

Reviewers: Nikola Balvin, Samuel Bickel, David Parker

© 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. EVALUATIVE CRITERIA: A BRIEF DESCRIPTION

Evaluation relies on a combination of facts and values (i.e., principles, attributes or qualities held to be intrinsically good, desirable, important and of general worth¹ such as 'being fair to all') to judge the merit of an intervention (i.e., a programme or policy). Evaluative criteria specify the values that will be used in an evaluation. While evaluative criteria can be used in different types of evaluations, this brief specifically addresses their use in impact evaluations (i.e., studies that provide information about the long-term effects produced by an intervention; see Brief No. 1, Overview of Impact Evaluation).

UNICEF uses a range of evaluative criteria to guide its evaluations; not all of them are used in every evaluation, as some are appropriate to certain interventions and/or types of evaluation only. The [Terms of Reference](#) (ToR) for the evaluation must specify the relevant evaluative criteria to use.

Evaluative criteria used in UNICEF impact evaluations of interventions² include the following:

The standard OECD-DAC criteria³

- **Relevance:** The extent to which the objectives of an 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. Interventions must be both environmentally and financially sustainable. Where the emphasis is not on external assistance, sustainability can be defined as the ability of key stakeholders to sustain intervention benefits – after the cessation of donor funding – with efforts that use locally available resources.

Criteria about equity, gender equality and taking a human rights-based approach to programming (HRBAP)

- **Equity:** The basic fairness of the processes and outcomes of decision making. For UNICEF, this implies that all children have an opportunity to survive, develop and reach their full potential, without being subjected to discrimination, bias or favouritism.

¹ Stufflebeam, Daniel L., 'Evaluation values and criteria checklist', Western Michigan University Checklist Project, Kalamazoo, 2001. See http://www.wmich.edu/evalctr/wp-content/uploads/2010/05/values_criteria.pdf.

² Where there are important differences in how evaluative criteria are used in the evaluation of either a programme or a policy, they will be specifically mentioned in this brief.

³ 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>.

- **Gender equality:** For UNICEF, this means promoting the equal rights of women and girls, and supporting their full participation in the political, social and economic development of their communities.⁴
- **HRBAP:** Five core guiding principles underpin HRBAP and can be used as evaluative criteria – normativity, non-discrimination, participation, transparency and accountability.⁵

Evaluative criteria⁶ for humanitarian assistance (used in addition to some or all of the above)

- **Coverage:** The need “to reach major population groups facing life-threatening suffering wherever they are, providing them with assistance and protection proportionate to their need and devoid of extraneous political agendas”.⁷
- **Coordination:** The need for different actors involved in an emergency response to coordinate; the intervention of a single agency cannot be evaluated separately as what may seem appropriate from one actor’s point of view may be inappropriate from the point of view of the overall system.
- **Protection:** The provision of security and protection of the target population (such as protection from armed elements operating within the project area, or providing security within a displaced persons/refugee camp) is critical to the effectiveness of humanitarian action/relief assistance.
- **Coherence** (i.e., policy coherence): The need to ensure consistency across security, development, trade and military policies and humanitarian policies.

These evaluative criteria should be thought of as ‘concepts’ that must be addressed in the evaluation. They are insufficiently defined to be applied systematically and in a transparent manner to make evaluative judgements about programmes or policies, however. Under each of the ‘generic’ criteria, more specific criteria such as benchmarks and/or standards⁸ – appropriate to the type and context of the intervention – should be defined and agreed with key stakeholders.

For example, in the case of a global evaluation of the application of HRBAP to UNICEF programming, the evaluators identified different understandings of what constitutes HRBAP. They conducted consultations and consensus building to prioritize five core principles (as listed above) and subsequently defined key indicators to assess the specific UNICEF programmes under evaluation.

⁴ See: United Nations Children’s Fund, Working for an Equal Future, UNICEF Policy on Gender Equality and the Empowerment of Girls and Women, UNICEF, New York, 2010.

⁵ See: United Nations Children’s Fund, Global Evaluation of the Application of the Human Rights-based Approach to UNICEF Programming, Final Report – Volume I, UNICEF, New York, 2012. (Note: a sixth principle of indivisibility/interdependence was considered to be inherent in the principle of normativity.)

