



Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management

August 2013

This Guidance Document was prepared by the UNEG Impact Evaluation Task Force. Helpful guidance from Dr. David Todd, Dr. Patricia Rogers, Burt Perrin, Dr. Michael Spilsbury and Dugan Fraser is gratefully acknowledged.

Contents

Summary.....	4
Introduction	5
1 Definitions and role of impact evaluation	6
2 Impact Evaluation Design	11
2.1 Range of Design Approaches	12
2.2 Theory of Change	14
2.3 Evaluability Assessment in Impact Evaluation Planning	17
2.4 Gender Equality and Human Rights	18
3 Common Methods in Impact Evaluation.....	19
3.2 Quantitative Methods	19
3.3 Qualitative Methods	21
3.4 Participatory Methods to Establish Stakeholder Perceptions	21
3.5 Methods and Validity	22
3.6 Choosing the Mix of Methods for Impact Evaluation.....	23
4 Quality Control for Impact Evaluations	24
4.1 Specific Quality Control Criteria at the Design Stage.....	24
4.2 Quality Control Requirements and Approaches for Impact Evaluation.....	25
4.3 Quality Control of Evaluation Standards.....	26
4.4 Managing a Quality Control System	27
5 Impact Evaluation of Normative Work	29
6 Impact Evaluation of Multi-Agency Interventions.....	30
6.1 Types of Multi-Agency Interventions	31
6.2 Impact Evaluation Issues Specific to Multi-Agency Interventions	34
6.3 Agreement on Purpose and Roles in Multi-Agency Impact Evaluations	35
Annex 1: Works Cited	38
Annex 2: Agency Specific Definitions of Impact cited by UNEG Members.....	41

Summary

This Guidance Note, used in conjunction with many other recent resources on impact evaluation, provides a sound starting point for UN evaluation bodies wishing to commence conducting impact evaluations.

A summary of the key points:

- There is rising interest and a growing body of expertise and experience in Impact Evaluation among evaluators in the UN system.
- The concept of impact used by most UNEG member bodies is derived from the “DAC definition”.
- Impact evaluation can be used for different purposes. Accountability and lesson learning are two aspects, which have been emphasized. The evaluation purpose should form the basis of its design and methods.
- A fundamental element of impact evaluation is establishing cause and effect chains to show if an intervention has worked and, if so, how.
- Different impact evaluation designs provide varying approaches to establishing how and to what extent, interventions have caused anticipated and/or unanticipated effects.
- A “mixed method” approach utilizing quantitative, qualitative, participatory and blended (e.g. quantifying qualitative data) approaches is now widely accepted as advisable to address the types of interventions that are now predominant in international development.
- A Theory of Change approach has become accepted as a basic foundation for most types of Impact Evaluation.
- Impact evaluation of UN normative work needs to go beyond establishing institutional impact to identify changes in people’s lives.
- Quality control is very important for impact evaluation and systems need to be specified and managed to different aspects and characteristics of such evaluations.
- Joint impact evaluation of Multi-Agency Interventions can deliver additional findings, beyond those arising from the evaluation of individual components. However, they also have costs and must be systematically managed.

Introduction

The purpose of this guidance note is to describe and define impact evaluation for member organizations of the UN Evaluation Group (UNEG); and to articulate some of the main theoretical and practical considerations when carrying out impact evaluations.

Interest in impact evaluation has arisen in response to increasing emphasis in international development circles on the principles of Evidence Based Policy and Results Based Management. At the same time, understanding of the role of development assistance has changed, with an increased perception that aid rarely achieves results on its own. Rather, development is attained as a result of strong national ownership and leadership of change processes, supported by international partners, who should operate in a harmonized fashion in order to maximize the benefits of their support.

Impact evaluation has come under increasing scrutiny, since its elevated profile has appeared in parallel with enhanced understanding of the complexity of the issues it addresses, as a result of substantial and heated debate among practitioners and development institutions.

UNEG created the Impact Evaluation Task Force (IETF), which has been exploring the issues around Impact Evaluation in the UN system since 2009. It initially conducted research among UNEG member evaluation units to establish the current status of and experience with impact evaluation in their programmes. On the basis of this, a Concept Note was circulated to set the ground for future work on the issue. This work has proceeded through a substantial exercise of desk research, drafting and consultation among IETF members, culminating in this Guidance Note.

At the same time, UNEG created other bodies, notably on Multi-Agency Interventions and on UN Normative Work, whose findings related to impact evaluation have been summarized in this Guidance Note. The Note also draws on many other recent documents on impact evaluation and seeks to provide an introduction to the topic, without going into extensive details of specific design and methodological issues. These are to be found in numerous more detailed papers, which are cross-referenced in the text, for those who want to explore particular topics in more depth.

1. Definitions and Role of Impact Evaluation

Attempts to establish one universally agreed definition of impact evaluation have not been productive. This is because different, but overlapping elements of such evaluations have been emphasized by various stakeholders. Furthermore, methodological discussions around impact evaluation have raised fundamental and sensitive issues of the relationship between qualitative and quantitative methods in the social sciences, which cannot be resolved in the evaluation arena.

A paper published by the Center for Global Development in 2006 claimed that there is an absence of strong evidence on what works or does not work in the international development arena.¹ This sparked a heated debate among practitioners, notably between those who claimed the exclusive right to be considered “rigorous” because of their adoption of the methodology of Randomized Control Trials and those who considered that a broad range of other methods can also be pursued in a rigorous manner. Over time, discussions have become more balanced and several recent papers have provided useful overviews of the range of methods in common use in impact evaluation. This Note draws upon some of these recent documents and tries to make use of those elements which are most relevant to UNEG members.

In terms of definitions, the main debates have focused around two types. The first of these has come to be known as “the DAC definition”. This was not a definition formally approved or prescribed as correct by the DAC. Rather, it was a formulation, which received the assent (or at least no objection), of the then 30 DAC member states and agencies, (including representatives of the UN system and Development Banks), for inclusion in its Glossary of Evaluation Terms.² The DAC defines impact as: “*Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended*”. The DAC definition of impact forms the core of many definitions of impact evaluation adopted by development institutions, often with minor modifications or additions.³

This definition has several important elements. Impact is about “effects produced by a development intervention”. It is therefore about “cause and effect” and thus specifically addresses the issue of *attribution*,⁴ which incorporates the concept of *contribution*. The latter concept has been widely adopted among UN implementers and evaluators as providing an accurate approach to assessing the difference most UN interventions make. However, it should be noted that attribution-based definitions of impact do not require that effects be produced solely or totally by the intervention. They anticipate the co-existence of other causes,

¹ Center for Global Development. *When will we ever learn?* Washington DC, 2006.

² Development Assistance Committee, Organisation of Economic Cooperation and Development. *Glossary of Key Terms in Evaluation and Results Based Management*. Paris, 2001.

³ Annex 1 lists some definitions of impact evaluation used by UN Agencies.

⁴ The DAC Glossary defines attribution as the “ascription of a causal link between observed (or expected to be observed) changes and a specific intervention”.

so that the intervention will have contributed to the demonstrated effects. The DAC impact definition specifically includes the possibility of partial attribution, or contribution, through its inclusion of secondary and indirect effects.

Another important aspect of the DAC definition of impact is that it focuses on “long term effects”. According to the DAC Glossary, *outcomes* are the “likely or achieved short-term and medium-term effects of an intervention’s outputs”. The DAC definition therefore draws attention to a longer time scale, in which short and medium term effects (outcomes) have played some part in the generation of “long-term effects” (impacts). It should be noted that the concept of a “long-term effect” does not define when in the overall results chain such an effect can begin, but highlights its duration.

Additional aspects of the definition, which need to be addressed by any impact evaluation are negative and unanticipated consequences of an intervention. These are different and both can be important in any intervention. As an example of negative, but anticipated effects we can consider infrastructure projects; such as roads, dams and storm water drains. It is known in advance that such projects may require some people to be relocated; and measures are built into the overall implementation plan to mitigate the harmful effects through compensation and support measures. Any impact evaluation therefore needs to assess to what extent the negative aspects have been appropriately addressed.

A GEF biodiversity project offers an example of unanticipated negative consequences. The project aimed to generate income for a Protected Area and surrounding communities through eco-tourism activities. However, an offshoot of these activities was that local indigenous people became involved in alcohol abuse and sexual services, with associated health effects. The GEF impact evaluation of the project commissioned an additional specialist study to assess these effects⁵, so that they could be included in the overall evaluation of the results of the intervention.

The second main strand of definitions focuses on specifically comparing the differences between what actually happened and what would have happened without the intervention, through the specification of some form of “counterfactual”.

The International Initiative for Impact Evaluation (3i) impact evaluation⁶ definition of impact in its Impact Evaluation Glossary⁷ is similar to that of the DAC, namely: “How an intervention alters the state of the world. *Impact evaluations* typically focus on the effect of the *intervention*

⁵ GEF Evaluation Office. *Impacts of Creation and Implementation of National Parks and of Support to Batwa on their Livelihoods, Well-Being and Use of Forest Products*. Namara, A. 2007.

⁶ The 3ie is an organization which was founded as part of the process of highlighting the importance of impact evaluation in the international development community’s moves towards enhanced use of Results Based Management and Evidence Based Policy principles.

⁷ 3ie. *3ie impact evaluation glossary*. International Initiative for Impact Evaluation: New Delhi, India. 2012.

on the *outcome* for the *beneficiary population*". The core concept associated with this approach is that of attribution, which the 3ie Glossary defines as: "The extent to which the observed change in *outcome* is the result of the *intervention*, having allowed for all other factors which may also affect the *outcome(s)* of interest".

Although neither the DAC Glossary, nor the 3ie Glossary of evaluation terms has a specific entry for contribution, both of their definitions of attribution incorporate this concept. In considering available terminology relevant to impact evaluation, it is therefore clear that there is no need for a separate definition of contribution, since it is already covered under attribution.

