



UNEG
United Nations Evaluation Group

UNEG Handbook for Conducting Evaluations of Normative Work in the UN System

This Guidance Document was prepared by the UNEG Task Force for the Evaluation of Normative Work with the help of Mr. Mark Stiles, which is gratefully acknowledged.

November 2013

Table of Contents

Acronyms and Abbreviations	4
Introduction.....	5
Chapter 1. Determining the Evaluability of Normative Work.....	12
Chapter 2. Preparing the Evaluation	20
Task 2.1: Identifying stakeholders and their role in the evaluation.....	20
Task 2.2: Determining the purpose and key questions of the evaluation	22
Task 2.3: Selecting the criteria with which to frame the evaluation.....	26
Task 2.4: Crafting indicators.....	29
Task 2.5: Selecting the appropriate evaluation design and methodology.....	32
Task 2.6: Selecting the evaluation team	42
Task 2.7: Preparing the evaluation terms of reference	43
Chapter 3. Collecting and Analysing Data	47
Task 3.1: Collecting data	47
Task 3.2: Analysing data.....	48
Task 3.3: Identifying lessons and best-fit practices	50
Chapter 4. Preparing the Evaluation Report	54
Task 4.1: Preparing the evaluation report	54
Chapter 5. Following Up on the Evaluation	57
Task 5.1: Disseminating the evaluation report	57
Task 5.2: Preparing a management response.....	58
Task 5.3: Using the evaluation.....	58
Task 5.4: Evaluating the evaluation.....	58
Documents Reviewed	60

Tables

Table 1. Examples of Normative Work in the UN System	6
Table 2. Steps in an Evaluability Assessment	12
Table 3. Typical Challenges at the Evaluability Stage and Possible Solutions	14
Table 4. Example of a Completed Template for Categorizing Stakeholders	20
Table 5. Key Evaluation Questions by Category of Normative Work	24
Table 6. Indicators for Advocacy Work of a Normative Nature	30
Table 7. Evaluation Matrix of Normative Outcomes Related to Biodiversity Standards	32
Table 8. Choices in Types and Designs for Evaluations of Normative Work	33
Table 9. Evaluation Methods for Initiatives Supporting the Development of an International Norm, Standard or Convention	35
Table 10. Evaluation Methods for Initiatives Supporting Governments to Integrate an International Standard into Legislation, Policies and Development Plans	37
Table 11. Evaluation Methods for Initiatives to Assist Governments to Implement Legislation, Policies and Development Plans Based on International Standards	38
Table 12. Evaluation Methods for Use in Special Circumstances	39

Figures

Figure 1. Evaluation Cycle	11
Figure 2. Example of a Theory of Change	16
Figure 3. Common Methodologies by Type of Evaluation	39

Boxes

Box 1. Scenario Calling For an Alternative Approach to Evaluating Normative Work	17
Box 2. Lessons from Evaluability Assessments	18
Box 3. Quality Check – Conducting an Evaluability Assessment	19
Box 4. Uses of Evaluation of UN Normative Work	23
Box 5. UNEG/DAC Evaluation Criteria with Normative Questions	27
Box 6. Questions to Frame an Evaluation of a Programme Designed to Strengthen National Institutional Support for Cleaner Production Standards	28
Box 7. Lessons in Preparing Evaluations of Normative Work	44

Box 8. Quality Check – Preparing Evaluations of Normative Work	45
Box 9. Steps in a Contribution Analysis	49
Box 10. Lessons in Collecting and Analysing Data	51
Box 11. Quality Check – Collecting and Analysing Data	52
Box 12. Lessons – Preparing the Evaluation Report	55
Box 13. Quality Check – Preparing the Evaluation Report	55
Box 14. Lessons – Evaluation Follow-Up	59

Acronyms and Abbreviations

AI	Appreciative Inquiry
CEDAW	Convention on the Elimination of All Forms of Discrimination against Women
CP	Cleaner Production
CSO	Civil society organization
FAO	Food and Agriculture Organization of the United Nations
GEF	Global Environment Facility
IAEA	International Atomic Energy Agency
IDRC	International Development Research Centre
ILO	International Labour Organization
MDGs	Millennium Development Goals
MSC	Most Significant Change
OECD/DAC	Organisation for Economic Co-operation and Development/Development Assistance Committee
ToR	Terms of Reference
UN	United Nations
UNDP	United Nations Development Programme
UNEG	United Nations Evaluation Group
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children’s Fund
UNIDO	United Nations Industrial Development Organization
WHO	World Health Organization

Introduction

Background

1. The United Nations Evaluation Group (UNEG) is a professional network that brings together the units responsible for evaluation in the United Nations system and in international organizations and funds working closely with the UN system. It aims to strengthen the objectivity, effectiveness and visibility of the evaluation function and to advocate for the importance of evaluation for learning, decision-making and accountability. As part of its mandate, UNEG formed a Task Force on Evaluation of Normative Work in response to an increased call for such evaluations and a dearth of relevant resources.
2. At the UNEG Annual General Meeting, 2012, the Task Force, in consultation with UN Heads of Evaluation, defined normative work as:

The support to the development of norms and standards in conventions, declarations, 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. (UNEG, 2012, p. 5)

3. Note that the definition specifies three categories of normative work:
 - a) the development of norms and standards;
 - b) the support to governments and others to integrate the norms and standards into legislation, policies and development plans; and
 - c) the support to governments and others to implement legislation, policies and development plans based on the international norms, standards and conventions.
4. Some norms, such as conventions, are legally binding upon ratification and oblige State parties to comply with the terms of the norms. In addition, certain norms provide for the State parties' reporting obligations within the UN system concerning the application of the said norms.¹ This, for example, includes reporting on agreed environmental targets. Other norms may consist of general guidelines or directions that are applied on a voluntary basis.²

¹ Some of the supervisory/monitoring mechanisms within the UN system were established long ago. For instance, for the ILO, the State parties' reporting obligations were set out in the Treaty of Versailles in 1919, and the International Labour Conference Committee on the Application of Standards and the Committee of Experts on the Application of Conventions were established as early as in 1926.

² The UNEG Handbook on Integrating Human Rights and Gender Equality in Evaluation (2011) provides the following clarification with regard to the binding legal effect of international instruments related to human rights: "International human rights law is not limited to the rights enumerated in treaties. It also comprises rights and freedoms that have become part of customary international law, binding on all States, including those that are not party to a particular treaty. Judicial decisions of the international or regional courts and of international

5. The UN’s leadership role in normative work remains one of its strongest comparative advantages. Through consensus, it sets universally applicable norms and standards for peace and security, economic and socio-cultural development, human rights, the rule of law, health and environmental sustainability, and others.
6. The vast majority of organizations within the UN system, including funds, programmes, specialized agencies and affiliated organizations, are involved in normative work. For many, normative work is at the core of their mandate. For example, specialized agencies, such as the World Health Organization (WHO), the United Nations Educational, Scientific and Cultural Organization (UNESCO), the United Nations Industrial Development Organization (UNIDO), the Food and Agriculture Organization of the UN (FAO), and the World Intellectual Property Organization (WIPO), collaborate with other UN organizations and Member States to develop international conventions, technical standards and codes of practice and to support their implementation. The International Labour Organization (ILO) works with Member States to adopt international labour standards drawn up by representatives of tripartite constituents (governments, employers and workers). Others, such as the United Nations Children’s Fund (UNICEF) and the United Nations Development Programme (UNDP), cooperate with Member States to achieve international standards in relation to child rights, poverty reduction, Millennium Development Goals (MDGs), democratic governance and response to crisis situations.³
7. Table 1 illustrates some of the different forms of normative work of UN organizations. Although not a definitive classification, the table gives a sense of the scope and variety of normative work within the UN system.

Table 1. Examples of Normative Work in the UN System

Types	Description	Examples
Development of international conventions, protocols and declarations	Conventions are the basis of international laws, while protocols and declarations can be non-binding. Member States participate in their development and are responsible for implementation.	The Convention on Biodiversity; the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW); International Labour Organization (ILO) conventions with respect to labour standards and international laws; the UNESCO Convention concerning the Protection of the World Cultural and Natural Heritage; the Convention on the Rights of the Child; the United Nations Convention on the Law of the Sea; the International Health Regulations and the Framework Convention on Tobacco Control (WHO), the Universal

monitoring bodies also have a significant role in international human rights law as they provide further clarifications on the scope of States obligations and the content of the rights. There are many other non-binding universal and regional instruments (declarations, principles, guidelines, standards, rules and recommendations) relating to human rights. These instruments have no binding legal effect, but have an undeniable moral force and provide practical guidance to States in their conduct.”

³ International Organizations and Funds members of UNEG can also have a role in normative work, for instance, the International Organization for Migration (IOM) for migration-related matters.

		Declaration of Human Rights and the International Covenants on Civil and Political Rights, and on Economic, Social and Cultural Rights; and the Montreal Protocol on Substances that Deplete the Ozone Layer.
Establishment of norms, standards, international codes of conduct and guidelines	Many UN organizations are involved in establishing norms, standards and international codes of conduct. In some cases, these are adopted by State parties in the form of legislation. In other cases, they may be incorporated into policy or programme practices, or viewed as exemplary practices.	CODEX ALIMENTARIUS, the joint World Health Organization (WHO)-Food and Agriculture Organization (FAO) food standards code; United Nations Industrial Development Organization's (UNIDO) work on Standards, Metrology, Testing and Quality Infrastructure and energy efficiency; the FAO Code of Conduct for Responsible Fisheries; the International Civil Aviation Organization's standards, regulations and procedures for aviation safety, security, efficiency and regularity; the International Atomic Energy Agency (IAEA)'s nuclear safety standards; the UN Code of Conduct on Sexual Exploitation and Abuse.
Monitoring and reporting on the implementation of conventions, norms, and other obligations	In general, the reporting obligations concern ratified conventions. However, monitoring of the State parties' application of norms could go beyond that.	The inter-agency work on monitoring and reporting mechanisms regarding violation of child rights in situations of armed conflict. The monitoring and reporting by the Office of the High Commissioner for Human Rights (OHCHR) on the implementation of conventions, norms and other obligations directly on the basis of the High Commissioner's mandate. The work of WHO with respect to the eradication and elimination of selected diseases agreed by the World Health Assembly, and that of the World Trade Organization in monitoring trade policies, regulations and practices and in strengthening trade capacity; and, pursuant to the ILO Declaration on fundamental principles and rights at work (1998), the examination of periodic reports submitted by State parties on the measures taken to give effect to unratified core ILO conventions and recommendations.
Advocacy	Advocacy work can take many forms, including monitoring, reporting on and promoting norms and standards directly via dialogue and communication and indirectly by mobilizing NGOs.	UNICEF's advocacy work with respect to the Convention on the Rights of the Child. The advocacy work of the UN to support the MDGs. The IAEA's work to promote the safe use of nuclear power. UNIDO's work on promoting green industry.
Development and dissemination of normative products,	This includes publications of norms and standards, the production and dissemination of statistical information and the maintenance of databases,	Almost all UN agencies engage in this work. For example, FAO's reports on the global status of fisheries, forestry, agriculture, food security; ILO's annual reports on the Application of International Labour Standards; UNDP's Human Development Index; UNESCO's