⁶ Paraphrased from the chapter ‘Linking evaluation criteria to evaluation questions: Additional criteria for humanitarian assistance programmes’ in the *UNICEF Monitoring and Evaluation Training Resource*. The document is still under revision; no page numbers were indicated and only an excerpt was provided.

⁷ Organisation for Economic Co-operation and Development Development Assistance Committee, *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies*, OECD-DAC, Paris, 1999.

⁸ A benchmark or index is a set of related indicators that provides for meaningful, accurate and systematic comparisons regarding performance; a standard or rubric is a set of related benchmarks/indices or indicators that provides socially meaningful information regarding performance (see Brief No. 11, Developing and Selecting Measures of Child Well-being and Brief No. 4, Evaluative Reasoning).

Main points

1. Evaluative criteria specify the values that will be used in an evaluation of interventions (i.e., programmes and policies).
2. UNICEF uses a range of evaluative criteria to guide its evaluations; not all of them are used in every evaluation, as some are appropriate to certain interventions and/or types of evaluation only. Impact evaluations should consider:
 - relevance, effectiveness, efficiency, impact and sustainability
 - equity, gender equality and human rights
 - coverage, coordination, protection and coherence (especially in regard to humanitarian assistance).
3. 'Generic' evaluative criteria must be qualified with more specific detail in the ToR of an evaluation or during the initial stages of the evaluation process to apply them systematically and in a transparent manner to make evaluative judgements about the intervention.

2. WHEN IS IT APPROPRIATE TO USE EVALUATIVE CRITERIA?

All impact evaluations conducted in the UNICEF context should consider the OECD-DAC criteria and identify which of these are relevant to the particular evaluation. The OECD-DAC criteria reflect the core principles for evaluating development assistance⁹ and have been adopted by most development agencies as standards of good practice in evaluation. Criteria of equity, gender equality and human rights are also a requirement for UNICEF impact (and other types of) evaluations.

Depending on the type of intervention (e.g., a case of humanitarian assistance, as referred to above) and/or the type of evaluation (e.g., a process evaluation), additional criteria may apply or particular criteria should be focused on. For example, the OECD-DAC criterion of impact is irrelevant to a process evaluation, as this type of evaluation looks at how an intervention is being implemented (e.g., how services are delivered, whether clients are satisfied with the services provided, what the management practices are) rather than whether or not it has produced the intended results.

3. HOW TO USE EVALUATIVE CRITERIA

Linking evaluative criteria to evaluation questions

The ToR for the evaluation should specify the overall purpose and scope of the evaluation as well as the generic evaluative criteria to be addressed. The ToR should also set out the [key evaluation questions](#) (KEQs) although these may be developed or refined further as the first step in the evaluation process (see Brief No. 1, Overview of Impact Evaluation).

To keep the impact evaluation focused, only a limited set of KEQs or high-level (macro level) evaluation questions should be specified. For example:

- KEQ1. What was the quality of the intervention design/content? Relevance, equity, gender equality, HRBAP.

⁹ Organisation for Economic Co-operation and Development – Development Assistance Committee, *Principles for Evaluation of Development Assistance*, OECD-DAC, Paris, 1991. See <http://www.oecd.org/dac/evaluation/50584880.pdf>.

- KEQ2. How well was the intervention implemented and adapted as needed? Effectiveness and efficiency.
- KEQ3. Did the intervention produce the intended results in the short, medium and long term? If so, for whom, to what extent and in what circumstances? Effectiveness, impact, equity, gender equality.
- KEQ4. What unintended results – positive and negative – did the intervention produce? How did these occur? Effectiveness, impact, equity, gender equality, HRBAP.
- KEQ5. What were the barriers and enablers that made the difference between successful and disappointing intervention implementation and results? Relevance, equity, gender equality, HRBAP.
- KEQ6. How valuable were the results to service providers, clients, the community and/or organizations involved? Relevance, equity, gender equality, HRBAP.
- KEQ7. To what extent did the intervention represent the best possible use of available resources to achieve results of the greatest possible value to participants and the community? Efficiency.
- KEQ8. Are any positive results likely to be sustained? In what circumstances? Sustainability, equity, gender equality, HRBAP.