Whereas the DAC Glossary has no specific definition of Impact Evaluation, the 3ie Glossary does: a "study of the *attribution* of changes in the *outcome* to the *intervention*. Impact evaluations have either an *experimental* or *quasi-experimental* design". It therefore specifies that, in order to qualify as an impact evaluation, methods based on comparison between the "factual" and a counterfactual established through experimental design or statistical controls counterfactual must be used. It is mainly on this issue that the (polemical) debates on impact evaluation in recent years have centered. Some of those advocating a statistical counterfactual have claimed for their work the exclusive right to be considered "rigorous". According to this view, only particular quantitative social science methods have "rigour," whilst the results of qualitative or simple statistical analysis can be considered inexact or impressionistic.

In considering the heated debates on impact evaluation, it can therefore be said that there are (at least) two common approaches, which have been considered by their proponents to be examples of Impact Evaluation. The common element is a strong focus on tracing cause and effect, to demonstrate if an intervention actually produced results. Whereas under the DAC definition, impact could in principle be evaluated solely on the basis of the factual, according to the 3ie Glossary, the determination of impact requires explicit comparison with a counterfactual, however this is constructed.

The two approaches towards impact evaluation are not mutually exclusive, but overlap at certain points. Thus an approach using a statistical counterfactual could be used during project implementation, immediately at its end (at Terminal Evaluation stage) and/or some years later. The DAC definition could also be applied at different stages, since a "long-term effect" might be generated at any time. Furthermore, it neither specifies nor rules out the use of a counterfactual-based approach, whether statistically or otherwise pursued.

Most UN Agencies adopt the DAC definition of impact and apply it to impact evaluation, with some adaptations to account for specifics of their key target groups,⁸ including:

- Causal pathways from outputs to impacts, which can be fairly straightforward or more complicated, and effects that become manifest relatively quickly or over longer timeframes;

⁸ Some agency-specific definitions are listed in Annex 1.

- Different levels of analysis: national, institutional, community, household, etc.;
- Different types of intervention that require tailor-made approaches to assess impact (ranging from administrative reform, support to national legislation, to farmer subsidies and humanitarian aid). Given the above, the focus of an impact evaluation can differ widely from one evaluation to another; correspondingly, there may be substantial variation in the mix of methods applied in the evaluation through which the ‘why’ and ‘how’ of an intervention can be explored, and that also may capture the form and extent of indirect and secondary effects.

Role of impact evaluation

Impact evaluation is ideally embedded within broader monitoring and evaluation systems. Together with evaluations based at the outcome and output level, impact evaluations help to demonstrate the effectiveness of an intervention in relation to its objectives; to inform decisions about the continuation (or discontinuation), expansion, or replication of a programme or project; and to contribute to the global evidence base of ‘what works’ and ‘what works for whom in what situations’.

Additionally, impact evaluation enables a better understanding of the process(es) by which impacts are achieved and to identify the factors that promote or hinder their achievement as important feedback into ongoing or future initiatives, including adapting successful interventions to suit new contexts.

Ideally, Impact Evaluation can build upon a substantial base of existing information, to consider the specific issues it can best address. The “key questions”,⁹ to which impact evaluation may provide invaluable (and perhaps unique) answers include the following:

- Did the intervention make a difference?
- What specific contribution did the project make? (Alternatively couched as “What specific part of this difference can be attributed to the project?”)
- How was the difference made?
- Can the intervention be expected to produce similar results elsewhere?

These questions cover a broad range of issues from accountability (particularly value for money) to lesson learning (for replication and scaling up of the effects of the intervention).

Accountability issues may encourage a focus on the first two questions and on specifying cause and effect, rather than on explaining how and why change came about. Questions concerning how much an intervention contributed are often approached through counterfactual-based statistical methods as at least one of their methodological strands. The

⁹ See, for example, “Broadening the Range of Designs and Methods for Impact Evaluation,” DFID Working Paper No. 38, April 2012, P37.

third and fourth questions are appropriate for detailed examination of processes, mechanisms and contexts. They will best be answered through qualitative methods, to uncover underlying processes and their relationship to such contextual factors as national or institutional culture.

None of these questions can be simply answered and each might be approached through one or more evaluation methods. The emerging consensus in literature on impact evaluation appears to be that most questions can best be answered by “mixed methods”. This might involve a mix of both quantitative and qualitative methods, or a mix of specific approaches within either of the two categories. Furthermore, approaches which “blend” methods, such as quantifying some aspects of qualitative data are also increasingly seen as valuable.

The use of impact evaluations among the UN agencies is varied, and its use is expanding. In 2009, the UNEG Task Force on Impact Evaluation conducted a survey of current impact evaluation practices among UNEG members and obtained responses from 28 member organizations. Of these nine had conducted or were about to conduct specific impact evaluations. Others felt that they have partially addressed impact issues as part of other types of evaluation. The nine organizations were: FAO, GEF, IFAD, ILO, OIOS, UNEP, UNICEF, UNIDO and WFP. Since 2009, the number of impact evaluations carried out by these and other UN agencies has increased.

2. Impact Evaluation Design

An impact evaluation design must choose the best means of meeting its objectives, as defined by the key questions it is attempting to answer and by the stakeholders commissioning and conducting the work. It consists of four basic elements:¹⁰

- The evaluation questions
- The theory of cause and effect, which will be accepted as providing sufficient answers to the questions
- Definition of the data necessary to examine the theory
- Framework for analyzing the data to provide adequate explanation of performance against the theory.

A given set of evaluation questions could be answered by a range of evaluation designs. Which design is chosen as best depends on a number of factors, including the context of the evaluation, preferences and persuasions of the commissioning institution and of the evaluators (e.g. in terms of experimental or theory-based approaches), available time, resources and budget. Within a broad design type, (e.g. Theory Based Evaluation) a variety of methods may be used (e.g. document review, case studies, and surveys). Some methods may be components of many or most designs. Thus, a Theory of Change will be an essential part of a Theory Based Evaluation, but may also be found in a design focused on Randomized Controlled Trials. All designs are likely to commence with documentary review.

For impact evaluation to be useful, it is important to adopt methods and approaches that can indicate why a given approach did or did not result in impact, along with implications of this for future directions. For example, an intervention may not have resulted in impact because there were flaws in its underlying assumptions, often referred to as “theory failure,” that will always prevent it from achieving the intended effects. In other cases the logic of the normative work made sense, but lack of impact was due to poor implementation, weak awareness raising or lack of funds, leading to overall “implementation failure”. Clearly, responses to theory or implementation failure should differ. Impact evaluation will be most useful when it can identify factors contributing to successful implementation at the institutional and other levels and the likelihood of sustained benefits to people, as well as at what stages blockages emerge and what can be done to overcome these.

A fundamental characteristic of Impact Evaluation, as indicated by the basic design elements, is its focus on “cause and effect” and on assessing to what extent results can be attributed to the intervention, and what role was played by other factors. There are different types of causal

¹⁰ For a more detailed discussion of design issues see DFID 2012, Chapter 3.

relation, which will require different nuances of impact evaluation design and methods to address. This is illustrated in Table 1 below.¹¹

Table 1: Types of cause-effect relationship in different intervention types

Cause – Effect Relationship	Example of Intervention Type
One cause (the intervention) associated with one outcome	A livelihood programme targeting early reduction of income poverty
One cause (The intervention) associated with multiple outcomes	A road infrastructure programme, which aims to improve travel and transport, commerce and access to basic services
Multiple causes (from one or more interventions) associated with multiple outcomes	A “deepening democracy” programme, which combines support for election processes with training members of parliament and encouraging a culture of political accountability; in order to improve governance, policy making and distribution of national services and benefits
Multiple causes (or interventions) associated with one main outcome	Improving maternal health through one or more interventions to improve neonatal services, health education, and midwife training; and targeting of low income families for health and nutrition assistance

2.1 Range of Design Approaches

The recent rich spate of discussion of Impact Evaluation has produced substantial agreement on the overall range of impact evaluation designs and methods available, but authors have categorized them somewhat differently, depending on their particular perspectives. A recent DFID Working Paper¹² provides the following (Table 2) useful overview of how the main repertoire of design approaches can be used to address the four key questions, which impact evaluation is expected to help answer.

It can be seen from the Table 2.¹³ that there is a substantial range of design approaches available under the broad category of impact evaluation. Furthermore, these design approaches can be combined to ensure that their respective strengths can be used to build up a comprehensive picture of such issues as what has happened, how and why? If we consider the four basic evaluation questions (and the assumptions which underlie them) we can see the match between questions and designs. Once the evaluation design or designs have been selected to answer the key questions of the impact evaluation, the methods necessary to deliver

¹¹ Source: DFID 2012, Table 3.2, P20.

¹² DFID 2012, P24.

¹³ Source: DFID 2012, P48.

according to each design can be selected. This process can be implemented through the use of a detailed evaluation matrix, which relates the specific questions of the impact evaluation to the designs and methods necessary to answer them to the satisfaction of those commissioning the study. This exercise also enables an assessment to be made of the extent to which the design and methods need to be tailored to the available resources and of how best to retain the validity and breadth of findings in the “real world” in which the evaluation must be conducted.¹⁴

Table 2: Impact evaluation designs for key questions

Key Evaluation Question	Related Evaluation Question	Underlying assumptions	Requirements	Suitable designs
To what extent can a specific (net) impact be attributed to the intervention?	What is the extent of the perceived impact? What are other causal or mitigating factors? How much of the impact can be attributed to the intervention? What would have happened without the intervention?	Expected outcomes and the intervention itself clearly understood and specifiable. Likelihood of primary cause and primary effect. Interest in particular intervention rather than generalization.	Can manipulate interventions. Sufficient numbers (beneficiaries, households etc.) for statistical analysis.	Experiments. Quasi-experiments. Statistical studies. Hybrids with ‘Case’ based and participatory designs.
Has the intervention made a difference?	What causes are necessary or sufficient for the effect? Was the intervention needed to produce the effect? Would these impacts have happened anyhow?	There are several relevant causes that need to be disentangled. Interventions are just one part of a causal package.	Comparable cases where a common set of causes are present and evidence exists as to their potency.	Experiments. Quasi-experiments. Theory based evaluation, e.g. contribution analysis. Case-based designs, e.g. QCA.