including knowledge products and information resources	and increasingly the production of web-based and e-learning resources aimed at supporting the implementation of norms and standards.	Education for All Global Monitoring Report; UNICEF's Child Rights Index; UNIDO's work on industrial statistics; United Nations Environment Programme's (UNEP) Annual Environmental Review; the nuclear information databases of the IAEA; WHO's International Classification of Diseases, the List of Essential Drugs, Infant Growth Tables, or the Global Health Report; and the statistical information produced and disseminated by some of the regional commissions of the United Nations.
Promotion and capacity strengthening in relation to cross-cutting norms and standards	This includes various forums, sometimes agency specific and in other cases involving multiple agencies, intended to serve as neutral gatherings for policy and technical dialogue.	Promotion of and follow-up on progress made towards the MDGs; the Millennium Declaration on the protection of vulnerable groups and related interventions; promotion of sustainable development, including green growth; UNESCO's work on culture and sustainable development; UNIDO's Global Forum Function; UN system promotion and implementation of gender equality and gender mainstreaming under the guidance of UN Women.
Facilitating inter-governmental dialogue and coordination	One of the greatest strengths of the UN system is its convening power. Many UN organizations facilitate intergovernmental dialogue on norms and standards and coordinate activities regionally and globally.	Intergovernmental meetings and negotiations; various thematic commissions, such as the UN Commission on Sustainable Development, Statistical Commission, Commission on Population and Development, Commission for Social Development; and much of the work of the UN's five regional commissions, and that of the Commission on the Status of Women; and expert group meetings.

8. Not only do many UN organizations assist Member States and partner institutions to adopt norms and codes of practice, but UN organizations themselves are held to standards: human rights, human rights-based approach to programming, gender equality and environmental sustainability, to name a few. It can be argued, therefore, that the evaluation of normative work is applicable across the entire UN system because almost all UN organizations are involved in normative initiatives in one way or another.

Purpose of the Handbook

9. The purpose of this handbook is to guide UN organizations in conducting evaluations of normative work. Its objectives are to:
- Provide an integrated approach to the evaluation of normative work in the UN;
 - Provide hands-on methodological guidance, concise practical examples and tools for conducting evaluations of normative work; and

- Highlight relevant lessons and best-fit practices of the evaluation of normative work inside and outside the UN system.
10. The handbook is intended primarily for evaluation professionals in the UN who are either tasked with the management and/or conduct of evaluations of normative work. It may also be useful to UN staff responsible for designing, managing and monitoring programmes and projects that integrate the norms, standards and codes of practice; and to the UN's partners in its normative work, particularly to governments and civil society organizations (CSOs). The handbook might also be applied when evaluating the internal performance of UN organizations with respect to standards for gender equality, human rights, ethics, conduct of humanitarian assistance and environmental sustainability.
 11. Readers may ask: "Why yet another evaluation guide; what makes the evaluation of normative work different?"
 12. As indicated above, evaluations of normative work constitute a significant part of the evaluations carried out by UNEG member organizations, according to a recent UNEG survey. Although stand-alone evaluations of normative work are less frequent, that may change as UN organizations devote more of their time and resources to upstream work in the future. Agencies that have been successfully evaluating their normative work over the years also perceive the need for a handbook that captures and systematizes effective practices for the sake of knowledge management.
 13. Evaluation of normative work is often challenging. Causality⁴ in normative work is often difficult to establish as one moves up the logic chain,⁵ and particularly challenging at the impact level⁶. Although there may be a direct link between the actions of a UN organization and a national government agreeing to a particular norm, many factors beyond the control of the organization come into play in determining the extent to which the government translates the agreed norm into national policies, legislation and action on the ground. The outcomes and impact of most normative work often depend on the enabling environment and a host of complementary actions by government, UN agencies, and others. Although this does not necessitate a different approach to evaluation or different methods, it does require a great deal of care when planning an evaluation and heightened rigour in sorting out causality.
 14. Much of the normative work carried out in the UN system is difficult to assess because its tactics, outputs and outcomes are constantly shifting and not always well documented. Such is often the case when building consensus for a global standard and when strengthening the

⁴ Causality refers to the relationship between events where one set of events (the effects) is a direct consequence of another set of events (the causes).

⁵ A logic chain, results chain or logic model is a visual tool that identifies inputs, activities, outputs, outcomes and impact and their causal, sequential relationships.

⁶ Impacts are defined as "the positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended" (Development Assistance Committee of the Organisation for Economic Co-operation and Development [OECD/DAC], 2002, p. 24). Impacts are usually measured sometime after the project or programme is over.

political will for the implementation of an international norm. Much of this work is non-linear and constantly shifting. Although it can be argued that many UN programmes are subject to unpredictable contextual variables, much of the UN's normative work takes that possibility to a much higher level.

15. Furthermore, normative work, such as that related to human rights or environmental sustainability, takes a long time to have impact; rarely is it visible within the time-frame of an organization's programme cycle. Hard data are often lacking where UN organizations assume a convener or coordination role, or when they advocate for the adoption of international protocols, guidelines and codes of practice. One reason for this is that the intervention's outcomes⁷ are often not clearly stated in the form of a logic model or theory of change⁸.
16. Finally, there are often no specific data available to evaluate normative work, especially when normative work is not the main focus of the intervention. Sometimes UN staff themselves may not be cognizant of the standards or have them foremost in their minds as they go about their work;⁹ sometimes they pay insufficient attention to integrating gender equality and environmental sustainability standards in their work; and other times stakeholders may have different understandings of the meaning of some normative work categories, such as 'advocacy', 'social mobilization' and 'rights-based approach' to programming (Stiles, 2002). In addition, a good deal of normative work takes place outside of projects and programmes funded from extra-budgetary sources. Not only does this contribute to inadequate information, it can also lead to unclear or widely dispersed accountability (Stiles, 2002). These factors and others create pronounced methodological challenges for the evaluation of normative work.
17. This handbook draws on a wide range of resources, including the *Handbook on Integrating Human Rights and Gender Equality in Evaluation* (UNEG, 2011) and the *UNEG Good Practice Guidelines for Follow up to Evaluations* (UNEG, 2010a). The terminology used throughout the handbook is consistent with definitions established by Organization for Economic Co-operation Development Assistance Committee (OECD/DAC). Terms, such as 'outputs' and 'impact', are defined where they first appear.

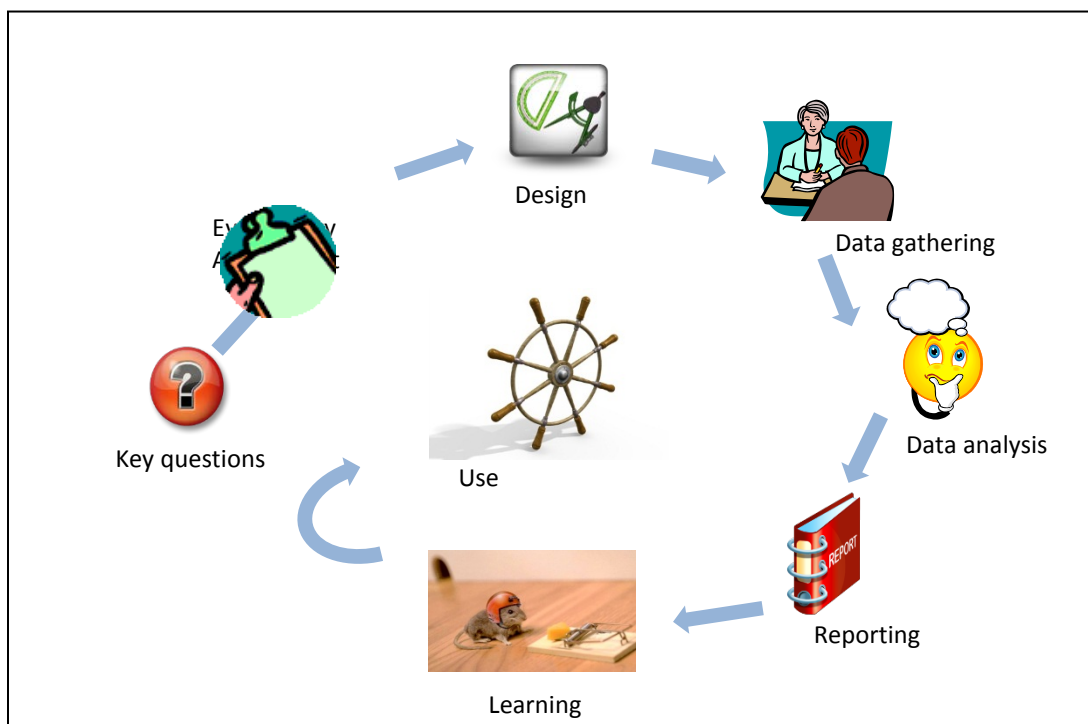
⁷ Outcomes are "the likely or achieved short-term and medium-term effects of an intervention's outputs" (OECD/DAC, 2002, p. 28). The term, 'outcomes', is used interchangeably with 'results'. Short-term or immediate outcomes should be achieved part way through a project or programme; medium-term or intermediate outcomes should be achieved toward the end.

⁸ Theory of change or programme theory is similar to a logic model, but includes key assumptions behind the causal relationships and sometimes the major factors (internal and external to the intervention) likely to influence the outcomes.

⁹ See, for example, Universal Management Group (2012). *Global Evaluation of the Application of a Human Rights Based Approach to UNICEF Programming*. Executive Summary New York: UNICEF.