A range of more detailed (mid-level and lower-level) evaluation questions should then be articulated to address each evaluative criterion in detail. All evaluation questions should be linked explicitly to the evaluative criteria to ensure that the criteria are covered in full. Boxes 1 to 5 give examples of the OECD-DAC criteria of mid-level (meso level) questions that are then unpacked further into lower-level (micro level) questions. The terms high/macro, mid/meso and lower/micro level are not intended to reflect a hierarchy of importance but rather increasing levels of specification or operationalization.

Box 1. Examples of meso and micro level evaluation questions to address 'relevance'

Was the intervention aligned to the country's commitments and strategic plan for increasing children's welfare?

- Were the objectives of the intervention relevant to identified national needs and priorities?

Was the intervention design technically sound?

- Were the objectives and the design of the intervention relevant to the context and to the needs of recipients?
- Was there a clear intervention logic and locally relevant evidence base connecting intervention activities and operations with the realities faced by recipients? Was a clear rationale provided for the selected intervention activities?
- Were the activities and outputs of the intervention consistent with the overall goal and the attainment of its objectives? Were they also consistent with the intended outcomes and impacts?
- Did the allocated funding adequately reflect the level of needs and the operating environment?

Was the intervention supportive of gender equality and other human rights standards?

- To what extent were equity, gender equality and HRBAP goals and processes incorporated into the planning of the intervention?
- Was the intervention designed to provide for equal participation by all relevant groups (e.g., men and women, boys and girls)?
- Did the intervention promote more equal access by men and women, boys and girls to the benefits of the activity, and more broadly to resources, services and skills?

Was the intervention adjusted throughout its implementation period to align it with emerging priorities/needs and to ensure support for best practice?

- To what extent did the intervention include activities and processes/mechanisms to support the effective design and implementation of the programme and to elicit feedback?

Box 2. Examples of meso and micro level evaluation questions to address 'effectiveness'

Was the intervention implemented according to plan?

- Was the intervention implemented according to plan? If not, why not? And what was done about it?
- To what extent did contextual factors help or hinder intervention implementation?

Was timely corrective action taken where necessary?

- Was additional support identified or provided to overcome implementation challenges? What form did this support take, who provided it and to what effect?
- What were the risks to achieving the intervention's objectives? Were the risks managed appropriately?

Were intended results achieved?

- To what extent were the objectives of the intervention achieved? What were the major factors influencing the achievement or non-achievement of objectives?

Box 3. Examples of meso and micro level evaluation questions to address 'efficiency'

Did the intervention use the available resources in the most economical manner to achieve its objectives?

- Were objectives achieved on time?
- Did the intervention have sufficient and appropriate staffing resources? Did the implementation of the intervention make effective use of time and resources to achieve results?
- To what extent has effective coordination and collaboration with existing interventions and partners been addressed and achieved?

- Was the intervention designed and/or amended throughout the implementation period to provide the best value for money?
- Has the intervention been shown to be as efficient or more efficient than appropriate alternative approaches in the same context?
- Are there other feasible ways to implement the intervention that would be more economical?

Box 4. Examples of meso and micro level evaluation questions to address 'impact'

Was the intervention scaled up sufficiently to achieve the intended impacts (e.g., key health indicators, inequities)?

- Was there evidence of change – positive or negative? If so, what contributed to this change? If not, why not?
- Were there any unintended changes – positive or negative – in the lives of recipients and in their environment? What were they? Were they directly or indirectly related to the programme or due to external factors?

Were results achieved in adherence to equity, gender equality and other human rights?

- How well did the intervention succeed in involving women and men, and rights-holders as well as duty-bearers, especially the most vulnerable?
- To what extent did different groups, including children and other vulnerable groups, benefit in different ways from the intervention?
- To what extent did different groups, including children and other vulnerable groups, increase their capacity to relate differently to other groups in the intervention?
- How were the costs and burdens of participation in the intervention distributed across different groups of participants? What about the comparative benefits, if any, arising from their participation?

Did the operating context have an influence on the intervention or vice versa?

- What role did the local and national context play in either supporting or hindering change?
- Did the intervention have an impact on the broader operating environment for working towards children's welfare in the country? If so, what form did this take?