¹⁴ See Bamberger, M. and Rao, V. and Woolcock, M. Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development. World Bank. 2010.

How has the intervention made a difference?	How and why have the impacts come about? What causal factors have resulted in the observed impacts? Has the intervention resulted in any unintended impacts? For whom has the intervention made a difference?	Interventions interact with other causal factors. It is possible to clearly represent the causal process through which the intervention made a difference – may require ‘theory development’.	Understanding how supporting & contextual factors connect intervention with effects. Theory that allows for the identification of supporting factors -proximate, contextual and historical.	Theory based evaluation especially ‘realist’ variants. Participatory approaches.
Can this be expected to work elsewhere?	Can this ‘pilot’ be transferred elsewhere and scaled up? Is the intervention sustainable? What generalizable lessons have we learned about impact?	What has worked in one place can work somewhere else. Stakeholders will cooperate in joint donor/ beneficiary evaluations.	Generic understanding of contexts e.g. typologies of context. Clusters of causal packages. Innovation diffusion mechanisms.	Participatory approaches. Natural experiments. Synthesis studies.

Given the fact that there is a plethora of design approaches, which have been found to contribute towards sound impact evaluation, it is perhaps surprising that relatively few impact evaluations are undertaken, including within the UN system. Although the number of agencies carrying out impact evaluations has increased in recent years, those that conduct specific impact evaluations are not yet the majority. A number of agencies include impact, either directly or through the criterion of the sustainability of benefits, among the issues to be addressed in their regular evaluations. Budgets spent on specific impact evaluations by UNEG members vary hugely, from \$25,000 to over \$ 220,000. In discussing the opportunities for impact evaluation within the UN system, the current very low level of funding available for impact evaluation needs to be kept in mind, to prevent unrealistic expectations or proposals.

2.2 Theory of Change

There is a growing consensus that a Theory of Change approach provides a sound basis for impact evaluations adopting qualitative or quantitative approaches, or a mix of the two.

A Theory of Change may also be referred to as a program theory, results chain, program logic model, and intervention or attribution logic. In international development evaluation circles,

these terms seem to be used interchangeably. However, academic analysts may draw subtle distinctions among them. In order to avoid extensive discussions of “correct” use of terminology, it is therefore important to define, early in the preparation for an impact evaluation exactly what terms are being used with what meanings.

A Theory of Change is a model that explains how an intervention is expected to lead to intended or observed impacts. The theory of change illustrates, generally in graphical form, the series of assumptions and links underpinning the presumed causal relationships between inputs, outputs, outcomes and impacts at various levels. Many other factors may be incorporated into the model; including “impact drivers”, “assumptions” and “intermediate states” between core steps in the model (e.g., between outputs and outcomes). One effective approach to articulating the theory of change is to work backwards. This involves starting with the desired impact, identifying the various factors that can influence this, and what will need to happen, at various stages, for intervention inputs, outputs and outcomes to be able to contribute to this impact.

Woolcock¹⁵ has emphasized the importance of determining the timeframes and trajectories of impact that we should expect. He notes that, while some projects can be expected to yield high initial impacts, others may take far longer to show results, not because they are ineffective, but because the change they are targeting is inherently long-term in its nature. This needs to be kept in mind with regard to impact evaluation, in order to avoid drawing falsely negative conclusions concerning progress at the time of evaluation.

The process leading to the articulation of the Theory of Change is also important. Sometimes a ToC model is prepared by an evaluator, mainly based upon a review of documentation and perhaps supplemented by some interviews. But there is a danger that this can result in a model for which there is no ownership, and that may not reflect the reality of what is taking place. While of course a review of documentation represents one essential step, very often how an intervention is implemented in practice may vary, sometimes considerably, from how it is described on paper. Stakeholders are more likely to be aware of this, as well as of important nuances, than an evaluator with limited involvement in the content area.

Thus, to the extent possible, a participatory approach should be followed to articulate the ToC, with the role of the evaluator primarily as a facilitator of the process. A group process can help create a shared perspective regarding the nature of the intervention and how it is expected to lead to impact, including identification of various intermediate steps, the roles of other actors, and other factors that may have to be in place. At a minimum, key personnel within the UN agency should be involved in the process, preferably including people who can bring in different perspectives. Other UN agencies that have a role to play in the development and/or implementation of the initiative should also be involved. And as suggested later, other partners, who need to play a role in implementation, including Government bodies, NGOs, and

¹⁵ Woolcock, M. *Towards a Plurality of Methods in Project Evaluation: A contextualised Approach to Understanding Impact Trajectories and Efficacy*. Working Paper 73, University of Manchester: Brooks World Poverty Institute.

other international organizations, should also be given an opportunity to be involved in some way.

At the same time, consensus should not be forced. If there are differing views about potential outcomes and the needed pathways and intermediate steps to achieve these, consideration of these alternative views or assumptions may represent a potential focus for the evaluation, where the validity of these competing assumptions can be examined empirically. Indeed, sometimes it can be useful to develop alternative theory of change models, one representing a presumed pathway to success, and the other where different impacts, including possible negative effects, may result.

One of the benefits arising from articulating the theory of change, in particular if a participatory approach is taken is that this can help surface implicit assumptions and beliefs. Frequently these implicit views are not thought through or shared even with close colleagues. This can result in individuals and programs operating upon differing assumptions without realizing this, often leading to working at cross purposes and/or with basic considerations such as gender equality being forgotten.

Box 1: Outline Theory of Change for UNDESA Statistical Work

- The Statistics Division (SD) of the Department of Economic and Social Affairs (DESA) provides technical analysis on various statistical issues where norms need to be developed or elaborated.
- Member States take notice of this analysis in their deliberations at the intergovernmental level.
- Member States are influenced positively by this analysis.
- Member States then agree on the basis of this analysis to elaborate or agree to some norms and promulgate these norms as declarations, conventions or resolutions.
- National authorities become aware of these norms.
- National authorities incorporate these norms in their national planning efforts.
- These norms are used by national authorities in their national planning efforts.
- The use of these norms at the national level leads to better identification of target population X with a given development need.
- National authorities are able to better use their limited resources making use of this norm.
- X number of citizens in a Member State benefit because of the use of this norm, (positive and intended impact), or
- The use of this norm by a Member State leads to confusion as the old norm was too well established and the civil servants of the given Member States were not convinced of the utility of the new norm.

A useful strategy for articulating the theory of change is to encourage stakeholders to map out the necessary steps between the initial output and the eventual impact, wherever and whenever this is expected to arise. This process can involve getting into considerable detail about the expected causal pathway. Box 1 (on page 16), prepared by the UN Secretariat, is an example of a bare-bone, stripped-down illustration of intermediate steps by which the statistics work of the Department of Economic and Social Affairs (DESA) is expected, ultimately, to lead to benefits for citizens in Member States (but may not do so).

Most of the steps listed in this example refer to changes at the institutional level. Such changes are a key aspect of many UN-supported activities (particularly in normative work), since they are often necessary for successful implementation of improved policies and/or for effective service delivery. When evaluating institutional change, it is important to consider multiple and sometimes simultaneous casual pathways. For example, advocacy may involve direct engagement at senior levels of government, developing support throughout the administration and community mobilization.

While the theory of change represents an invaluable tool for articulating the various steps involved in bringing about change at the institutional level, its focus should not be limited just to this level. It should also indicate the expected pathways whereby changes at this level are expected to lead to ultimate long-term and down-stream impacts, for example on peoples' livelihoods. The exercise should provide a frame of reference for evaluating the relevance of pursued actions and changes at the institutional level, even though it may not always be possible to fully assess changes at more down-stream levels. Involving partners in this process should also help to unpack in greater detail the links between a broader set of actions, or inputs, and changes throughout the causality chain. The theory of change should identify these various pathways, and how they are expected to interact with one another, as well as with other factors, including supporting or opposing actions by other actors.

2.3 Evaluability Assessment in Impact Evaluation Planning

Typically, an evaluability assessment includes several steps and has a number of outputs. Among these, the evaluability assessment will include the mapping, systematization and analysis of any baseline and/or monitoring data that were produced by the managers of the intervention/body of work to be evaluated; these data will be important to inform the development of the impact evaluation tools. The main output of the evaluability assessment should be a full approach paper, including an evaluation matrix, that sets out in a detailed and explicit manner the analytical and methodological approach of the evaluation.

The development of the theory of change is a key part of the evaluability assessment. A ToC is particularly useful in identifying potential evaluation questions and in helping to determine what it is realistic or possible to assess at given points of time in the programme cycle and with defined resources. In particular, the theory of change should specify how far along the results chain it can be realistic to expect changes attributable to the intervention to have occurred at any given point in time and this can aid in identifying how best to focus the evaluation. Development of the Theory of Change should therefore be a major part of the

evaluability assessment, discussed later, which forms the preparatory phase of all complex evaluations. For the impact evaluation of very large or complex interventions, the evaluability assessment may be a study in itself. More often, it is undertaken by the evaluation office as part of its preparation for the impact evaluation and to facilitate development of its detailed Terms of Reference.

By identifying what is possible to evaluate at a given point in time, highlighting those evaluation questions that are most critical, and specifying assumptions in the programme logic most in need of empirical verification, an evaluability assessment can identify priorities for impact evaluation. Even when it may be premature to assess long-term impact specifically, an evaluability assessment should identify how progress towards impact can be assessed, and those assumptions in the theory of change that are most in need of objective verification.