18. The handbook is organized into five chapters, modelling the evaluation cycle, as illustrated in Figure 1, beginning with determining the evaluability¹⁰ of the normative work in question:

Figure 1. Evaluation Cycle



¹⁰ Evaluability is “the extent to which an activity or a programme can be evaluated in a reliable and credible fashion” (OECD/DAC, 2002, p. 21).

Chapter 1. Determining the Evaluability of Normative Work

19. Determining evaluability means assessing the situation to see if the evaluation is feasible, affordable and of sufficient value to proceed. It includes determining whether the intervention's outcomes are adequately defined and verifiable. An evaluability assessment is particularly important when considering an evaluation of normative work because of the methodological challenges involved as outlined in the introductory chapter.
20. Table 2 shows six practical steps in determining the evaluability of normative work.

Table 2. Steps in an Evaluability Assessment

Steps	Questions	Data sources
1. Analyse the context of the initiative	<p>What is the political, economic, social and cultural context of the normative work?</p> <p>Who are the key stakeholders and beneficiaries?</p>	<p>Background documents</p> <p>Project/progress reports</p> <p>Colleagues</p> <p>Web</p> <p>Books</p> <p>Media reports</p>
2. Identify the nature of the normative work pertaining to the initiative	<p>What is the nature of the normative work?</p> <p>In which of the three categories of normative work does it fit, based on the UNEG's definition?</p> <p>Is it central or peripheral to the initiative?</p>	<p>Background documents</p> <p>Project/progress reports</p> <p>Manager of the initiative</p>
3. Determine whether the normative work is featured in the initiative's design, particularly in the logic model, theory of change, programme development framework and/or the performance measurement framework. ¹¹ If there is no theory of change, develop one and verify it with those responsible for the intervention. ¹²	<p>Are there clear and realistic outcomes for the normative work at the various levels?</p> <p>Are assumptions clearly articulated in the programming logic?</p> <p>Are there valid indicators for each normative outcome?</p>	<p>Background documents (project/programme documents, including logic model etc.).</p>

¹¹ A performance measurement framework outlines in a matrix the intervention's expected outcomes, indicators, data sources, data collection methods, frequency of collection and persons or organizations tasked with data collection. The matrix serves as the basis for a detailed monitoring plan.

¹² A recent comprehensive guide to preparing theories of change is: Funnell, S. and Rogers, P. (2011). *Purposeful Program Theory*. San Francisco: Jossey-Bass.

4.	Determine whether there are solid, readily available data in relation to the normative work	<p>What monitoring data exist? What is its quality and coverage?</p> <p>What background information is available?</p> <p>Where can narrative and anecdotal information be found?</p>	<p>Background documents</p> <p>Progress reports</p> <p>Colleagues</p> <p>Manager/specialists</p>
5.	Determine the use of the evaluation	<p>What purpose will the evaluation serve?</p> <p>How will it be used?</p> <p>What interests do the stakeholders have in the evaluation? What are their expectations and information needs?</p> <p>Who stands to benefit from the evaluation?</p> <p>How receptive are the primary users of the evaluation to learning and change?</p>	<p>Senior management</p> <p>Specialists</p> <p>Key partners and stakeholders</p>
6.	Roughly estimate the time and resources needed for the evaluation and determine whether both are available	<p>When does the evaluation need to be completed, e.g., for decision-making or budgetary purposes?</p> <p>What expertise is needed for the evaluation?</p> <p>What will the evaluation cost?</p> <p>Are the required resources – money and expertise – available and affordable?</p>	<p>Evaluation colleagues</p> <p>Documents on similar evaluations</p>

21. The fourth step – determining whether sufficient data are available to proceed – is relevant to evaluations of all forms of normative work and particularly important when evaluating UN support provided to the development of norms and standards. Key informants – people with rich, historical information about the UN’s role – may have changed over time and documented evidence pertaining to the nature and quality of the UN’s work may be in short supply. In such cases, the evaluation manager has either to focus the evaluation on areas where there are sufficient data or recommend not to proceed with the evaluation.

22. The fifth step – determining the use of the evaluation information – is particularly important. If the evaluation of normative work serves little practical purpose, if the key stakeholders show little interest in learning from it, and if there is limited opportunity to bring about

positive change, then it is doubtful the evaluation should proceed. This step blends with tasks outlined in the next chapter, such as seeking out stakeholder views on the issues of greatest importance for the evaluation, determining how the evaluation should be designed, when it should take place and how the information generated can be used to inform decisions and improve performance. Such information will help to focus the evaluation on the issues most relevant to its users.

23. It may also be useful to plan an inception phase as part of the evaluability assessment, particularly for large complex evaluations, such as global, regional, country and thematic evaluations of normative work. An inception mission can sometimes help strengthen ownership of the evaluation by its intended users, and it can resolve some of the typical challenges evaluators face at this stage, as described in Table 3.

Table 3. Typical Challenges at the Evaluability Stage and Possible Solutions

	Challenges	Possible solutions
1.	The intervention’s context is complex.	<ul style="list-style-type: none"> • Allocate sufficient time and resources to understand the context. • Undertake a web search and/or a literature search. • Undertake an inception mission. • Plan to hire local experts who can bridge the contextual gaps.
2.	The normative work appears to be a peripheral part of the intervention.	<ul style="list-style-type: none"> • Try to understand the reasons why the normative work might be peripheral by contacting those responsible for the intervention’s design and implementation. • Determine whether there is value in evaluating the normative work, e.g. to inform future programming where normative work will have a more central role.
3.	The normative work is not reflected in a logic model, theory of change or country programme framework.	<ul style="list-style-type: none"> • Try to find out why the normative work is not reflected by contacting those responsible for the intervention’s design and implementation. • Construct a retrospective theory of change to include the normative work and validate it with those responsible for the intervention.
4.	The normative work is reflected in the logic model, theory of change or country programme framework, but these are seriously flawed.	<ul style="list-style-type: none"> • If the flaws are major, such that many of the outcomes, indicators and assumptions are unrealistic, where feasible, negotiate with the programme stakeholders and revise them.
5.	Documents provide little information about the normative work and its outcomes.	<ul style="list-style-type: none"> • Contact those responsible for implementation to understand the nature of normative work. Look for information from other sources, such as national governments, reports of other UN bodies, development organizations and local NGOs. • Focus on expected outcomes over which the implementing organization¹³ had control (rather than high-level impacts).

¹³ The term ‘implementing organization’ usually refers to the UN organization that is involved in normative activities either as part of a project or programme or informal advocacy and facilitation, such as convening meetings and developing standards with panels of technical experts. It can also refer to normative work that a UN

6.	There is insufficient time or resources to thoroughly evaluate the normative work.	<ul style="list-style-type: none"> • Scale back the evaluation and focus it on short- and medium-term normative outcomes rather than normative impacts. • Look for ways to save time and resources, such as a desk study, self-evaluation or 1-2 information-rich case studies.
----	--	---

“The terms of reference for OPS4 [*Fourth Overall Performance Study of the GEF*] were highly ambitious. At several junctures, OPS4 encountered the limits of what it could do with existing data and evaluative evidence within the time available and the budgets for the sub-studies. These limitations meant that, on some important points, this report is not able to answer all key questions fully.”

- Global Environmental Facility, *Progress toward Impact. Fourth Overall Performance Study of the GEF*, 2010, p. 4

24. Rectifying the third problem – the absence of normative work in the logic model, theory of change or programme frameworks – must be done in consultation with the programme manager and key programme stakeholders. Ideally, such a model or framework should be produced in a well-facilitated workshop with the full participation of key stakeholders. Although a workshop takes time and resources, it could add value by strengthening the capacity of stakeholders and by fostering ownership of the evaluation by its intended users.

25. When developing a theory of change, it is important to identify the level of control the UN

organization has over the short- and medium-term outcomes of the intervention. This will later help determine the scope and focus of the evaluation. If the intervention falls within the first category of normative work – the creation of norms, standards and international instruments – then the evaluation should focus on process and governance issues, and perhaps the relevance of the norm, standard or international instrument. If the intervention is to support a government in adopting and implementing a particular norm as a duty bearer, then the evaluation should focus on the outcomes related to that support.

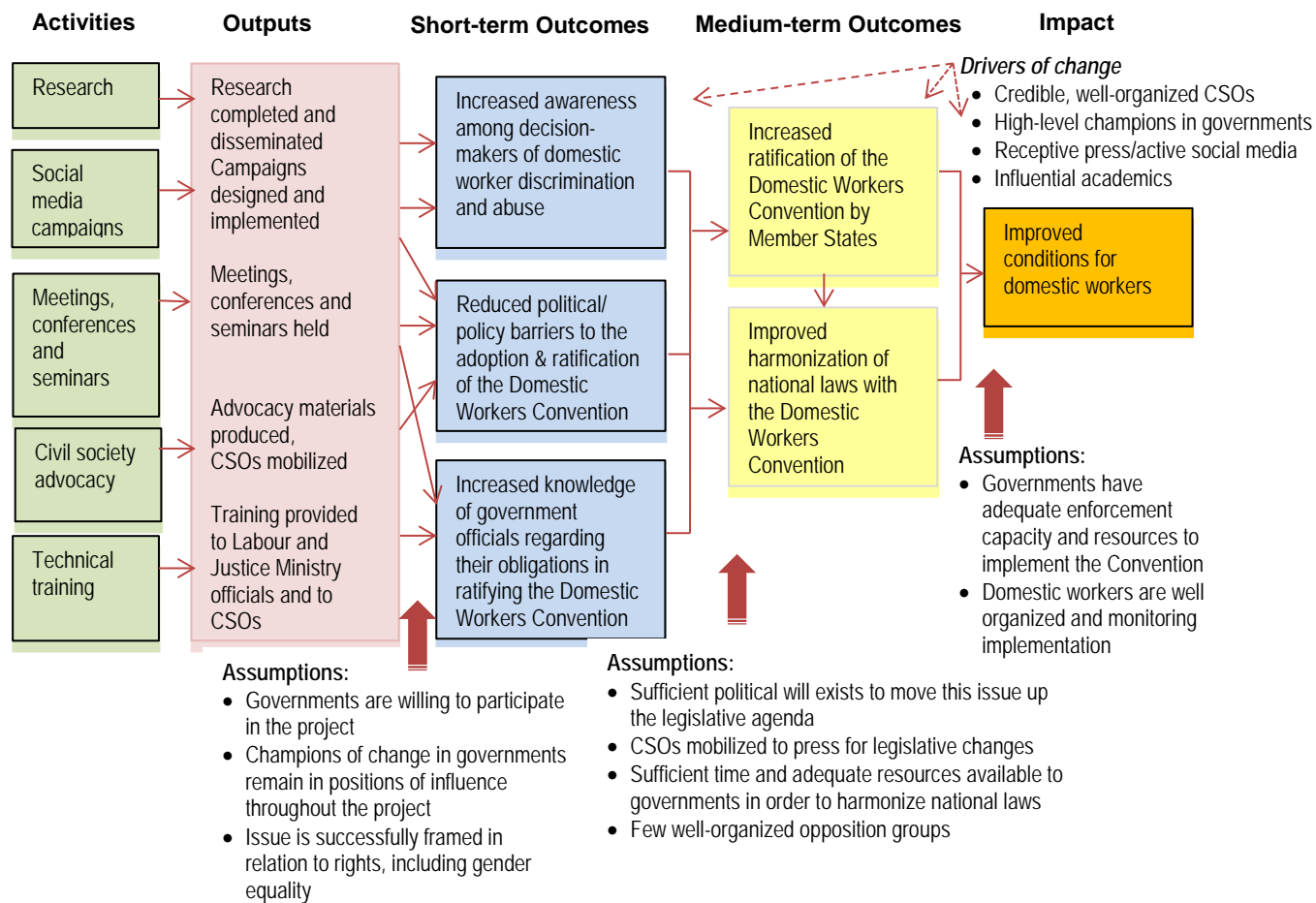
26. Immediate or short-term outcomes usually specify changes in knowledge, awareness and access to resources on the part of the intervention’s beneficiaries. Intermediate or mid-term outcomes usually specify changes in the behaviour or practices of individuals and groups and – for long programmes – the capacity of organizations and institutions. The implementing organization normally has considerable control and influence over these and the short-term outcomes. It is at the impact or long-term outcome level where the most challenging attribution problems occur.¹⁴ For example, although it is realistic that the normative work of a UN organization could result in strengthened policies, regulations and laws aligned with international standards, it would be a stretch to think that the UN organization could contribute directly to improving the lives of the target populations, at least not within the usual project or programme timeframes and budgets. Likewise, it would be unrealistic to expect a UN organization to achieve on its own macroeconomic impacts, such as increased wage earnings, reduced vulnerability of the poor or equal pay for women.