Box 5. Examples of meso and micro level evaluation questions to address 'sustainability'

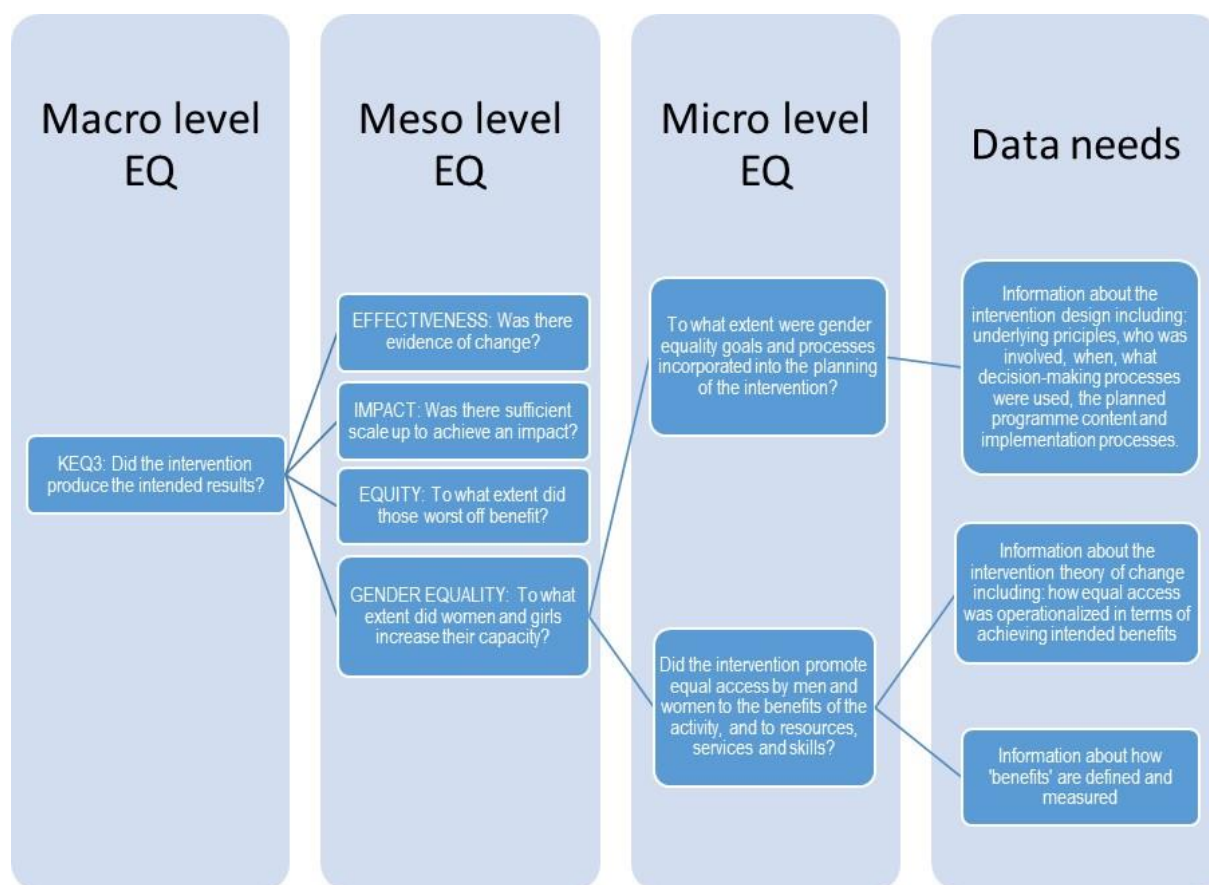
Were results achieved in a sustainable manner?

- To what extent can the activities and the benefits of the intervention continue after external funding has ceased?
- Are any areas of the intervention clearly unsustainable? What lessons can be learned from such areas?
- To what extent do intervention recipients and/or partner country stakeholders have ownership, capacity and resources to maintain the activity results after external funding ceases?
- Did the intervention contribute to capacity building of local organizations to continue to deliver quality services for children? If so, what form did this take?
- What were the major factors that influenced the achievement or non-achievement of sustainability of the intervention?