2.4 Gender Equality and Human Rights

Gender equality and human rights (GE and HR) are both substantive areas of normative work and crosscutting issues, which should be mainstreamed in all UN initiatives and that should be assessed in all UN evaluations, including impact evaluations. The UNEG Handbook “Integrating Human Rights and Gender Equality in Evaluation - Towards UNEG Guidance” notes that “All UN interventions have a mandate to address Human Rights and Gender Equality issues”.

The Handbook identifies the following principles for integrating human rights and gender equality in evaluation:

- Inclusion
- Participation
- Fair power relations
- Mixed evaluation methods

These principles, which largely correspond to good evaluation practice, are translated into various aspects of the evaluation process. Examples include the conduct of an evaluation stakeholder analysis from a HR and GE perspective, the development of evaluation criteria and questions that specifically address HR and GE, the collection of disaggregated data, but also the recruitment of an evaluation team with knowledge of and commitment to HR and GE.

This may prove challenging in some situations. For example, basic data that an evaluation should ideally draw upon may not have been disaggregated – or even exist in any form. This may require additional data collection through specific methods; such as, for example, through surveys and analysis of existing documentation (e.g. both informal and formal records of meetings) that talk about gender and human rights differences. A variety of qualitative techniques, including community meetings, focus groups, key informant interviews and Most Significant Change reports, can also be used to obtain retrospective data.

Below are examples of questions, which may help address HR/ GE principles in impact evaluations:

To what extent has the UN agency incorporated HR/GE principles in inter-agency work: e.g. the development of institutional monitoring and reporting mechanisms for workers' or children's rights?

To what extent have governments and other institutional partners incorporated and applied HR/GE principles in their implementation of normative work?

A theory of change may be explicit in the original intervention design, but often is not. For example, proposals for change may assume that increasing women's income-generating capacity will lead to empowerment – which may or may not be true. Or that laws ensuring human rights (in a constitution, for example) are sufficient to guarantee their fulfilment. More frequently, proposals for change focus on one dimension (for example; economic, skills training, infrastructure); which is necessary but not sufficient, while ignoring other key factors (e.g. access to markets, self-confidence or other social and cultural phenomena). A very important role of evaluations is to draw attention to implicit theories of change and their strengths and weaknesses. Often human rights and gender equality are absent in a theory of change, or expressed in a way that does not lead to concomitant action. For example, projects or programmes might note that woman-headed households are poorer than others, but include no activities designed to address this inequality. Alternatively, a programme of land reform that pays attention to gender equality might not only enact rights to land, but may also ensure that the registration system includes a category for joint ownership, identifies the gender of the owner, communicates and promotes women's rights to land ownership and the advantages of joint registration, and provides disaggregated information about changes in the ownership of land by gender.

3. Common Methods in Impact Evaluation

It has been shown above that there is a range of impact evaluation designs. There is also a range of methods that can be used within these designs. Methods are flexible and can be used in different combinations within impact evaluation designs to answer the specified evaluation questions.

3.1 Quantitative Methods

Experimental and quasi-experimental quantitative designs are appropriate for questions concerning whether an intervention has made a difference and the extent to which a specific impact can be attributed to an intervention. Leeuw and Vaessen¹⁶ have noted that methods suited to such designs are particularly appropriate for impact evaluations of “single-strand

¹⁶ NONIE. *Impact Evaluations and Development. Nonie Guidance on Impact Evaluation.* (Leeuw, F. and J. Vaessen, J.) 2009.

initiatives with explicit objectives — for example, the change in crop yield after introduction of a new technology, or reduction in malaria prevalence after the introduction of bed nets. Such interventions can be isolated, manipulated, and measured, and experimental and quasi-experimental designs may be appropriate for assessing causal relationships between these single-strand initiatives and their effects”. Further, as White and Phillips¹⁷ have indicated, these methods are most suited for evaluations with both “large N” and “large n”. Both the overall population affected and the sample groups must be large.

These quantitative methods use sophisticated statistical procedures to address three basic issues, namely:¹⁸

- “The establishment of a counterfactual: What would have happened in the absence of the intervention(s)?
- The elimination of selection effects, which might lead to differences between the intervention group (or treatment group) and the control group
- A solution for the problem of unobservables: The omission of one or more unobserved variables, leading to biased estimates”.

Statistical methods used in experimental and quasi-experimental designs include: Randomized Controlled Trials (RCTs), pipeline approaches, propensity score matching, judgemental matching, double difference, regression analysis, instrumental variables and regression discontinuity analysis. Increasingly, impact evaluations using these methods also incorporate at an early stage preparing an overview of the intervention based on the construction of a Theory of Change (see 5.2.1 below).

The intervention characteristics (“single-strand initiatives with explicit objectives”), which promote use of these methods, are not the normal business of many UN organizations. Furthermore, the requirement for “large n” studies using complex statistics calls for substantial finance (and evaluation management resources), which is rarely available. An IETF study¹⁹ showed that (as at 2010) only 3 UNEG member bodies had commissioned Impact Evaluations using experimental or quasi-experimental methods. With this background in mind, it is important to focus guidance for UNEG evaluation practitioners on issues, which actually confront them. This conforms with an important observation made by Bamberger et al.,²⁰ namely: “For many evaluation professionals, particularly those working in developing countries, the debates on the merits and limitations of statistically strong impact evaluation designs are of no more than academic interest as many may never (and are highly unlikely to)

¹⁷ White, H. and Phillips, D. *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework*. 3ie. New Delhi. 2012

¹⁸ NONIE, 2009, P23.

¹⁹ UNEG. *Concept Note: Impact Evaluation among UNEG Members*, Annex 2, Impact Evaluation in UNEG members. New York. 2010.

²⁰ Bamberger, M. and Rao, V. and Woolcock, M. *Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development*. World Bank. 2010.

have an opportunity to apply any of these designs during their whole professional career”. This Guidance Note does not therefore examine these methods further, but refers readers to the NONIE Guidance (Ps. 23-31), which provides a good introduction to them.

In addition to the more complex statistical approaches described above, basic quantitative analysis of existing databases and/or survey analysis can make an important contribution to developing the overall “story” of the difference made by an intervention. This analysis may draw on descriptive statistics, such as cross-tabulations; and/or inferential statistics, such as analysis of variance to compare the means of several different groups.

3.2 Qualitative Methods

Many types of intervention are not appropriate for complex quantitative approaches, such as experimental or quasi-experimental methods. These would include: “programs with an extensive range and scope that have activities that cut across sectors, themes, and geographic areas. These can be complicated — multiple agencies, multiple simultaneous causes for the outcomes and causal mechanisms differing across contexts and complex (recursive, with feedback loops, and with emergent outcomes)”.²¹ Much of the work of UN bodies is in complicated and/or complex situations,²² an aspect that needs to be built into impact evaluation designs. Sometimes, it may be possible to break down such interventions into simpler components, which lend themselves to quantitative analysis. However, for a great many UN interventions, quantitative methods will answer only part of the questions related to impact. This will place a premium on evaluation designs, which are centered on qualitative methods.

There is a range of qualitative methods, which have been found useful in Impact Evaluation, including: Realist Evaluation, General Elimination Methodology,²³ Process Tracing,²⁴ and Contribution Analysis. Information on their characteristics and potential scope and use, can be readily found in guidance documents, (see Annex 1, List of Works Cited).

3.3 Participatory Methods to Establish Stakeholder Perceptions

Various participatory methods,²⁵ including: Most Significant Change Method, Success Case Method, Outcome Mapping, and Method for Impact Assessment of Programmes and Projects (MAPP) do not focus on explicit attribution of cause and effect, although they may contribute to an understanding of this. Rather, they attempt to establish the factors that have contributed towards change by talking directly to stakeholders, using “semi-structured” approaches, rather than “structured” survey instruments. These methods are primarily qualitative, although small

²¹ NONIE, 2009.

²² See, Rogers, 2009

²³ See also White and Phillips, (P9, 10, 38,39)

²⁴ See, for example, White and Phillips (P10, 11, 40, 41)

²⁵ See NONIE Guidance, Annex 9.

scale quantitative approaches (e.g. surveys) may be used, as well as methods (such as Qualitative Comparative Analysis²⁶), which quantify key elements emerging from qualitative methods.

3.4 Methods and Validity

The various methods examined above have different comparative advantages and the contribution, which each can make, can be organized under four different types of validity:²⁷

- Internal validity: establishes that the causal relationships verified or measured by the evaluation correctly describe the links between outputs, outcomes and impacts.
- Construct validity: establishes that the variables selected for measurement appropriately represent the underlying processes of change.
- External validity: establishes the extent to which the findings from one evaluation can be generalized to inform similar activities.
- Statistical conclusion validity: for quantitative approaches, establishes the degree of confidence about the relationship between the impact variables and the magnitude of change.

In terms of the major impact evaluation approaches discussed, Randomized Control Trials are regarded as particularly strong on internal validity, since they eliminate other factors, which might affect the identified causal linkages. However, they have relatively low external validity, since their resources are mainly focused on ensuring an accurate counterfactual to the specific intervention under examination and they do not consider similar interventions elsewhere.²⁸

Qualitative and participatory methods, on the other hand, focus on the details, complexity and meanings of change and may therefore be highly effective in terms of construct validity. However, since the findings of these methods are also context specific, they are just as prone as quantitative methods to low external validity. In many cases, the most effective way of boosting external validity may be through some form of synthetic review of existing evidence from other evaluative sources.

In order to ensure a certain degree of internal (external, construct) validity of findings, the Guidance advocates for a mix of methods, “triangulating” the findings of different methods by

²⁶ See DFID 2012, Appendix and White and Phillips, (P56-60).