organization does through NGO partners. It is understood that national governments, as duty bearers, have the ultimate responsibility to implement international norms, standards and conventions.

¹⁴ Attribution is “the ascription of a causal link between observed (or expected to be observed) changes and a specific intervention” (OECD/DAC, 2002, p. 17).

27. A related common problem is logic models and programme theories that attempt to capture every possible normative outcome rather than those at the core of the normative work. The logic model or theory of change should focus on the most important outcomes to be achieved given the resources and scope of the intervention. Ideally, the programme logic of a typical project in the UN system should have one impact statement, 2-3 medium-term outcomes and 3-5 short-term outcomes. Get the core outcomes right; any additional outcomes identified through monitoring and evaluation are a bonus.¹⁵
28. A number of guides – some of which are listed at the end of this chapter – may help evaluation practitioners prepare logic models and theories of change. Figure 2 illustrates a theory of change for a fictitious, long-term project aimed at getting more countries to ratify the ILO Domestic Workers Convention.

Figure 2. Example of a Theory of Change



¹⁵ The evaluation of UNESCO’s standard-setting work of the Culture Sector (Part I) – 2003 Convention for the Safeguarding of the Intangible Cultural Heritage (2013) constitutes a good example of an evaluation, for which a retrospective theory of change was prepared and used throughout the evaluation process: unesdoc.unesco.org/images/0022/002230/223095E.pdf

29. The absence of critical data – the fifth problem – may be cause for postponing the evaluation. Data might be lacking because the intervention had not been systematically monitored and no monitoring data on progress made and outcomes achieved had therefore been collected. It might also be due to the fact that medium-term outcomes and impact, such as the actual benefits to citizens and national economies, are unlikely to be seen for many years, as a 2009 FAO evaluation concluded:

The creation of an international instrument is at best an interim outcome. [A]ny determination of the actual usefulness of international instruments must depend on the outcomes and impacts that follow after instrument creation: entry into force, international operations, national implementation and on-the-ground application. These outcomes will typically appear only gradually over the one to two decades following an international instrument's adoption. (FAO, 2009, pp. 2-3)

30. Or, for instance, an intervention that involved a mass media campaign had not yet produced much media usage data. In such a case, it might make sense to delay the evaluation until those data are collected.
31. A good deal of the UN's normative work takes place in volatile environments where outcomes are difficult to establish and where causality is challenging. Such conditions may call for non-traditional approaches to evaluation. Consider the scenario in Box 1.

Box 1. Scenario Calling For an Alternative Approach to Evaluating Normative Work

Country x has signed but not ratified the UN International Covenant on Civil and Political Rights (ICCPR). For several years, UN organizations and others have been quietly lobbying the national government to comply with Article 27 of the ICCPR, which deals with the rights of minority groups to cultural, religious and linguistic freedom. Little of that advocacy has been documented, and collaboration among the UN and donor agencies has been sporadic. Local and international NGOs have also been advocating for change, more vocally than the UN and bilateral donors. The lead UN organization involved has received considerable extrabudgetary funding for this work. It has a logic model but it is out of date, and the organization has not systematically collected data against indicators. Results-based management capacity is weak and reporting has been largely activity-based. The national government has established economic development programmes for minorities. It promised new legislation, but pulled back whenever there were protests, some of which were violent. The lead UN organization is negotiating a sizable grant from a donor for a new initiative aimed at supporting the country to comply with Article 27 and to align its domestic policies and laws accordingly. The UN organization would like to evaluate its programming so as to learn from it and apply lessons to its new initiative.

32. Assuming an evaluability assessment concluded that there was little value in conducting a conventional evaluation, what are the alternatives? The conditions described in the scenario fit well with the criteria for developmental evaluation. Developmental evaluation is evaluation that helps in *developing* programmes and projects, and that is particularly well suited to programme innovation in complex environments (hence the subtitle, “Applying Complexity Concepts to Enhance Innovation and Use” of Michael Patton’s 2011 book, *Developmental Evaluation*). In developmental evaluation, the evaluator typically becomes part of the implementation team, fully participating in decisions and facilitating discussions about what and how to evaluate. The evaluator helps team members conceptualize, design

and test new approaches in a continuous process of experimentation and adaptation, always keenly sensitive to unexpected results and changes in the environment. The evaluator's primary function is to infuse team discussions with evaluative thinking, to bring solid data to strategy deliberations, and to facilitate systematic, evidence-based reflection and decision-making. In short, developmental evaluation aims at helping programmes in complex environments get off to a good start and to perform well to the end through a process of continuous learning, adaptation and innovation.

33. Developmental evaluation may be most applicable to the first category of normative work, the development of international norms, standards and conventions, where it could provide rich data with which to inform decisions regarding process, participation and communication. Obviously, given that in developmental evaluation the evaluator tends to become part of the implementation team, this type of evaluation is not applicable when the evaluation is expected to be strictly independent, as it is usually the case in the UN system.
34. In the end, the decision to proceed with an evaluation – conventional or non-conventional – is a judgement call. There is no formula or universally applicable tools to make this determination.

Box 2. Lessons from Evaluability Assessments

A participatory approach to the development of a theory of change can help uncover implicit assumptions and beliefs that are often not shared among individuals and groups involved in an initiative.

Prepare for monitoring and evaluating normative work when designing a programme by incorporating the normative work in the theory of change and the performance measurement framework. Not only will it help to ensure baseline data and a regular flow of performance information through monitoring, but it will also increase the likelihood of having good data available in time for an evaluation.

An evaluability assessment can be useful regardless of whether the evaluation proceeds or not because it often precipitates substantial improvements to programme designs by uncovering muddled goals, flawed logic, unreasonable assumptions and conflicting or unrealistic expectations about what the programme has set out to achieve (Rog, 1985).

Evaluability assessments often work best when conducted as a team. The team should include the lead evaluator and the evaluation manager, programme implementers, administrators and other stakeholder groups (Morra-Imas and Rist, 2009).

Effective theories of change are often multidimensional. For example, a theory of change related to women's empowerment that focuses on only one dimension, such as economic development and skills training while ignoring such factors as access to markets and socio-cultural impediments, is insufficient (UNEG, 2011).

Box 3. Quality Check – Conducting an evaluability assessment

- ✓ Has the evaluability assessment been carried out following the 6-step process described in this handbook?
- ✓ Have the questions been asked and answered at each step?
- ✓ Has the assessment been carried out with input from the lead evaluator and key stakeholders?
- ✓ Has a theory of change been developed that clearly captures the programme logic and its underlying assumptions?

Useful Resources

- Dozois, E., Langois, M. and Blanchet-Cohen, N. (2010). *DE.201. A Practitioner's Guide to Developmental Evaluation*. Montreal: The J. W. McConnell Family Foundation. Available at: [<www.mcconnellfoundation.ca/>](http://www.mcconnellfoundation.ca/)
- Funnell, S. and Rogers, P. (2011). *Purposeful Program Theory*. San Francisco: Jossey-Bass.
- Gamble, J. (2008). *A Developmental Evaluation Primer*. Montreal: The J. W. McConnell Family Foundation. Available at: [<www.mcconnellfoundation.ca/>](http://www.mcconnellfoundation.ca/)
- GrantCraft (2006). *Mapping Change: Using Theory of Change to Guide Planning and Evaluation*. Available at: [<www.grantrcraft.org>](http://www.grantrcraft.org)
- Patton, M. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford Press.
- Reisman, J., Gienapp, A., and Stachowiak, S. (2007). *A Guide to Measuring Advocacy and Policy*. Baltimore: Annie E. Casey Foundation.
- United Nations Children's Fund (2010). *Advocacy Toolkit: A guide to influencing decisions that improve children's lives*. New York: UNICEF. Available at: [<www.unicef.org/evaluation/files/Advocacy_Toolkit.pdf>](http://www.unicef.org/evaluation/files/Advocacy_Toolkit.pdf)
- W. K. Kellogg Foundation (2004). *Using Logic Models to Bring Together Planning, Evaluation, and Action. Logic Model Development Guide*. Battle Creek, Michigan: W.K. Kellogg Foundation. Available at: [<www.wkkgf.org/knowledge-center/resources/2006/02/wk-kellogg-foundation-logic-model-development-guide.aspx>](http://www.wkkgf.org/knowledge-center/resources/2006/02/wk-kellogg-foundation-logic-model-development-guide.aspx)

Chapter 2. Preparing the Evaluation

35. Once the evaluability assessment is concluded and the evaluation manager determines that the evaluation will proceed, the next step is to prepare the terms of reference (ToR). This chapter describes the tasks involved, drawing on lessons from previous evaluations of normative work. The tasks are not necessarily sequential: much depends on the circumstances, the policies of the evaluation unit and the evaluation manager's personal judgement.