An evaluation plan should be developed in response to the evaluation ToR. This plan should describe the key and detailed evaluation questions linked to generic and specific evaluative criteria (or how the latter will be determined as part of the evaluation process); the data needed to answer each of the evaluation questions, and the associated evaluation research design (see Brief No. 6, Overview: Strategies for Causal Attribution); and data collection/collation and analysis methods (see Brief No. 10, Overview: Data Collection and Analysis Methods). The evaluation plan should also include responsibilities (i.e., who will do what), timelines and deliverables so that it can also be used by evaluation commissioners/managers to keep the evaluation process on track.

Figure 1 summarizes how evaluative criteria link to evaluation questions (at the macro, meso and micro level) and data needs for the example of KEQ3 (see above): Did the intervention produce the intended results in the short, medium and long term? If so, for whom, to what extent and in what circumstances?

Figure 1. Illustrative example of linking an evaluation question to an evaluative criterion and specific data needs



Note: EQ = evaluation question.

Using specific benchmarks or standards to judge intervention performance

Good evaluation questions are not just about ‘What were the results?’ (i.e., descriptive questions about how things are and what has happened) but also ‘*How good* were the results?’ (i.e., evaluative questions about the overall conclusion as to whether or not a programme or policy achieved what it set out to do well).

The ToR for the evaluation may include specific benchmarks or standards by which the intervention must be judged, or if these do not already exist, how they should be identified and agreed upon during the evaluation process. For example, if effectiveness is defined as the extent to which objectives have been achieved and one of the objectives is ‘increased time in school for children’, then the specific criteria for judging the worth of the intervention may focus on school retention (i.e., staying on until the later years) and school attendance (e.g., all children attending at least 90 per cent of the time, or 90 per cent of students attending all of the time).

Benchmarks and/or standards exist for particular sectors and/or settings relating to children. These are usually based on evidence and developed in consultation with relevant stakeholders. The key purpose of benchmarks and/or standards is to serve as specific (realistic rather than aspirational) targets against which progress can be monitored (locally, nationally and/or globally) and achievements evaluated. When

they are well defined and implemented accordingly, they can be used to compare and learn from performance across programmes/policies, situations and/or specific geographical areas. For example, benchmarks and/or standards were defined for:

- [WHO immunization coverage targets](#): The rationale is that minimum coverage is crucial for the effective prevention of disease outbreaks.
- [UNICEF national child protection systems](#) (see box 6): The rationale is that national child protection systems are recognized as the most effective and sustainable means to protect children from all forms of maltreatment. The absence of such a system frequently leads to incomplete coverage, and to individual violations being addressed as separate, unrelated problems.¹⁰

Evaluative criteria are not only useful for evaluating interventions but they should also be explicit in the [theory of change](#) (see Brief No. 2, Theory of Change) and operationalized in the intervention content and in the way the intervention is implemented. Box 6 outlines a UNICEF example of proposed benchmarks and standards for evaluations in the child protection sector.

Box 6. Benchmarks and standards used in child protection service delivery and evaluation

Goal

To support personnel working in child protection service delivery, it is critical for countries to establish comprehensive systems that include the respective legal requirements, a continuum of related education and a system of regular inspection of skills levels.

Programme benchmarks

UNICEF EAPRO proposed the following **benchmarks** for a professional education and continuing development system:

- A university degree programme in social work, the curriculum for which includes courses on social services, developmental issues, protective and preventive topics, and therapeutic interventions.
- A vocational qualification programme in social work or child development, which has a curriculum approved by relevant authorities.
- A system of accreditation of social work skills, which is based on competency tests within relevant training programmes.
- Training on tackling abuse, violence and exploitation for education workers (such as teachers), health professionals and/or other professionals who work with children.
- A specific training programme for staff within ministries who hold a lead interior and home affairs role or have lead justice roles on children and justice.

Evaluative judgements

Standards were then specified to assess the status of a country's professional training system:

- Standard A education system includes **all** of the items in the benchmark criteria.
- Standard B education system includes **at least three** of the items in the benchmark criteria.
- Standard C education system includes **one or two** of the items in the benchmark criteria.

¹⁰ United Nations Children's Fund, *Measuring and Monitoring Child Protection Systems: Proposed Regional Core Indicators for East Asia and the Pacific*, Strengthening Child Protection Systems No. 3, UNICEF EAPRO, Bangkok, 2012, foreword, p. vii. See http://www.unicef.org/eapro/Measuring_and_monitoring.pdf.