²⁷ As discussed in Chapter Five of the NONIE Guidance.

²⁸ RCTs control for observable and unobservable characteristics in the sample which may affect outcomes. Without further knowledge on the relationships between these characteristics, the distribution of these characteristics in the sample, and the distribution of these characteristics in other populations, the findings of the analysis cannot be generalized to these other populations.

comparing them with each other. For most development evaluations, the phenomena to be evaluated are sufficiently complex to make a mixed method approach essential.

3.5 Choosing the Mix of Methods for Impact Evaluation

When undertaking impact evaluation, even more so than in other types of evaluation, it is important to do more than list a set of methods and assume that this amounts to a methodology. Rather, within the chosen evaluation design, there should be an explicit overarching methodological framework, which enables the individual methods to be brought together to produce a meaningful overall analysis that can evaluate whether the intervention has had impacts, or made a contribution towards them. It is essential to tailor the particular evaluation design and mix of methods, to the particular situation, context, questions and issues of most concern.

For these and related reasons, this Guidance Note has emphasized the importance of starting with the Theory of Change and of using this as a basis for identifying and prioritizing questions for impact evaluation that can be appropriately explored at the chosen point in time. The discussion above has identified various design options (e.g., experimental, case-based) and principles that should be taken into consideration when choosing an appropriate mix of methods.

In each case, it is essential to consider the strengths and limitations of potential methods and their best fit with the requirements of the evaluation being designed. Given that all methods have strengths and limitations, it is important to use a mix of different methods, both quantitative and qualitative, to provide for triangulation and to balance off the limitations of any single method.

For example, policy change, which is a frequently intended institutional impact, is influenced by many factors. Responses to a questionnaire (as an example of a possible method to include in impact evaluation) may or may not acknowledge the influence of any UN intervention on the development of a new policy. Politicians and officials may sometimes be reluctant to acknowledge that others have influenced their work. Furthermore, government policies are invariably influenced by multiple factors and the role of external stakeholders is often not explicitly acknowledged, particularly in formal documents.

Nevertheless, it may be possible to relate particular events, such as work on new legislation or policies initiated, to other activities that can be documented, such as representations by UN officials and public advocacy campaigns. This may require significant probing; as well crosschecking with different stakeholders, to be able to identify how the UN has contributed to this. In some situations, it might be possible to develop a time-series, to relate changes over a period of time to when specific actions occurred. Much of the most pertinent documentation (e.g. minutes, memos) may only be identified during open-ended interviews and discussions (if at all). In-depth interviews with a range of stakeholders are among the best means to gather evidence about how actions came about, including perceptual data and to build up a meaningful picture.

The exact choice of methods will vary from evaluation to evaluation and should be identified during the development of the evaluability assessment. This Note stresses the importance of planning adequately for impact evaluation and has emphasized the need to articulate the program logic and identify the intermediate steps and the types of data that are needed at various stages for meaningful impact evaluation.

This does not mean that once the theory of change has been identified, the evaluation design developed, and data collection mechanisms put into place, these should be regarded as definitive or unchangeable. It is useful to periodically consider if the assumptions underlying the approach to impact evaluation are still valid, and if any changes may be needed. For example, at the early stages in the evaluation, findings may raise some questions about certain assumptions in the theory of change. It might be appropriate then to make some modifications to the theory of change, and perhaps also to the evaluation approach as applicable.

In particular, it is important to bear in mind the iterative, circular nature of complex undertakings, such as those often supported by the UN. Circularity does not mean going around in circles or being confused. In the context of systems thinking, this refers to the inter-relatedness of the various components of a system, that most frequently are not unidirectional but instead have feedback loops. Interactions often may occur in unpredictable ways. This means that throughout the conduct of the impact evaluation, evaluators and managers of the exercise should maintain an open mind on issues to be pursued and analyzed, including necessary reviews of original assumptions and approaches to the evaluation. Flexibility in methods should be ensured in so far as possible given usual time and resource constraints.

4. Quality Control for Impact Evaluations

Impact evaluation provides answers to questions concerning the ultimate benefits to which an intervention has contributed. It is therefore highly sensitive and must be able to withstand intense scrutiny. These characteristics make effective quality control essential.

4.1 Specific Quality Control Criteria at the Design Stage

In view of the over-riding importance of the initial choice of design and methods for an impact evaluation, there should be additional Quality Control at this stage. This should cover the specific areas shown in Table 3 on the next page.

Table 3: Issues for quality control at IE design stage to assure validity and rigour²⁹

Contribution	Explanation	Effects
Is the design able to identify multiple causal factors?	Does the evaluation make it clear how causal claims will be arrived at?	Are long term effects identified?
Does the design take into account whether causal factors are independent or interdependent?	Is the chosen design able to support explanatory analysis (e.g. answer how and why questions)?	Are these effects related to intermediate effects and implementation trajectories?
Can the design analyze the effects of contingent, adjacent and cross-cutting interventions?	Is theory used to support explanation? (E.g. research-based theory, Theory of Change), if so how has theory been derived?	Is the question 'impact for whom' addressed in the design?
Are issues of 'necessity', 'sufficiency' and 'probability' discussed?	Are alternative explanations considered and systematically eliminated?	

4.2 Quality Control³⁰ Requirements and Approaches for Impact Evaluation

To a large extent, experimental and quasi-experimental methods have established techniques to determine such aspects as validity, reliability and bias.³¹ These techniques are largely impenetrable to all but specialist statisticians or academics with a strong background in quantitative methods. In this respect, quality of impact evaluations with high statistical content can best be assured through recruitment of a specialist adviser or panel, to meet at key stages to review the conduct and, later, outputs of the work.

Although there are many quality requirements, these are often basic evaluation practice, rather than specific to impact evaluation. However, given the potential importance of impact evaluation in assessing the contribution of UN bodies to their long-term change goals, quality requirements acquire an even higher profile in such exercises. Quality questions begin at the design stage of the impact evaluation and follow through to completion. They can be broken down into specific quality questions, covering the various standards shown in Tables 4 and 5 on the next pages.

²⁹ Source: DFID 2012, P75

³⁰ Often referred to as Quality Assurance. In this paper, the two terms are used interchangeably.

³¹ See NONIE Guidelines, Section 4.2 for a discussion of these issues.

Table 4: Quality control criteria for overall technical implementation of impact evaluation

<p><i>Choice of designs and methods</i></p>	<ul style="list-style-type: none"> • Are designs and associated methods put forward established, well documented and able to be defended? • Do the chosen designs take into account the Evaluation Questions and intervention attributes?
<p>Reliability</p>	<ul style="list-style-type: none"> • Are they able to explain how an intervention contributes to intended effects for final beneficiaries? • Do the Evaluation Questions allow for success and failure (positive and negative effects) to be distinguished?
<p><i>Proper application of designs and method</i></p> <p>Robustness</p>	<ul style="list-style-type: none"> • Are the ways that designs and methods are applied clearly described and documented? • Does the application of designs and methods and subsequent analysis follow any protocols or good practice guidelines? • Is the evaluation team knowledgeable about the methods used?
<p><i>Drawing legitimate conclusions</i></p> <p>Transparency</p>	<ul style="list-style-type: none"> • Do conclusions clearly follow from the findings? • Has the evaluation explained the effects of the programme? • How are evaluative judgements justified? • Have stakeholder judgements been taken into account when reaching conclusions? • Are the limitations of the evaluation and its conclusions described?

Source: DFID 2012, P74.

As shown in Table 4 above, there are a number of key quality questions, which cover the full life-cycle of the impact evaluation: from its initial choice of design and methods, through the application of these during the work, to the final process of drawing conclusions in a legitimate manner. Quality Control should ensure the reliability, robustness and transparency of the evaluation methods used throughout the process.

4.3 Quality Control of Evaluation Standards

An additional set of Quality Control criteria should be in place to ensure that the evaluation standards appropriate to the UN system are maintained throughout the study. Although these are to a large extent covered by the UNEG Norms and Standards, they can be clearly codified

for verification and control in the circumstances of impact evaluation as shown in Table 5 below.³²

Table 5: Quality control issues for evaluation norms and standards

Country-based criteria	Ethical criteria	Institutional criteria
Have country-based stakeholders (government and civil society) been actively involved and consulted in formulating evaluation questions?	Have the evaluators made explicit their interests and values as they relate to this intervention?	Are there any joint or partnership arrangements in place – joint evaluations, consortia involving local partners?
Have country based administration and information systems been used as far as possible?	Have arrangements been put in place to monitor and remedy bias or lack of impartiality?	In what ways has the evaluation contributed to evaluation capacity building in-country?
Has the evaluation been set into the country context and other country based evaluation taken into account?	Have confidentiality and risks to informants been taken into account?	What has the evaluation done to feed into policy making both among donors and in partner countries?
	Have feedback and validation procedures that involve stakeholders been specified and used?	

4.4 Managing a Quality Control System

Effective management of a Quality Control System for an Impact Evaluation requires substantial inputs from several sets of stakeholders: the evaluation office itself, the impact evaluation team (usually external consultants) and any internal or external Quality Assurance (QA) advisers recruited to act individually or as a Panel. It is therefore important to ensure that any such system designed can achieve the best balance between degree of QA offered and the available resources. As noted earlier, the budgets actually available for impact evaluation within the UN system have historically mainly clustered around the \$25,000 to \$30,000 mark. Unless this figure increases substantially, the only option to ensure adequate control of quality will be for Evaluation Office staff to undertake this role themselves (assuming that their time is not included in the initial budget figure). However, this will only work if they have sufficient specialist knowledge of the key issues of impact evaluation design and implementation.