Task 2.1: Identifying stakeholders and their role in the evaluation

36. It is important at this stage to think about who should be involved in the evaluation and when and how they should participate with a view to maximizing the use of the evaluation. Identifying the rights holders and duty bearers and their roles in the evaluation is important particularly for normative work supporting the adoption or implementation of international norms and conventions. Begin with a stakeholder analysis as in Table 4 below, which is based roughly on the scenario presented in Box 1 in the previous chapter. Importantly, the stakeholder analysis will identify the users of the evaluation as distinct from the beneficiaries of the intervention being evaluated.

Table 4. Example of a Completed Template for Categorizing Stakeholders

Beneficiaries of the intervention	Users (clients) of the evaluation
<p>Primary (direct/indirect)</p> <p>Government counterparts (direct) Groups representing minorities (direct)</p> <p>Minorities (indirect)</p>	<p>Primary</p> <p>Local managers of the UN implementing organization Senior managers of the UN implementing organization (headquarters and region) Government counterparts</p>
<p>Secondary</p> <p>Senior government officials (direct) Minority cultural specialists, linguistic scholars and religious leaders (direct and indirect)</p>	<p>Secondary</p> <p>Collaborating donors Collaborating NGOs Other UN organizations Groups representing minorities Academics and evaluation professionals</p>
<p>Others</p> <p>Other minority groups or government officials not reached by the intervention</p>	

37. The different categories of normative work are likely to have different stakeholders. For example, developing norms and standards often involves technical working groups from participating countries with the UN providing administrative support. Normative work that involves the support to national governments as duty bearers to ratify, integrate and

implement the norms often involves a much wider range of stakeholders such as government line departments, CSOs, research institutions and technical bodies. For example, UNESCO's 2010 evaluation of its Strategic Programme Objective 11, "Sustainably Protecting and Enhancing Cultural Heritage", involved more than 100 staff and external experts and almost 200 implementing partners and beneficiaries. UNICEF's 2009 evaluation of DevInfo, a database system for monitoring human development, involved internal and external experts and a wide range of users of the system in six countries.

38. When evaluating normative work where many stakeholders are involved, a mapping exercise is a useful tool as it produces a visual representation of the stakeholders and the relationships among them. A stakeholder map can reveal important contextual information, as well as issues worthy of exploration. In a recent FAO evaluation, the mapping exercise was also used to gather the perceptions of a wide range of stakeholders on the relevance and effectiveness of FAO's work relating to tenure, rights and access issues.¹⁶ A theory of change may help in identifying stakeholders against processes and results at different levels.
39. Evaluation managers should include in the stakeholder analysis individuals who may have been excluded from participating in the normative work, or who may have been negatively affected by it. They also need to make sure that perspectives of both women and men are well represented when deciding who should participate in an evaluation of normative work. This does not necessarily mean that there must be an equal number of women and men; it does, however, mean that the selection should enable the voices and perspectives of both women and men to be heard.
40. Whether to select from the left-hand, right-hand or both columns in Table 4 depends on the purpose of the evaluation and the level of participation needed. At one end of the participatory continuum is *empowerment evaluation* in which the evaluator turns decision-making control, data collection and data analysis over to the primary beneficiaries of the programme (Fetterman, 2001). Although difficult to do well, this model of evaluation is applicable in situations where strengthening the evaluation capacity of stakeholders, such as the rights holders, is a primary objective. At the other end of the spectrum is *third-party, independent evaluation*. Here, the external evaluator is in full control with relatively little participation from those involved in the normative initiative. This approach – common to many UN evaluation units – is particularly well suited in circumstances where there may be controversy or high stakes that necessitate independence on the part of the evaluators and where accountability is the key aim.
41. For many evaluations in the UN system, the preferred model lies somewhere in between, for which the participation of the intended users (right-hand column of Table 4) is essential in order to maximize use of the evaluation. The Paris Declaration principles of 'ownership' and 'mutual responsibility' reinforce this choice, intimating that evaluations

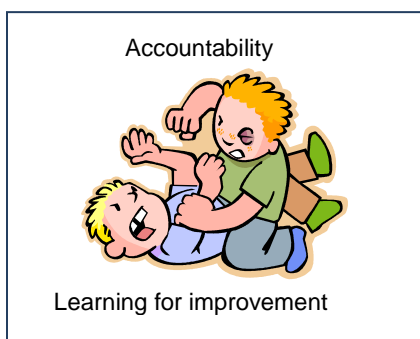
¹⁶ Strategic Evaluation of FAO work on tenure, rights and access to land and other natural resources (January 2012), Annex 3: Stakeholder Perception Study. Available at: www.fao.org/docrep/meeting/025/mC957E.pdf

ought to be conducted jointly with governments or, at the very least, with governments actively participating.

42. One mechanism to enhance user participation, which is pertinent to all three categories of normative work, is the evaluation reference group. How to set up and operate an evaluation reference group is not discussed in this handbook since other, more general evaluation handbooks provide plenty of guidance on this issue.
43. The extent to which the evaluation manager involves stakeholders in data collection and interpretation depends on the sensitivity of the evaluation issues, how the evaluation is to be used, and the time and resources available. There are no set rules. Sometimes it may be necessary to give a little on methodological rigour in order to maximize participation.

Task 2.2: Determining the purpose and key questions of the evaluation

44. Determining the purpose of the evaluation is perhaps the most important task of the evaluation manager during the planning stage. Although it seems obvious, this step is often overlooked or treated lightly. Clarity of purpose is particularly important when evaluating normative work, owing to the challenging nature of the work and its primacy in the UN system. What is the rationale for the evaluation? How is the evaluation information to be used? Is the evaluation forward-looking and improvement-oriented, or is it aimed at determining the value of past normative initiatives only? The answers to these questions will shape the evaluation approach, design and methods.
45. Most evaluations of normative work in the UN system have a two-fold purpose: *accountability* and *learning* and there is often tension between the two. The design of an evaluation where the emphasis is on accountability is likely to be different from one that emphasizes learning and improvement. Designing for accountability is relatively straightforward; the evaluation manager can turn to the UNEG/DAC criteria (relevance, effectiveness, efficiency, sustainability and impact), develop questions around each, and structure the evaluation accordingly. In such a case, stakeholder participation becomes secondary. If, however, the evaluation is forward-looking and learning-oriented, there are a host of issues that could be explored (Box 5), some but not all of which fit neatly into the standard UNEG/DAC criteria¹⁷. In this case, participation of the intended users of the evaluation is crucial.



46. The evaluation manager should, wherever possible, strike a balance between the two purposes. Donors who fund evaluations in the UN system are often driven by accountability concerns, such as results achieved and value for money. However, often a strong case can be made to weigh the evaluation of normative work towards learning for

¹⁷ <www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>

programme improvement, given the non-linear, often qualitative nature of the work and the dearth of information on effective practice. This is not to suggest abandoning accountability concerns but rather to find the right balance.

47. Again, much depends on the circumstances. For instance, if senior managers were contemplating scaling up or moving the normative work into a new phase, then the purpose of the evaluation could be to inform those decisions by identifying the conditions needed to make the normative work successful in the future. For example, the purpose of ILO's mid-term evaluation of Better Work Haiti was "to review the progress made toward the achievement of the immediate objectives of the project and identify lessons learned from its key services implemented to date, its approach taken toward stakeholder capacity building, and its general engagement with national and international stakeholders" (Nexus Associates, Inc., 2012). On the other hand, if the programme was ending with no chance of replication, the primary purpose could be to determine what results were achieved. Box 4 highlights other potential uses for the evaluation of normative work in the UN system.

"Although most policy makers and stakeholders are primarily interested in beneficiary level impact ... policy interventions are primarily geared at inducing sustainable changes at the institutional (government) level ("second-order" effect), with only indirect effect at the beneficiary level."

- Leeuw and Vaessen, 2009, p. 9

Box 4. Uses of Evaluations of UN Normative Work

- To identify performance lessons for the purpose of programme improvement
- To make decisions about scaling up or replicating the normative work and the resources required
- To better understand the forces driving adoption and implementation or non-adoption and non-implementation of the norm(s)
- To help make organizational decisions, such as whether to decentralize responsibility for the normative work
- To determine what partnerships and timelines are required in order to achieve the desired normative outcomes and impact

48. Once the purpose of the evaluation is clear, the key questions it should strive to answer will follow. Here, the evaluation manager has to be disciplined and ask: "What are the *three* most important questions this evaluation can and should answer?" In the end, there may be quite a few more, but the top three ought to be unequivocal. They should be the questions that relate most closely to the purpose of the evaluation and it should be clear to anyone who reads the ToR that the three questions are the priority and that the value of the evaluation is staked on answering them.
49. What are those key questions? They depend on the circumstances. Here is where a reference group or an inception mission could help. The temptation might be to turn first to the UNEG/DAC criteria for questions. The danger is that the evaluation can end up with too

many questions and a lot of information to collect that has little practical use, and not enough information on what is most important; in short, an evaluation that is data rich, but information poor.

50. Table 5 provides examples of questions relevant to the three general categories of normative work in the UN system.

Table 5. Key Evaluation Questions by Categories of Normative Work

Categories	Examples of key questions
1. Support for the development of an international standard	<ul style="list-style-type: none"> • What was the strategy/process for developing the standard, and how successful was it? • What results were achieved in terms of strengthening the capacity of Member States to participate in international forums to develop the standard? How inclusive was the participation? • How successful was the intervention at reaching consensus on the standard with Member States taking ownership of the process and outputs? • How inclusive was the UN organization's support?
2. Support to encourage governments to integrate the standard into legislation, policies and development plans	<ul style="list-style-type: none"> • How effective were the interventions at convincing governments to integrate the international standard into legislation, policies and development plans? • What strategies, methods and approaches worked well, not so well, and why or why not? • How successful was the intervention at identifying and assisting Member States to manage risks and overcome barriers to integrating the standard in legislation, policies and plans?
3. Support to assist governments to implement legislation, policies and development plans based on the international standard	<ul style="list-style-type: none"> • How effective was the intervention in developing Member States' capacities to prepare and implement legislation, policies and development plans related to the standard? • Did the UN organization have the right partners and the right mix of expertise in providing technical support? • How sustainable were the outcomes? • What was the impact at the population level?