- Standard D education system includes **none** of the items in the benchmark criteria.

An example of how these standards may be used to make a judgement about the worth of a country's professional development programme might be by informing a rating system. For example:

- **Excellent** rating – clear example of exemplary performance or best practice in this domain; no weaknesses; meets standard A.
- **Very good** rating – very good or excellent performance on virtually all aspects; strong overall, but not exemplary; no weaknesses of any real consequence; meets standard B and, additionally, has passed all specified quality requirements for training.
- **Unacceptable** rating – major weaknesses in all areas of the programme; meets standard C or D.

Source: United Nations Children's Fund, *Measuring and Monitoring Child Protection Systems: Proposed Regional Core Indicators for East Asia and the Pacific*, Strengthening Child Protection Systems No. 3, UNICEF EAPRO, Bangkok, 2012, foreword, p. vii. See http://www.unicef.org/eapro/Measuring_and_monitoring.pdf.

Thinking through how to define – and thus how to judge – whether or not a programme or policy is successful should be done up front. That way, data collection can be geared to the mix of evidence required to make appropriate performance judgements. It is also more transparent and objective to define 'success' before evidence is gathered. Involving relevant stakeholders in defining 'what success looks like' for a certain intervention in its particular context improves the acceptance of evaluation findings (by reducing misunderstandings and disagreements about interpretation later on) and supports a commitment to taking the necessary action in regard to the programme. Brief No. 4, Evaluative Reasoning sets out a step-wise approach for deciding (1) on what dimensions the intervention must do well (i.e., establish specific criteria of worth about 'equity'), and (2) how well the intervention must perform on those dimensions (i.e., construct a standard or [rubric](#) or use an existing one).

4. ETHICAL ISSUES AND PRACTICAL LIMITATIONS

Ethical issues

All UNICEF evaluations should follow the *Ethical Guidelines for Evaluation*¹¹ set out by the United Nations Evaluation Group (UNEG). These guidelines specify both a code of conduct for evaluators as well as ethical obligations to participants in the evaluation (see also Brief No. 1, Overview of Impact Evaluation). The latter include: respect for dignity and diversity; rights; confidentiality; and 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.

The UNICEF Office of Research - Innocenti released in October 2013 [Ethical Research Involving Children](#), an online resource¹² that brings together expert thinking on key ethical issues involving children and how these might be addressed in different research (and evaluation) contexts.

¹¹ United Nations Evaluation Group, *Ethical Guidelines for Evaluation*, UNEG, New York, 2007.

¹² 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

In addition to the general ethical issues referred to above, there are particular issues relating to how evaluative criteria are chosen, including:

- Who should be involved in defining success? It must also be decided who has the final say if conflicting priorities/views arise or if contradictory evidence emerges.
- How to define success? Defining success only in terms of 'average effects' or 'aggregate social indicators' may omit the experience of certain subgroups (such as boys or girls, the most vulnerable or the poorest).
- How were the intended results achieved? For example, is it unethical to increase vaccination coverage through authority-prescribed participation to combat a disease outbreak?

Practical limitations

Common practical limitations in addressing evaluative criteria include:

- The difficulty of moving from the generic wording of evaluative criteria (such as the OECD-DAC criterion of 'sustainability') to more specific criteria that take into account different stakeholders' views on what are important dimensions and what thresholds the intervention should surpass.
- A mismatch between the timing of the impact evaluation and when relevant impacts or sustainability evidence will be visible. Measuring 'proxies' (i.e., what is feasible to measure rather than what is ideally measured) may further complicate definitions and negotiations about what success looks like.
- There is often limited information available about the quality of certain components or services implemented as part of a large-scale programme. This information is difficult to reconstruct or understand retrospectively, and thus hard to capture adequately when specifying evaluative criteria.

If the resources and time frame for the evaluation allow, some of these limitations may be rectified (e.g., by extending the evaluation time frame or scope) but it is likely that conclusions drawn will have to be based on imperfect information. Hence, it is important to be transparent about the limitations of the evaluation and to describe how these may have affected the overall findings, conclusions and recommendations.