In situations of scarce resources, the most important contribution to be made in terms of QC is likely to be through the original development of an impact evaluation concept paper and methodology. This will be done by specialists in charge of the exercise (if it is a self-evaluation) or by the evaluation office (if it is an independent one). The paper should include:

³² Source DFID 2012, P76

- The nature of the evaluation (key questions);
- A draft theory of change (or a change model);
- A description of the methods to be used and how they link to the key evaluation questions;
- Whether (and if so how) to treat counterfactuals and/or how to use such methods as causal contribution analysis;
- Which partners should be associated with the impact evaluation and in what roles;
- How impact evaluation results will be used;
- The Quality Assurance/Control plan for the impact evaluation.

Such a well-developed Concept Paper provides a firm basis for: drawing up Terms of Reference, if consultants are to be used; informing and assessing proposals in the event of competitive bidding; and briefing consultants during the inception phase of the work. Thereafter, regular quality control measures can be put in place to ensure that the evaluation team work according to the concept paper.

Furthermore, the Tables 3 to 5 above can be used to create QC Checklists. Depending on resources, these can form the basis of periodic meetings between the impact evaluation implementation team, the Evaluation Office and the QC panel or Adviser. If resources are not available for such regular meetings, the issues highlighted in the Tables can be used to create tailored evaluation-specific check lists, which can be scored (even if only as “Yes” or “No”) on the basis of documents produced during the course of the evaluation, supported by (occasional) in-person and regular virtual meetings between the EO and the evaluation team members (together with any Quality Adviser(s)).

Important contributions can also be made through the process of circulating all impact evaluation documents at draft stage to key stakeholders for review, comments and feedback. Such stakeholders can include the UN Country Team, representatives of Government bodies and other institutions affected by the evaluation and (rarely) of bodies representing direct beneficiaries (where there are such). Feedback can also be provided verbally, in writing or through meetings of knowledgeable and experienced peers, particularly at headquarters level, where EOs are often located.

The circulation process is particularly important at the stage of the Draft Final Report, which needs detailed review by all key stakeholders. Appropriate comments should be integrated into the report by the evaluation team, in a transparent manner, with a clear audit trail.

Although they can play an important role, such QC procedures may also have some negative consequences. For example, different viewpoints or emphases between the evaluation team and external quality advisers may impose time delays and use up valuable human resources. A

system for negotiating agreement or (as necessary) imposing decisions needs to be agreed in advance by all parties.

5. Impact Evaluation of Normative Work³³

UN Normative work has been defined by UNEG as follows:

The support to the development of norms and standards in conventions, declarations, resolutions, regulatory frameworks, agreements, guidelines, codes of practice and other standard setting instruments, at global, regional and national level.

Normative work also includes the support to the implementation of these instruments at the policy level, i.e. their integration into legislation, policies and development plans, and to their implementation at the programme level.

Normative work has the potential to result in high level impacts that can affect the lives of millions of people. Broadly speaking, existing evaluation approaches and methods can be applied to the evaluation of normative work. Nevertheless, the nature of this type of work does differ in some significant ways from that of initiatives, including projects and programmes that have a direct impact at the community, household or individual level.

Impact evaluation of normative work refers to identifying the lasting and significant changes of this work at all levels in the results chain, including at its end, e.g. on people's livelihoods, their empowerment, increased biodiversity and healthier ecosystems. However, this impact comes to fruition through a variety of intermediate steps and a complex results chain, in which institutions are often the intended primary initial beneficiary of normative work ("institutional impact") and play a major role in influencing the down-stream effects of the NW itself. Institutions are indeed the first and direct focus of impact of NW and this level can have considerable intrinsic value in itself, such as when government policies, practices, or organizational cultures are changed in response to the NW itself.

Given the long-term trajectory of much normative work and the indirect way in which impacts further down the results chain may come about, it is not always possible for an evaluation to identify all longer-term impacts, depending upon its timing and resources. Nevertheless, it is widely recognized that impact evaluation of normative work, to the extent possible, should go beyond establishing institutional impacts and that it is appropriate to combine assessment of institutional impact with identification of subsequent resulting changes. Assessment of the latter would also allow identifying the validity, merit or worth of the normative work itself. For example, adoption of standards with respect to food quality by itself does not guarantee that these standards are indeed appropriate and lead to the intended effects (e.g. do they really result in safer food as obtained and consumed by individuals and families). Equally, the

³³ This section draws extensively on the draft paper prepared for the impact evaluation TF by Mr. B. Perrin. More detailed guidance on Evaluation of Normative Work is to be found in the UNEG Guidance Note on Evaluation of Normative Work.

standards might result in unintended negative effects that undermine the overall positive impact. For example, labour protection standards lead to improved safety for those covered, but may lead to some employers and workers moving instead into the informal economy with lesser protection.

The assessment of attribution means demonstrating that the identified impact resulted in some way from what was done (i.e. the normative work); and would not have happened otherwise (i.e. the counterfactual). When evaluating the impacts of NW, this becomes increasingly indirect and difficult further down the results chain, due to the increasing number and types of factors influencing impact. In this context, the most feasible, and important approach is identifying what actually took place and indicating how the normative work (along with other actions) influenced or contributed to the observed change.

In this context, models and approaches that can acknowledge and reflect complicated and complex causal relationships are likely to be more appropriate than those that are more applicable to simple linear relationships. Data on activities, outputs and intermediate outcomes, irrespective of who is directly responsible for them, are essential components of the impact evaluation of normative work. It is also essential to test out and document the assumed causal links and relationships, at all levels, leading up to impacts.

6. Impact Evaluation of Multi-Agency Interventions³⁴

The defining feature of a multi-agency intervention is that multiple actors, such as UN agencies and other international bodies, national public institutions and civil society organizations, contribute to the execution of activities towards a common overarching goal. The collaborating agencies share responsibility for the intervention's overall achievements and shortcomings. This is different to the situation where one agency is responsible for implementation and delegates responsibility for execution to another agency.

UN initiatives are increasingly moving away from discrete projects implemented by a single agency to comprehensive programmes at the country, regional and global levels which bring together expertise from different organizations. There are also multi-agency initiatives such as the Delivering as One initiative and the UN Development Assistance Frameworks (UNDAF), around which countries are increasingly arranging their programmes and joint programmes addressing a wide range of different issues and priorities. These often include a variety of initiatives that cut across sectors, themes and geographic areas. The Delivery as One initiative, launched in 2007 is an important component of the United Nations' response to the challenges of a changing world. The initiative seeks to enhance the coherence, efficiency and effectiveness of the UN at a country level and to reduce transaction costs for host countries.

³⁴This section draws extensively on the draft paper prepared for the IETF, by Prof. P. Rogers, P. and Mr. D. Fraser: Impact Evaluation of Multi-Agency Interventions. Guidance Note on Planning, Managing, and Utilizing Impact Evaluations of Multi-Agency Interventions. UNEG IETF Revised Draft.

6.1 Types of Multi-Agency Interventions

Multi-agency interventions are expected to achieve results that are more than the sum of their parts, thanks to the synergistic effects of their programme components. Thus, an important aspect of multi-agency impact evaluations should be a focus on assessing the additional value from agencies working together.

Multi-agency interventions differ in the ways they work together and the degree of ‘jointness’. These have important implications for impact evaluation. If agencies are already working closely together, it is likely to be both more useful to undertake a joint evaluation of their work and easier in terms of already having worked on common definitions of terms and data systems.³⁵

When the multi-agency intervention is in the form of a joint project or programme there are often good reasons to conduct a joint impact evaluation including:^{36,37}

- To enhance joint work by increasing understanding of priorities, the shared and separate issues involved in an intervention’s effectiveness and appropriateness, and by aligning recommendations;
- To reduce the evaluation burden on partner governments and aid recipients from multiple, separate evaluations;
- To improve the impact evaluations, by bringing together wider expertise, sharing information about evaluation methods and processes, enabling cost-sharing and broadening ownership of findings.

However, it should not be assumed that a joint impact evaluation will always be the best option. If agencies have very different information needs and processes, the evaluation risks incurring significant transaction costs without producing additional useful information for either agency.³⁸ Another option is to have a smaller joint impact evaluation, which addresses specific issues of joint concern, supplemented by separate evaluations undertaken by the different agencies. Figure 1 on the next page illustrates these options.

³⁵ More detailed guidance on Joint Evaluations is to be found in the UNEG Resource Pack of Joint Evaluations.

³⁶ Feinstein, O. and Ingram, G. (2003). *Lessons Learned from World Bank Experiences in Joint Evaluation*. DAC Working Party on Evaluation. Available at <http://www.oecd.org/dataoecd/secure/15/13/31736431.pdf>

³⁷ Binnendijk, A (2000) *Effective Practices in Conducting a Multi-Donor Evaluation*. Paris: OECD/DAC.

³⁸ Toulemonde, J., Fontaine, C., Laudren E. and Vincke, P. (1999). ‘Evaluation in Partnership. Practical Suggestions for Improving their Quality’. *Evaluation*, 4 (2): 171-188.

Figure 1: Options for impact evaluations of multi-agency interventions

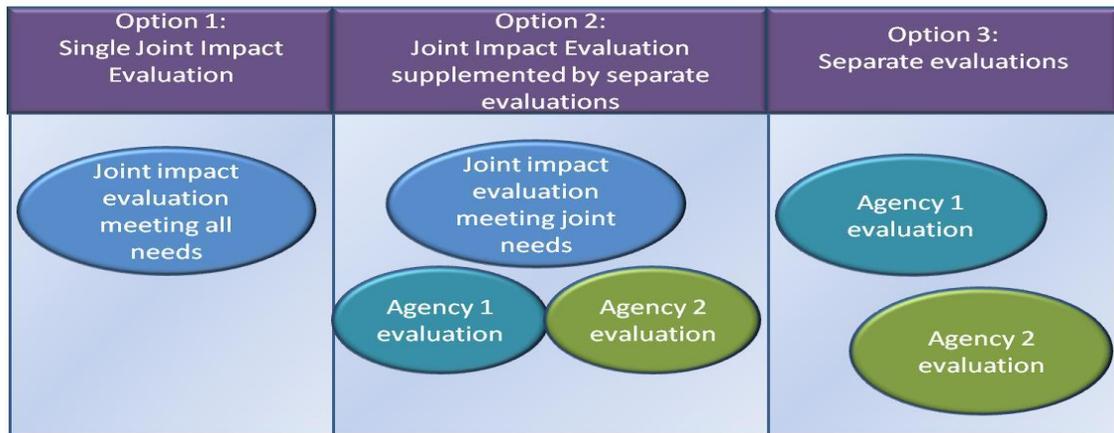


Table 6 shows six types of Multi-Agency Interventions which describe different levels of ‘jointness’ and different ways that agencies might work together. These are not intended as specific examples, but rather as starting points to allow teams designing or commissioning impact evaluations of multi-agency interventions to develop their own understanding of the intervention they will be evaluating.