51. Key questions related to normative work that was part of a larger initiative might be:

- To what extent were UN staff aware of the organization's normative function and held accountable for achieving normative results?
- How well was the normative role integrated in the planning of the intervention, also in terms of resources allocated and regular monitoring?
- What results were established and achieved specifically in relation to the normative work?
- To what extent has normative work contributed to the efficiency and effectiveness of the intervention(s) to which it relates?

52. An actual example of this comes from the Fourth Overall Performance Study of the Global Environment Facility (GEF) in 2010. Among its key questions were:
- To what extent has the GEF followed the guidelines of the conventions [such as, the Convention on Biodiversity and the Convention on Persistent Organic Pollutants] for which it is a financial instruments?
 - To what extent has the GEF been able to provide feedback to the conventions on their guidance, the implementation of that guidance, and the results achieved, including on multifocal area issues?
 - Is the governance system of the GEF adequate and up to international standards?
53. Questions aimed at evaluating advocacy work to convince governments to adopt an international standard could be:
- To what extent were the intervention's advocacy initiatives based on action research, feedback from target groups and other known effective advocacy practices?
 - How efficient was the advocacy/communication strategy at framing the issue and targeting key decision makers?
 - To what extent did the UN agency adapt its advocacy work in a timely and continuous manner in response to the changing political and economic environment?
54. If the evaluation is assessing normative products, databases and related resources, then there should be questions concerning quality, access and distribution (See, for example, the assessment criteria in Annex 8 of FAO's *Evaluation of FAO's Activities on Capacity Development in Africa* [FAO, 2010]).
55. UNEG's guide to integrating human rights and gender equality in evaluation states that "All UN interventions have a mandate to address HR [human rights] and GE [gender equality] issues" (UNEG, 2011, p. 16). It follows that evaluations of the UN's normative work should include an assessment of human rights and gender dimensions of the interventions. Examples of questions for this purpose are:
- To what extent did the intervention incorporate human rights principles and integrate gender equality considerations?
 - To what extent were the implementing partners accountable for incorporating human rights and gender equality in their normative work?
56. Questions concerning the intervention's design are often important because weak implementation is often linked to flawed design. The same applies to the partnerships involved in the normative work. Did the intervention have the right partners with optimal technical expertise? What was the level of ownership and the degree of buy-in among the partners? Risk identification and mitigation are an important part of design, so an evaluation of normative work could include questions in this realm.
57. Remember that most evaluations have limited resources: few can afford to assess many complex issues within a single evaluation. Better to answer a few key questions thoroughly

than to cover many superficially. Quality, not quantity, should be the watchword. The responsibility to focus the evaluation on what is most important and achievable, given the available time and resources, lies mostly with the evaluation manager.

Task 2.3: Selecting the criteria with which to frame the evaluation

58. The UN system uses the UNEG/DAC's evaluation criteria of relevance, effectiveness, efficiency, impact and sustainability to frame its evaluations, as shown in Box 5.

Box 5. UNEG/DAC Evaluation Criteria and Normative Questions

Relevance

- Has the normative work met needs that its beneficiaries/duty bearers have expressed and agreed to?
- Does the normative work make sense regarding the conditions, needs or problems to which it is intended to respond?
- Is the normative work consistent with the UN organization's mandate and global priorities?
- Is the normative work aligned with the priorities and policies of the host government?
- Is the normative work aligned with the United Nations Development Assistance Framework (UNDAF)?

Effectiveness

- What outcomes have the normative work achieved, expected and unexpected, positive and negative?
- Is the normative work reaching the intended beneficiaries, rights holders and duty bearers?
- What would the likely outcomes be had the intervention not occurred?
- Was the normative work adequately reflected in the programme logic model, theory of change and performance measurement framework?
- Is the normative work likely to achieve intended outcomes by the time the initiative ends?
- How effective was the UN organization's process in relation to stakeholder participation, inclusivity, accessibility, transparency and capacity development?
- To what extent does the normative work integrate gender equality, women's empowerment, human rights and South-South cooperation?
- What value has the UN organization added?
- To what extent are the normative products, including knowledge products and information resources, disseminated by the UN organization used by the various stakeholders?

Efficiency

- Are the resources and inputs (funds, expertise, and time) being used to achieve normative outcomes in an efficient and timely manner?
- Is the relationship between cost and results reasonable?
- Are there sufficient resources to achieve the intended outcomes of the normative work?
- How do the costs compare with similar normative work?
- Are the variances between planned and actual expenditures justified?
- To what extent are the mechanisms used for inter-governmental dialogue and cooperation efficient?

Sustainability

- What is the likelihood that benefits of the normative work will continue to flow long after its completion and without over-burdening local organizations and partner institutions?
- Is the capacity being developed adequate to ensure that institutions/organizations will take over and sustain the benefits envisaged?
- To what extent has the programme built in resilience to future risks?
- To what extent do partners and beneficiaries participate in and 'own' the normative outcomes?
- Does the normative work adhere to sustainable environmental practices and standards?

Impact

- To what extent has the normative work contributed to achieving results at the impact level?

59. The OECD/DAC criteria and questions should serve as a guide. If the evaluation has to include all criteria, the ToR should clearly state which ones are a priority. One way to avoid spreading the evaluation too thinly is to cover some of UNEG/DAC criteria – relevance and

efficiency, for example – with a desk study and perhaps a focus group interview, leaving more resources to delve deeply into the remaining criteria.

60. In framing the evaluation of normative work, the evaluation manager can apply other criteria. Some could be of a technical nature, perhaps dealing with scientific standards (e.g. *quality of science*¹⁸ underpinning standard-setting norms); others could relate to human rights, women's empowerment and principles of behaviour in humanitarian interventions, for instance. Some evaluations could choose to focus on management issues, such as the extent to which decentralization affects normative work, or the technical and managerial capacities governments need to implement standards. The evaluation should be tailored to meet what matters most. Box 6 contains a good example of a compact set of questions from an independent evaluation of the UNIDO-UNEP Cleaner Production (CP) Programme aimed at supporting national host institutions to provide CP services in accordance with recognized international standards.

Box 6. Questions to Frame an Evaluation of a Programme Designed to Strengthen National Institutional Support for Cleaner Production Standards

Relevance: Are the elements of the programme (i.e. the Cleaner Production (CP) concept, the CP services, the National Cleaner Production Centre institution, the global network and the technical assistance inputs) applicable and valuable for the intended beneficiaries (i.e. the private sector, government, academia and research institutes in the host country)?

Effectiveness: Does the design of the programme (i.e. national centres, global management and networking, and technical assistance) and its implementation enable the Centres and beneficiaries to achieve the programme's intended results (i.e. uptake of CP)?

Efficiency: Is the programme designed and implemented to achieve optimal benefit from its available resources? Are the Centres and other programme activities managed and administered in a manner that fosters service delivery to beneficiaries?

Sustainability: Is it probable or likely that the benefits (e.g. availability of CP services, environmental and productivity benefits in industry etc.) achieved from the programme will continue into the future?

Capacity Development: Does the programme develop essential capacities (e.g. in regards to resource productivity, environmental management, entrepreneurship, and/or public private partnerships) for local stakeholders to improve their current and future well-being?

Ownership: Do local stakeholders regard the programme as their own and do they make commitments to advance the programme's aims and objectives and act on its outputs?

Source: United Nations Industrial Development Organization and United Nations Environment Programme (2008). *Independent Evaluation UNIDO-UNEP Cleaner Production Programme*. Vienna: UNIDO, pp. 10-11.

¹⁸ Quality of science does not have a standard international definition. Some aspects of quality highlighted in the literature include: (i) Results meeting objectives and adequate documentation of methods and results (US Nuclear Regulatory Commission, 2005); and (ii) Replicability, novelty, independence, methods, and clarity of contribution (Gill and Gilbert, 2011).

Task 2.4: Crafting indicators

61. Ideally, the evaluation manager should have a theory of change that captures the intended outcomes and impact of the normative work as well as the underlying assumptions. There may be a performance measurement framework or a logical framework analysis with indicators. Better yet, the intervention may have some useful performance monitoring data based on the indicators. In reality, however, it is quite possible that none of these exists or if any do, they may be seriously flawed, in which case, the evaluation manager may have to prepare a retrospective theory of change and a new or revised set of indicators for each outcome and impact statement. The assumption is that most evaluations of normative work in the UN system want to say something about effectiveness. This usually involves assessing the extent to which the intervention achieved its expected results, hence the need for indicators. Indicators should answer the question: “What would we expect to see as verifiable evidence of the achievement of this outcome or impact?”
62. Developing indicators for normative work is challenging. Recall the time and effort that went into developing the indicators for the MDGs. A few general rules apply:
- **Limit the number of indicators to two or three per outcome/impact statement.** Doing so will keep the volume of data to be collected in check. Data collection must not be allowed to become too onerous or expensive. If one or two good indicators can provide conclusive evidence of the intended change, leave it at that.
 - **Include at least one qualitative indicator per outcome statement.** Qualitative indicators allow the evaluator to find depth in information and to tease out important nuances.
 - **Ensure that indicators are gender-sensitive,** wherever possible. A gender-sensitive indicator is a measure of change over time in relation to gender equality, and one that produces sex-disaggregated data. For example: *Extent to which women’s and men’s perspectives are heard in policy dialogue sessions.*
 - **Think SMART:** Indicators should be specific, measurable, attainable, relevant and, where possible, time-bound.
63. Table 6 below gives examples of indicators relevant to an advocacy campaign supported by a UN organization and aimed at encouraging Member States to adopt an international norm.