5. WHICH OTHER METHODS WORK WELL WITH THIS ONE?

All evaluations are tasked with asking and answering evaluative questions – not simply with measuring and describing results. Therefore all findings must be interpreted within an evaluative frame. This means not only saying what the results are, but also how good they are. Evaluative reasoning is needed to do this and is a requirement of all evaluations, irrespective of the methods or approach used (see Brief No. 4, Evaluative Reasoning).

As many impact evaluations in a UNICEF context will rely on using existing data sources, it is also important to include a standards-based assessment¹³ of the quality of the monitoring and evaluation system linked to the programme and wider policy environment. Using data of an unknown quality may compromise the validity of the evaluation findings and may result in misleading conclusions. Evaluators should use or collect the best quality data and make certain that the implications of any data quality concerns are minimized, where possible, and specifically noted in the evaluation report.

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](#)').

¹³ Tools based on agreed global standards for assessing a monitoring and evaluation system (including national, sub-national and service delivery levels) are available. For example, Joint United Nations Programme on HIV/AIDS, *12 Components Monitoring and Evaluation System Strengthening Tool*, UNAIDS, Geneva, 2009. See http://www.unaids.org/en/media/unaids/contentassets/documents/document/2010/2_MERG_Strengthening_Tool_12_Component_s_ME_System.pdf.

6. PRESENTATION OF RESULTS AND ANALYSIS

The evaluation report should be structured in a manner that reflects the purpose and KEQs of the evaluation (see also Brief No. 1, Overview of Impact Evaluation).

In the first instance, evidence to answer the detailed questions linked to the OECD-DAC criteria of relevance, effectiveness, efficiency, impact and sustainability, and considerations of equity, gender equality and human rights should be presented succinctly but with sufficient detail to substantiate the conclusions and recommendations. Laying out the evidence in this way will also help the [evaluation management team](#) to ensure that the criteria stipulated in the ToR are fully addressed.

The specific evaluative rubrics should be used to 'interpret' the evidence and determine which considerations are critically important or urgent. Evidence on multiple dimensions should subsequently be synthesized to generate answers to the high-level evaluative questions. More specific recommendations on how to write a report that is strong on evaluative synthesis are provided in Brief No. 4, Evaluative Reasoning (e.g., using explicitly evaluative language when presenting findings, rather than value-neutral language that merely describes findings).

7. EXAMPLES OF GOOD PRACTICES

Some specific examples of good practices are outlined in box 7.

Box 7. Example from an ongoing impact evaluation of joint gender programmes in the UN system

Good practices in the Terms of Reference

Seven KEQs were included. These were directly linked to the strategic priorities and emerging issues identified in the programme portfolio analysis, for example:

- To what extent have joint gender programmes (JGPs) been conceptualized, planned and designed jointly to respond to international, regional and national commitments on gender equality and women's empowerment (GEWE); to establish coherence and capitalize on the comparative advantages of participating UN agencies; and to integrate a human rights based approach to programming (HRBAP)?
- What are the key contributions and the added value in terms of short- and long-term, intended and unintended, positive and negative GEWE results achieved by JGPs to date at the national level?
- What is the influence of the specific country context and circumstances (e.g., conflict/post-conflict, fragile state, 'Delivering as One' status) on the achievement of JGP results and operational effectiveness?

The evaluation criteria for the assessment of JGPs were also outlined and defined within the context of the programme (rather than just using the generic description in the OECD-DAC criteria). For example:

- Relevance/coherence of the JGP planning, design and implementation processes to regional, national and international commitments, policies and priorities; to aid effectiveness principles; to UN mandates and the United Nations Development Assistance Framework; and to individual agency policies, mandates and comparative advantages in terms of their responsiveness and alignment with country needs on GEWE.

- Effectiveness and impact of JGPs in achieving their stated objectives on GEWE and any intended or unintended long-term effects at the national level and for gender mainstreaming within the UN system, including the use of innovative approaches.
- Sustainability of the results of JGPs given the level of national ownership generated, effective partnerships established and national capacity strengthened through JGP processes.

Additional criteria were introduced, for example, participation and inclusion of national duty-bearers and rights-holders – specifically those most marginalized – in JGP processes.