Table 6: Types of multi-agency interventions and their implications for impact evaluation

Type	Description	Implications for Impact Evaluation
1. Shared front end	Two or more programmes which are planned and delivered separately but which feature a shared entry point and co-location of services for members of the target group (including direct beneficiaries, NGOs and government departments and agencies). While there is some co-ordination between agencies in outreach and reception, the activities are actually quite separate and relate to quite different programmes and intended outcomes and impacts.	While it might be useful to conduct an evaluation of the costs and benefits of co-location, there would not be value in doing a joint impact evaluation of the different programmes.

2. Separate strands	Two or more programmes which contribute to a shared intended impact but which operate separately – for example, school infrastructure and child health, which can each contribute to improving educational outcomes. In these types of multi-agency programmes, the different agencies do not work together to achieve short-term outcomes. The agencies usually have separate funding for their activities. In these cases, the achievements of each separate agency at the lower levels of the results chain can be easily distinguished.	An evaluation of the entire intervention would probably add little of value in terms of improving knowledge of the separate programmes, although it might be useful in terms of providing an overall evaluation of success.
3. Relay	Interventions where the output from one agency becomes an input for another agency – for example, one agency produces plans, which another agency then uses to guide implementation, or one agency builds capacity of agencies, which then use this capacity to implement specific interventions.	An impact evaluation can provide evidence of the overall impact of the agencies’ work and improve their co-ordination.
4. Different sites	A large intervention implemented by different agencies at different sites – such as different local authorities, or different national governments. This can be thought of as a variation of the ‘relay’ type, but with multiple implementing agencies.	This requires a high level of co-ordination to develop a joint impact evaluation, and increases the likelihood that a single evaluation will not meet all the different needs of the different agencies.
5: Horizontal collaboration	<p>While the ‘relay’ type has two or more agencies working ‘vertically’, with results passing from one agency to the next, horizontal collaboration is where agencies are working together at the same level in the causal chain to produce outputs and outcomes.</p> <p>An example from refugee services is where one agency provides basic food, and another provides materials for cooking the food, and obviously these need to be coordinated to be effective.</p>	This highly inter-related intervention is one where agencies are likely to find it particularly useful to undertake a joint impact evaluation and to learn about improving the quality of their co-ordination and partnership.
6: Emergent partners and roles	Where agencies are working together in flexible and adaptive ways. This is more likely to be appropriate for new types of interventions, where the problems or opportunities, which they address, are less	This emergent type of intervention is the most difficult for multi-agency impact evaluation, as the evaluation design might need

	well understood, and where the plan for working together will need to be developed as it is implemented. As this happens, the agencies involved may well change, and their roles may change as well.	to change to accommodate changes in how the intervention is implemented and changes in the partners in the intervention and their needs and expectations for evaluation.
--	--	--

6.2 Impact Evaluation Issues Specific to Multi-Agency Interventions

Multi-agency interventions can present particular challenges for impact evaluations in terms of:

- Effective management – balancing clear management processes and adequate consultation
- Appropriate scope and purpose – negotiating between competing priorities and needs of the different agencies in terms of questions to be answered and timelines for decisions
- Clear theory of change/logic model– articulating how the multiple components of the intervention are understood to work together
- Explicit and defensible evaluative criteria and standards – negotiating “what success looks like”, in terms of which impacts are seen as important and what standards of performance are required
- Feasible data collection and analysis - accommodating differences in data definitions and formats and what are seen as appropriate indicators and measures
- Credible causal inference – meeting different organizations’ needs regarding causal attribution. Partner organizations may have different policies and understandings about what research designs are considered credible and appropriate. For some organizations, only RCTs (Randomized Controlled Trials) can provide a compelling argument about causal attribution; for others a range of research designs can be used. Given the variation in the way terms are used and the very different positions held by different agencies, it is essential that this issue is clearly discussed and that agreement is reached before deciding to proceed with a joint impact evaluation of a multi-agency intervention.

6.3 Agreement on Purpose and Roles in Multi-Agency Impact Evaluations

The intended use of a joint evaluation needs to be identified and addressed carefully during planning and throughout the evaluation, not only when an evaluation has been completed. A multi-agency impact evaluation will likely need to balance agencies' different intended purposes and priorities, so it is even more critical at project design stage to systematically identify who is expected to use the impact evaluation and for what purpose(s).

In multi-agency impact evaluations different agencies might have different criteria for evaluating interventions, based on their overall organization's goals. Alternatively, they might agree on criteria but not on standards. Involving the different agencies in the process of developing shared descriptors or rubric of what success means will identify whether or not it will be possible to develop a shared evaluative judgement.

While most impact evaluations are based on a theory of change, these are particularly useful for multi-agency impact evaluations, especially if they make clear how the different agencies are understood to work together. It is important that the different agencies share an understanding of the intervention and are able to develop a characterization of how the agencies' combined efforts are expected to produce greater benefits than from individual interventions.

Existing documentation may not be sufficiently specific about how the agencies are understood to work together, even if a theory of change has been developed. If the impact evaluation is being planned some time after the program has begun, it is also likely that intended results, roles and responsibilities will have become clearer or have shifted to some degree since the intervention started. Therefore it is likely that a combination of sources will be needed – including existing documentation and articulation of stakeholders' perceptual models.

As with any joint evaluation, in the case of joint impact evaluation it is usually advisable for one of the participating agencies to accept a lead role, particularly in terms of engaging on quality assurance matters with the service provider and in acting as a convener of strategic and important events. The full implications of the decision should be explored with the procurement functions of the agencies, so that there are no negative consequences for the implementing agency further down the line.

Issues to be addressed:

Is an impact evaluation really needed?

Some agencies participating in the intervention may not wish to conduct an impact evaluation while others do. Careful consideration is needed to determine whether an evaluation should go ahead without the participation of all agencies involved in the multi-agency intervention, particularly from a data access perspective.

Is there agreement about the main purpose of the evaluation – or scope to accommodate multiple purposes?

The purpose of an evaluation plays a key role in informing strategic decisions around the approach to be followed, including who will implement the evaluation and the methods to be used. As a result it is important that agencies collaborating in an impact evaluation agree on its purpose. They should be explicit about their intended uses for the evaluation, and to ensure that the evaluation will adequately meet these needs.

How will the key evaluation questions be decided?

In multi-agency impact evaluations it is important to have agreement about the key evaluation questions. This does not mean simply increasing the number of questions to accommodate all the different agencies, as this is likely to produce an unmanageable list for the evaluation to adequately address. Instead a workable compromise should be sought – which may include having supplementary components of the evaluation that are undertaken by different agencies.

How are the different agencies understood to contribute to the intended outcomes and impacts?

It is most useful, but rare, for a logic model of a multi-agency intervention to make explicit how the different agencies are understood to work together – showing clearly what type of multi-agency intervention it is. For example, a ‘separate strands’ multi-agency intervention would show the different agencies producing separate outputs, which later combine to produce the intended outcomes and impacts; a “relay” multi-agency intervention would show how the outputs from one agency are the inputs for another agency.

Are the criteria for evaluating the success of the intervention clear and agreed and is there agreement about the standard of performance required?

The criteria for success should be made explicit and reviewed by all evaluation stakeholders in order to ensure that there is consensus on the evaluation criteria. Each agency participating in the intervention will have its own particular areas of concern, depending on its specific mandate and this will determine what should be looked at to assess whether success has been achieved.

Each agency is also likely to have an institutional approach that stipulates what standards need to be met in relation to each criterion: in most instances these will relate to the norms and standards used to assess and guide performance, although different terminology may well be used in different agencies. In certain instances, these standards may be implicit and may not have been articulated in a written document, which should be done for the purposes of the evaluation. Making performance standards explicit and capturing them in a shared document will enable all evaluation stakeholders to understand what will be considered success (or not) and avoid disagreements during the analysis and reporting phase.

Is there agreement about how to synthesize evidence to form an overall judgement of success?

Synthesis of evidence to produce an evaluative judgement (whether of the whole intervention or aspects of it) is not a process of applying a formula, but of making transparent and defensible judgements. It is rarely appropriate to base the overall evaluative judgement of an intervention on a single performance measure. It usually requires synthesizing evidence about performance across different dimensions.

Annex 1: Works Cited

3ie. 3ie Impact Evaluation Glossary. International Initiative for Impact Evaluation: New Delhi, India. 2012

ADB and EBRD. Performance Evaluation Report: Kazakhstan And The Kyrgyz Republic: Almaty-Bishkek Regional Road Rehabilitation Project. 2009.

ALNAP 20th Biannual Workshop Report. 2007.
<http://www.alnap.org/pool/files/Jointevaluationsworkshop2.pdf>

Bamberger, M. and Rao, V. and Woolcock, M. Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development. World Bank. 2010.

Bennet, A. Process tracing and causal inference, in Henry Brady and David Collier (eds.) Rethinking Social Inquiry: Diverse Tools, Shared Standards. Rowman and Littlefield. 2010.