Table 6. Indicators for Advocacy Work of a Normative Nature

Outcomes	Indicators
1. Strengthened support for the international norm	<ul style="list-style-type: none"> • Percentage change in awareness of the campaign normative issue among target groups; number of target groups (e.g. policy makers and opinion leaders) aware of the campaign principles and key messages related to the norm • Number of key stakeholder groups agreeing on the definition of the problem related to the norm • Extent of media coverage of the normative issue • Diversity of groups involved in the issue as measured by number of 'unlikely allies' supporting the norm • Number of Twitter followers, discrete visitors to the campaign website, media stories per month
2. Strengthened organizational capacity of partners backing the norm	<ul style="list-style-type: none"> • Number of management improvements made in partner organizations to support the implementation of the norm • Ability of partner organizations to implement the norm • Level of stability of partner organizations to implement the norm
3. Strengthened alliances supporting the norm	<ul style="list-style-type: none"> • Number of strategic alliances in support of the norm • Level of collaboration among partners in support of the norm • Size and strengths of coalitions formed to support the norm
4. Strengthened base of support for the norm	<ul style="list-style-type: none"> • Diversity of groups involved in the issue as measured by number of 'unlikely allies' supporting the norm • Extent of media coverage of the normative issue • Number of target groups (e.g. policy makers and opinion leaders) aware of the campaign principles and key messages related to the norm • Number of Twitter followers, discrete visitors to the campaign website, media stories per month
5. Improved policies that reflect the norm ¹⁹	<ul style="list-style-type: none"> • Documented evidence that a new policy on the campaign issue is in development • A new policy reflecting the norm is adopted in the form of an ordinance, legislation, binding agreement • Level of funding applied to implementing the new or improved policy

Adapted from Reisman, Gienapp and Stachowiak, 2007.

64. Collaboration with stakeholders can help overcome the difficulty of crafting indicators for normative outcomes. Ideally, the evaluation manager should form a team for this purpose and the team should include subject matter specialists and at least one wordsmith who has a good deal of experience with indicators. A series of short workshops is best suited for this task because the indicators are likely to undergo several iterations before they are ready. Ideally, the key stakeholders, including the managers whose normative work is being evaluated, should validate the indicators once drafted.

¹⁹ Note that this is a medium-term (intermediate) or perhaps even a long-term outcome.

Through semi-structured interviews in several African countries in 2009, evaluators assessing FAO's capacity development work found that many key stakeholders were unaware of some of FAO's key normative products and/or had difficulty accessing them via the Internet. The finding came as a surprise. Access to, and distribution and use of, normative products became useful proxy indicators of normative performance and led to practical recommendations to rectify the shortcomings.

- FAO, 2010

65. Do not overlook the obvious when preparing indicators. For example, if one of the outcomes is "Increased adoption of food safety standards by Member States", then a simple indicator could be "number of targeted countries that have adopted the standards in the period under review". However, this is insufficient to tell the performance story because there may be some Member States that adopted the standard for reasons other than the intervention. In this case, another indicator is needed so that the evaluator has a second line of evidence. For example: "Number of experts and NGO partners who believe the adoption of the standards would not have happened without the

intervention". Although evaluators should always look for simple measures, sometimes complex measures are needed to tell the performance story particularly when assessing the impact of normative work.

66. Remember that the outcome statements and indicators should focus on the core of the normative work: that which is in the control of the implementing organization and which can reasonably be predicted. Given that most normative work takes place in dynamic environments, there will likely be unexpected outcomes and new indicators that come to light during the evaluation.
67. Most evaluation inception reports feature an evaluation matrix outlining the key questions, subquestions and corresponding indicators, data sources and evaluation methods. Table 7 provides an example from the effectiveness section of an evaluation matrix for a programme aimed at promoting biodiversity standards.

Table 7. Evaluation Matrix of Normative Outcomes Related to Biodiversity Standards

Effectiveness Criterion	Key Question & Subquestions	Indicators	Data Sources	Methods
<p><u>Outcomes:</u></p> <p>1. Increased uptake of biodiversity standards by participating national governments.</p> <p>2. Strengthened awareness of biodiversity standards on the part of national environmental managers.</p>	<p>What results has the intervention achieved?</p> <ul style="list-style-type: none"> • What would be the likely outcomes had the intervention not occurred? • What unexpected results has the intervention achieved? • What are the factors enhancing and impeding results? 	<p>1.1 Number of governments that adopted the standards</p> <p>1.2 Number of governments that enacted laws or regulations reflecting the standards</p> <p>1.3 Perceptions of the partners, including environmental NGOs, regarding the UN organization’s contribution to increased uptake and the factors enhancing and impeding uptake</p> <p>2.1 Number of managers (male/female) who attended awareness-raising events and who report increased awareness</p> <p>2.2 Perceptions of other stakeholders, including environmental NGOs, regarding environmental managers’ awareness of standards</p>	<p>Programme documents</p> <p>National authorities</p> <p>Programme manager and staff</p> <p>NGO representatives</p> <p>Experts and groups that did not participate in the intervention</p>	<p>Document review</p> <p>Interviews</p> <p>Focus group discussions</p>

68. If the intervention has prepared a performance measurement framework, there may be baseline data available. This would be helpful to the evaluators because it would allow them to assess the changes that have occurred since the intervention began. Where a UN organization is working to establish a standard, and no standard was in place beforehand, the baseline would be zero. Where there are no reliable baseline data, the evaluators must try to reconstruct the baseline, usually by means of a document review and interviews with managers. This can be a difficult undertaking, which is particularly challenging when records are imprecise, people who were there at the start of the intervention have moved on, and when the programming context is dynamic.

Task 2.5: Selecting the appropriate evaluation design and methodology

69. The evaluation manager has many choices in designing an evaluation of normative work and selecting methodologies. First, the evaluation manager must decide when to conduct the evaluation because the timing affects the type of evaluation to be carried out. This is an important issue because normative outcomes and impact can be long-term, emergent and unpredictable, as in the case of some initiatives aimed at developing an international norm,

and others that support Member States to align their policies and legislation with international conventions. There are no gold standards: choice is situational and driven by the purpose of the evaluation, its key questions, the time and resources available, and how the evaluation data are to be used. Table 8 provides some guidance to help evaluation managers make prudent choices.

Table 8. Choices in Types and Designs for Evaluations of Normative Work

Types	Description	When to use
Summative evaluation	An evaluation conducted at the end of an intervention, usually to provide information about its value and impact.	At the end of a normative intervention, but best planned as the intervention gets under way.
Formative evaluation	An evaluation conducted part way through an intervention, usually focusing on operational issues and oriented towards identifying strengths and shortcomings. Sometimes referred to as process evaluation and mid-term evaluation.	Part way through a normative intervention so that corrective action can be taken to strengthen outcomes and impact. Best planned as the intervention gets under way.
Developmental evaluation ²⁰	An evaluation that helps in developing an intervention and strengthening its effectiveness as it is implemented.	Well suited to normative work in complex systems where the interventions are experimental and the outcomes are unclear.
Designs		
Experimental	An ex-ante design requiring random assignment of a population to at least two groups, one of which serves as a control or counterfactual.	Use whenever affordable and practical in order to determine whether impact can be attributed to the intervention as the single cause, i.e. that the normative work was both necessary and sufficient to cause the impact. ²¹
Quasi-experimental	A design where an intervention and non-intervention or comparison group (counterfactual) are formed either before or after the intervention, but without randomly assigning the population to a group.	Use when this level of rigour is necessary and feasible in order to prove that the normative work was both necessary and sufficient for a net change at the impact level. Challenging to find comparison groups that are closely matched.
Non-experimental, including theory- and case-based approaches	An evaluation design where no attempt is made to compare intervention and nonintervention groups, and the emphasis is on description.	Use at a single point in time to assess what was planned, what took place, what changed and why.

70. Although experimental and quasi-experimental designs can be powerful ways to answer questions about effectiveness, particularly questions dealing with impact, they are best suited

²⁰ See Chapter 1 for more information on developmental evaluation and an example of where this type of evaluation is appropriate.

²¹ Only one (Coulibaly, 2010) of the more than 30 evaluations of normative work reviewed for this handbook used an experimental or quasi-experimental design.

to conditions where there is one primary cause and one main effect. Much of the UN's normative work, however, involves numerous actors, many potential causes and just as many possible effects. Think, for example, of UN initiatives to convince state governments to accede to international labour laws, human rights conventions or sustainable development protocols.²² A 'causal package' is usually involved in such normative changes; the contents of that package are likely to be different from one country to another; and 'success' may not always look the same in each country.

"Most components of the evaluation used small, purposeful samples of information-rich cases, rather than large, random samples that permit valid generalizations and tests of statistical significance."

- *Evaluation of UNICEF-Government of Pakistan Advocacy and Social Mobilization Strategy* (Stiles, 2002, p. 9)

71. Determining 'success' in initiatives designed to develop a norm or standard is particularly challenging. Sometimes compromises have to be made to achieve consensus, and not every norm or standard is achievable for developing countries. When the picture is not all black and white, it calls for qualitative inquiry with causality questions aimed at capturing the various shades of grey; for example: "What works for whom, why and under what conditions?" (Pawson and Tilley, 1997).
72. What methods should the evaluation manager use to evaluate normative work? There is no single best method or single best combination of methods: much depends on the purpose of the evaluation, its focus, the time and resources available and the intended use of the evaluation information. That said, in most instances a few **general rules** apply:
 1. Use **multiple methods** in order to make up for the weaknesses of any single method. A mix of methods brings an element of complementarity to evaluation: the results obtained from one can often help the evaluator to better understand the results obtained from another.
 2. In most circumstances, both **qualitative and quantitative data** are needed to evaluate normative work in order to untangle the causal package and delve into qualitative issues in depth.
 3. **Multiple lines of inquiry** and **multiple data sources** are advised because they allow for **triangulation**, "the use of three or more theories, sources or types of information, or types of analysis to verify and substantiate an assessment" thereby strengthening confidence in the findings (OECD/DAC, 2002, p. 37).²³

²² For an in-depth look at the stringent conditions needed for experimental and quasi-experimental designs, see: Stern *et al.* (2012). "Broadening the Range of Designs and Methods for Impact Evaluations." *Department for International Development Working Paper No. 38*. London: Department for International Development.

²³ For example, a 2007 evaluation of the political affairs work of the UN Security Council used self-administered surveys, in-person interviews, programme-data analyses, case studies, and a literature review, drawing on 11 different data sources (UN Economic and Social Council, 2007).

73. Following the above general rules, Tables 9, 10, 11 and 12 assesses methods for each category of normative work, beginning with initiatives supporting the development of an international norm, standard or convention.