The evaluation planning period also allowed for further evaluation criteria to be added before concluding the final scope of the evaluation.

In addition to the KEQs, more detailed corresponding questions were included in a ‘matrix of evaluation questions’ in an annex to the ToR.

Source: See evaluation-related terms of reference and thematic reports linked to the evolving United Nations Population Fund, ‘Joint Evaluation of Joint Gender Programmes in the UN System’, web page, UNPFA, <http://www.unpfa.org/public/home/about/Evaluation/EBIER/TE/pid/10096>.

8. EXAMPLES OF CHALLENGES

Some common weaknesses relating to addressing evaluative criteria were identified, for example, in the quality review of all 2012 evaluation reports.¹⁴

- “Failure to specify the underlying questions and criteria guiding the evaluation.”
- “Integration of the Human Rights-Based Approach to Programming (HRBAP), gender equality and equity dimensions continues to be generally weak in evaluations (i.e., the ratio of reports that successfully integrated these elements did not rise above 50%; gender was addressed in 46% of reports, 44% paid attention to HRBAP, and 41% incorporated a greater focus on equity)”.

It was also noted: “Evaluation reports tend to reflect what evaluators were asked to consider, including the suggested methods, evaluation criteria, data sources, and questions underscored in the [terms of reference]. As such, even though gender, human rights and equity concerns may be central to the values of the Agency, whether or not these are addressed in the evaluation reports tends to be dictated by the [terms of reference].”

Hence, it is important to provide consistent and explicit guidance, ToR and feedback to evaluators throughout the evaluation process.

9. KEY READINGS AND LINKS

BetterEvaluation, ‘Define ethical and quality evaluation standards’, web page, BetterEvaluation, http://betterevaluation.org/plan/manage_evaluation/ethical_evaluation.

BetterEvaluation, ‘Establish Decision Making Processes’, web page, BetterEvaluation, http://betterevaluation.org/plan/manage/who_controls.

¹⁴ A total of 79 evaluation reports were included in this review. Quotes are taken from the Executive Summary (p. iii) in Universal Management Group, ‘GEROS – Global Meta-Evaluation Report 2012’, final report, Universal Management Group, 2013. See http://www.unicef.org/evaldatabase/files/2013_GEROS_MetaEvaluation_finalReport.pdf.

BetterEvaluation, 'Evaluation Plan', web page, BetterEvaluation, http://betterevaluation.org/evaluation-options/evaluation_planning_templates.

BetterEvaluation, 'Specify the Key Evaluation Questions', web page, BetterEvaluation, http://betterevaluation.org/plan/engage_frame/decide_evaluation_questions.

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

Organisation for Economic Co-operation and Development – Development Assistance Committee, *Principles for Evaluation of Development Assistance*, OECD-DAC, Paris, 1991. See <http://www.oecd.org/dac/evaluation/50584880.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.

United Nations Evaluation Group, *Ethical Guidelines for Evaluation*, UNEG, New York, 2007. See <http://www.unevaluation.org/ethicalguidelines>.

GLOSSARY

| | |
|---|---|
| <p><u>Evaluation management team (EMT)</u></p> | <p><i>A team in charge of managing an evaluation. At UNICEF, its tasks include: selecting and supervising the work of the (usually external) evaluators; reviewing the ToR, inception report and final draft; providing technical guidance to the evaluators; managing operational aspects of the evaluation (budget, field visits, contact with relevant UNICEF counterparts); and quality assurance. In a joint evaluation, an EMT would comprise of UNICEF staff and representatives of each partner organization.</i></p> <p><i>Related: Evaluation steering committee.</i></p> |
| <p><u>Key evaluation questions (KEQs)</u></p> | <p><i>High-level (macro level) evaluation questions about overall performance, which the evaluation should aim to answer. KEQs are derived from the purpose of the evaluation.</i></p> |
| <p><u>Rubric</u></p> | <p><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></p> |
| <p><u>Terms of Reference (ToR)</u></p> | <p><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></p> |
| <p><u>Theory of Change</u></p> | <p><i>Explains how activities are understood to produce a series of results that contribute to achieving the final intended impacts. It can be developed for any level of intervention – an event, a project, a programme, a policy, a strategy or an organization.</i></p> |