Binnendijk, A. Effective Practices in Conducting a Multi-Donor Evaluation. Paris: OCED/DAC. 2000.

Brinkerhoff, R.O. The Success Case method: Find out Quickly what's working and what's not? San Francisco, CA: Berrett-Koehler. (2003)

Brinkerhoff, R.O. Telling Training's Story: Evaluation Made Simple, Credible, and Effective. San Francisco, CA: Berrett-Koehler. (2008)

Center for Global Development. When will we ever learn? Washington DC, 2006.

Davidson, E.J. (2009) Improving evaluation questions and answers: Getting actionable answers for real-world decision makers. Presentation at the American Evaluation Association conference. Available at <http://bit.ly/MRUqci>

Development Assistance Committee, Organisation of Economic Cooperation and Development. Glossary of Key Terms in Evaluation and Results Based Management. Paris, 2001.

DFID. Broadening the Range of Designs and Methods for Impact Evaluation. Working Paper No. 38, April 2012. (Stern, E. Stame, N. Mayne, J. Forss, K. Davies, R. Befani, B.)

Earl, S., Carden, F. and Smutylo, T. (2001) Outcome Mapping: Building Learning and Reflection Into Development Programs. Ottawa: International Development Research Centre.
<http://www.outcomemapping.ca/>

Evaluation in the context of global public goods, Rob van den Berg, Evaluation 2011 17: 405

Evalue Research (2010). Final Evaluation Report of the Recognised Seasonal Employer Policy.
<http://dol.govt.nz/publications/research/rse-evaluation-final-report/final-06.asp>

Feinstein, O. and Ingram. G. (2003). Lessons Learned from World Bank Experiences in Joint Evaluation. DAC Working Party on Evaluation.

- GEF Evaluation Office (2008) Joint Evaluation of the GEF Small Grants Programme. [http://sgp.undp.org/img/file/SGP%20Joint%20Evaluation%202008\[1\].pdf](http://sgp.undp.org/img/file/SGP%20Joint%20Evaluation%202008[1].pdf)
- GEF Evaluation Office. Impacts of Creation and Implementation of National Parks and of Support to Batwa on their Livelihoods, Well-Being and Use of Forest Products. Namara, A. 2007.
- Health Scotland. 'Multiple Results Chains Showing Partner Contributions to Shared Health Outcomes. 2009.
- Hughes, K. and Hutchings, C. Can we obtain the required rigour without randomization? Oxfam GB's non-experimental global performance framework. 3ie Working Paper No. 13. 2011
- Mayne Contribution analysis: addressing cause and effect, in K. Forss, M. Marra and R. Schwartz (eds.) Evaluating the Complex. New Brunswick, NJ: Transaction Publishers. 2011
- Mayne, J. Addressing attribution through contribution analysis: using performance measures sensibly. Canadian Journal of Programme Evaluation, 16:1–24. 2001.
- NONIE. Impact Evaluations and Development. Nonie Guidance on Impact Evaluation. (Leeuw, F. and J. Vaessen, J.) 2009.
- OECD/DAC. Guidance for Managing Joint Evaluations. 2006.
- Pawson, R. and Tilley, N., Realistic Evaluation. Sage. 1997.
- Rogers, P. Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation, in: Designing Impact Evaluations: Different Perspectives. 3IE. 2009.
- Scriven, M. Can we infer causation from cross-sectional data. The Evaluation Center, Western Michigan University, 2005.
- Toulemonde, J., Fontaine, C., Laudren E. and Vincke, P. Evaluation in Partnership. Practical Suggestions for Improving their Quality'. Evaluation, 4 (2): 171-188. 1999.
- UK Cabinet Office. Quality in Qualitative Evaluation: A framework for assessing research evidence. Government Chief Social Researcher's Office. Spencer L, Ritchie J, Lewis J and Dillon L. London. 2003.
- UNEG. Concept Note: Impact Evaluation among UNEG Members, Annex 2, Impact Evaluation in UNEG members. New York. 2010.
- Van den Berg, R.D. Evaluation in the context of global public goods. Evaluation 2011 17: 405-415.
- WFP and UNHCR Mixed Method Impact Evaluation The Contribution of Food Assistance to Durable Solutions in Protracted Refugee Situations: its impact and role in Chad.
- White, H. and Phillips, D. Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework. 3ie. New Delhi. 2012.
- Woolcock, M. Towards a Plurality of Methods in Project Evaluation: A Contextualised Approach to Understanding Impact Trajectories and Efficacy. Working Paper 73, University of Manchester: Brooks World Poverty Institute.

World Food Program Office of Evaluation (2010) Concept Note Impact Evaluations 2010-2011.
<http://bit.ly/KKhhYA>.

Annex 2: Agency Specific Definitions of Impact cited by UNEG Members³⁹

Agency	Definition Used
CTBTO	For the purpose of the verification of the Comprehensive Nuclear-Test-Ban Treaty (CTBT), the DAC definition, i.e. “positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended”, is adapted as “positive and negative, primary and secondary long-term effects produced by the development and operation of the CTBT verification system, directly or indirectly, intended or unintended”.
DPI	No specific impact evaluation activities or definition reported.
ESCAP	Changes and effects, positive and negative, planned and unforeseen, resulting from the program with respect to the ultimate beneficiaries and other affected stakeholders.
FAO	The OECD/DAC definition is considered broadly valid. FAO Office of Evaluation defines impact as ‘lasting and significant change’ in institutions, policies, individual capacities, livelihoods, production patterns and levels, food consumption and security, incomes, etc., that can be attributed to FAO or to which FAO has contributed.
GEF	DAC definition, modified as appropriate to focus on the global environmental objective of GEF activities.
IAEA	The long-term effect of change, direct or indirect, on the identified needs which, when combined with other efforts, results from Agency involvement.
IFAD	Impact is defined as the changes that have occurred in the lives of the rural poor (whether intended or unintended, positive or negative, direct or indirect) as a result of a development intervention.
ILO	The OECD/DAC definition is broadly accepted; ILO’s primary beneficiaries are governments and organizations, whereas the household and individual level are mostly reached through government’s action, thus not under ILO’s direct responsibility. The institutional level and the contribution aspects of the ILO are most relevant for the Evaluation Office.

³⁹ Source: Concept Note: Impact Evaluation among UNEG members. Annexes. (Table 1, P7-9). UNEG 2010, updated 2013

IOM	The OECD/DAC definition is broad enough; IOM's primary beneficiaries are migrants, although it also reaches out to governments with policy advice and to the public at large through awareness campaigns, particularly for counter trafficking activities and to combat illegal migration.
OCHA	OCHA does not yet undertake impact evaluations or have any formal definition of impact. Insofar as it has considered the issue, it leans towards an OXFAM UK definition: "The systematic analysis of the lasting or significant changes – positive or negative, intended or not, in people's lives brought about by a given action or a series of actions".
OIOS	According to the Inspection and Evaluation Division manual, impact refers to the ultimate, highest level, or end outcome that is desired. In OIOS inspections and evaluations, impact is considered part of effectiveness. This broader definition may be divided into smaller subsets (particular types of impact), such as impact on legislative frameworks, impact on behavioural norms, impact on the ways in which police and other uniformed services are trained, impact on the visibility of an issue, etc. These are frequently easier and more useful to assess than impact in general.
OPCW	No specific impact evaluation activities or definition reported.
UNCDF	Uses a definition thought to be derived from the UNDP Evaluation Policy: "Actual or intended changes in human development as measured by people's well-being".
UNCTAD	The OECD/DAC definition is fine; most of UNCTAD's work has governments and institutions as primary beneficiaries, although some technical assistance work also aims at enterprises and individuals.
UNDP	<p>UNDP does not use the word "impact". It defines all its results in terms of "outcomes". It evaluates effects of its programmes as outcomes. So it conducts outcome evaluations rather than impact evaluations. However, the distinction is to some extent semantic. In fact, some of the outcomes are expressed as long-term objectives and could also be seen as impacts.</p> <p>In its thinking about impacts, It supports the standard DAC definition, whilst following the UNDP nomenclature, which is based on outcomes. It focuses on actual or intended changes in human development as measured by people's well-being.</p>
UNEP	The OECD/DAC definition is broadly accepted; UNEP's primary beneficiaries are governments and institutions and the Programme should have a catalytic role; the Evaluation Unit considers that in a number of cases, the causal chain leading to impact on the environment of UNEP's work can be identified.
UN-ESCAP	Impact in ESCAP is defined as "Member States' achievements in bringing about benefits for ultimate target groups." Impact is thus considered a shared

	responsibility of ESCAP member States and the secretariat.
UNESCO	DAC definition
UNIDO	DAC definition
UNIFEM	DAC definition
UNFPA	No specific reflection in UNFPA on impact definition, by default the OECD/DAC definition is accepted but no impact evaluations are conducted.
UNICEF	DAC definition at the level of children and women in relation to the rights contained in the CRC and/or the goals/objectives established in the Millennium Declaration and the World Fit for Children Declaration.
UNODC	The highest result level currently defined is “Project Objective”, defined as “The long term benefit the target group will receive”.
UNRWA	Long-term changes, whether planned or un-planned, positive or negative, direct or indirect, that a programme or project helped to bring about
UNV	No specific impact evaluation activities or definition reported.
WFP	Lasting and/or significant effects of the intervention – social, economic, environmental or technical – on individuals, gender and age-groups, households, communities and institutions. Impact can be intended or unintended, positive and negative, macro (sector) and micro (household)
WHO	Given the wide range of work performed by WHO, at different levels of the health system, the Internal Oversight Service does not have a standard working definition of ‘impact’ that is applicable to the evaluations that it carries out.
WIPO	The OECD/DAC definition is fine; WIPO’s primary beneficiaries are governments and institutions.
WMO	No specific impact evaluation activities or definition reported.