Table 9. Evaluation Methods for Initiatives Supporting the Development of an International Norm, Standard or Convention

Methodology	Pros	Cons
1. <i>Literature review</i> focused on similar types of support and evaluations. Should include selected resources from the end of each chapter in this handbook.	May reveal lessons about successful and unsuccessful practices under similar circumstances.	Expensive unless focused and/or undertaken by a junior evaluator.
2. <i>Document review</i>	Usually necessary in order to understand the context, stakeholders, roles, activities and expenditures. Often helpful in answering questions pertaining to relevance and efficiency.	Time-consuming. May not yield much data related to the indicators of outcomes and impact.
3. <i>Key informant interviews</i>	Most effective means of collecting in-depth qualitative information, e.g. exploring what worked well, what did not, and why.	Time-consuming. May be expensive if key informants are many, and at a distance. Telephone and Skype interviews are generally less effective than face-to-face interviews but much cheaper. Challenging to control for bias. Requires skilled evaluators and takes considerable time to set up interviews.
4. <i>Focus group discussion</i>	Can yield in-depth qualitative data. Can be done with different groups of stakeholders in order to compare points of view on a normative issue. A good method to get the views of many in a short time. Best results often come when group members are involved in similar work and/or at the same level.	Typically requires 6-10 participants with similar backgrounds for several hours, which may be difficult to arrange. Two evaluators needed: one to facilitate; the other to take notes. Must be well facilitated to get full participation. Requires group members to be at the same level and/or involved in similar work. Unsuitable where there is a rigid hierarchy in the group.

5. <i>Self-administered survey</i>	Self-administered surveys come in many forms, including mail-in, e-mail and online surveys. E-mail surveys are suitable for obtaining information from large numbers. Online surveys are fast, low cost, and have a potential wide reach. Readily available software makes it easy to crunch numbers to determine variances, and to generate tables and charts.	In order to have good response rates from self-administered surveys, recipients must be motivated or obliged to respond. For online surveys, response rates often plunge after 10 minutes, so surveys must be kept short and confined mainly to closed-ended questions unless incentives are provided. Results are not representative and open to bias when participants are self-selected.
6. <i>Direct observation, e.g. when Member States meet to consider a standard</i>	Useful for assessing the process used in developing a norm or standard, stakeholder participation, the quality of partnerships and normative leadership. Should be semi-structured using an assessment checklist or grading matrix. Helpful to have two evaluators observe together and compare notes afterwards. ²⁴	Difficult to control for bias. Requires experienced evaluators.
7. <i>Cost-benefit analysis and cost-effectiveness analysis</i>	Useful for answering questions related to efficiency. Can be used to compare the cost of one intervention to another when the conditions are similar.	Accurate financial data may not be readily available. May require financial expertise for data analysis. Evaluators must be careful when interpreting the results of the analysis, because the benefits of normative work are often difficult to quantify. ²⁵
8. <i>Mapping, e.g. mapping of stakeholders or mapping of the process used to achieve agreement on a standard or norm</i>	Helps to identify multiple stakeholders, their roles and relationships, and the drivers/blockers of change. This method, can be done when preparing for an evaluation, can also be employed during the evaluation to create visual maps of stakeholders and their relationships in each participating country as a way of identifying the variety of challenges and opportunities Member States face in obtaining support for standards.	Not particularly helpful when only a few stakeholders are involved. Challenging to do when there are many stakeholders and/or when the normative work has involved a complicated process over a long period.

²⁴ In a recent evaluation, the Inspection and Evaluation Division of the Office of Internal Oversight used direct observation as one of several methods to review the Department of Economic and Social Affairs' normative work (United Nations Economic and Social Council, 2011).

²⁵ For an in-depth discussion of the difficulties involved in cost-benefit analysis, see *Cost-Benefit Analysis in World Bank Projects* (World Bank, 2010), available at: siteresources.worldbank.org/INTOED/Resources/cba_overview.pdf

74. Evaluation of the development of an international norm, standard, or convention is likely to focus on assessing the processes and/or governance structures used to achieve agreement or consensus on a standard, norm or convention. The evaluation would be largely qualitative and would require robust methods for qualitative inquiry such as those in Table 9. Agreement or consensus on a standard, norm or convention should be reasonably straightforward to determine. More challenging might be the unexpected outcomes, both positive and negative. Interviews and focus group discussions would be among the best methods to uncover unexpected outcomes, and it would be prudent to include some groups of stakeholders who may have been excluded or who chose not to participate in the normative process.
75. A typical methodological challenge for evaluation in this category of normative work is that the representatives of partner countries or organizations change frequently. In this case, the evaluation manager should try to find those people who have been involved long enough to have informed, historical perspectives on the normative work. This would require purposeful sampling, which is legitimate for qualitative inquiry. Although the evaluators must be careful not to over-generalize, they can extrapolate from the data gathered from a purposeful sample. Extrapolations are “modest speculations on the likely applicability of findings to other situations under similar, but not identical, conditions” (Patton, 2008, p. 459). Case studies may also be appropriate if the evaluation is attempting to compare multiple cases of normative work aimed at developing a norm, standard or international convention.

Table 10. Evaluation Methods for Initiatives Supporting Governments to Integrate an International Standard into Legislation, Policies and Development Plans

Methodology	Pros	Cons
1. <i>All in Table 9 may apply</i> ²⁶		
2. <i>Case studies</i>	Helpful where a primary objective is to gain in-depth understanding of the programming context and the reasons why a normative intervention did or did not succeed. Choosing 4-5 case units and analysing within each unit and across units can yield important qualitative and quantitative information about the conditions needed for the successful integration of international standards or about the kind of UN support needed to assist governments at the national level to implement a normative instrument.	Time-consuming and potentially expensive. Risky to generalize findings on the basis of only 2-3 cases. Need to focus on information-rich cases to justify the effort and cost.

²⁶ If the intervention involves strengthening the capacity of Member States, for example the capacity to revise legislation or to prepare development plans that integrate the norm, the literature review should examine international experience in this regard, including recent studies on nature of capacity development, such as that of Baser and Morgan (2008).

3. <i>Self-assessments</i>	Useful when assessing the knowledge and skills gained by individuals who have received training, coaching or mentoring as part of the normative work. Ideally, should be done before and after the training, coaching or mentoring.	Often positively biased, so evaluators should obtain a second line of evidence, such as from the supervisors of those who participated in training.
4. <i>Benchmarking</i>	Potentially helpful in comparing each government's policies, legislation or development plans against a benchmark or ideal example that fits all UN criteria for the norm or standard.	Need to recognize the extenuating circumstances in some countries, such as weak legislative drafting capacity or lack of political will, which may make it difficult to achieve the benchmark.

76. Where the evaluation is focused on the experience of integrating the standard in one or a few countries, it could be helpful for the evaluator to prepare a chronology charting the major interventions and the steps taken by each country to bring their policies, laws and regulations in line with the standard. The chronology²⁷ could help the evaluator to correlate data pertaining to the changes that have occurred across the normative interventions.

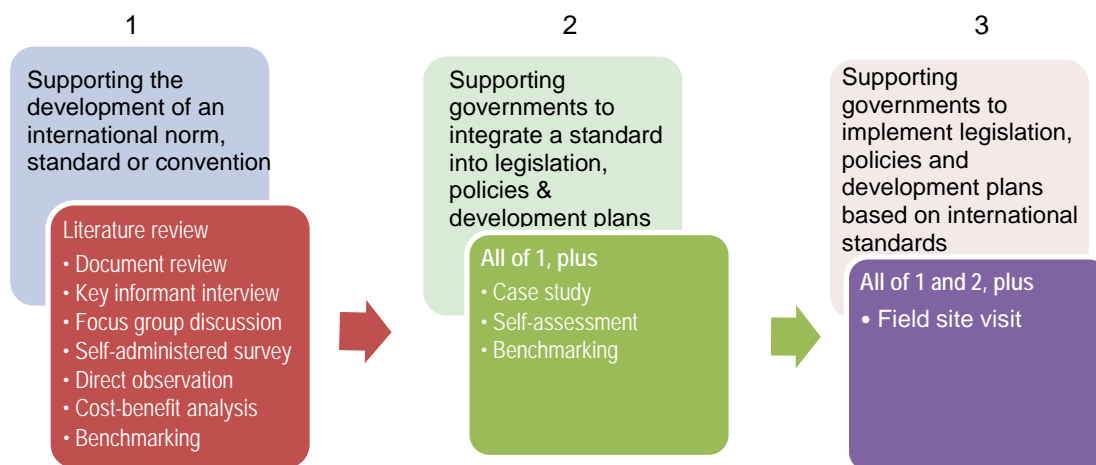
Table 11. Evaluation Methods for Initiatives to Assist Governments to Implement Legislation, Policies and Development Plans Based on an International Standard

Methodology	Pros	Cons
1. <i>All in Tables 9 and 10 may apply</i>		
2. <i>Field site visit</i>	Field site visits give the evaluator an opportunity to hear the perspectives of the ultimate beneficiaries/rights holders. They can help the evaluator to validate information obtained from other sources and other methods, and they can help the evaluator to gain a first-hand understanding of the constraints, such as lack of funds or technical skills, that national governments or institutions face when attempting to implement international standards. Most UN evaluations currently include field visits as part of their methodology.	Expensive when many countries are involved.

²⁷ An example of the use of a chronology and a related benchmarking (see item 4, Table 10) can be seen in the 2007 FAO's Evaluation of the Workings of the International Plant Protection Convention and its Institutional Arrangements, available at: <<ftp://ftp.fao.org/docrep/fao/meeting/012/k0233e02.pdf>>

77. Figure 3 provides a quick visual summary of the most common methodologies for each of the three types of evaluations of normative work in the UN system.

Figure 3. Common Methodologies by Type of Evaluation



78. The evaluator is not confined to these common methods; the special methods outlined in Table 12 below can be used effectively to evaluate normative work under the circumstances described.

Table 12. Evaluation Methods for Use in Special Circumstances

<p>Outcome mapping²⁸</p>	<p>Outcome mapping is a measurement system designed by the International Development Research Centre (IDRC) in Canada. Rather than assessing the products of an intervention (e.g. policy change), it focuses mainly on changes in behaviours of the people and organizations affected by the intervention. Outcome mapping establishes a vision of the human, social, and environmental betterment to which the project hopes to contribute and then focuses monitoring and evaluation on factors and actors within its sphere of influence. Outcome mapping pays less attention to the intervention's actual progress and more to the intervention's influence (both deliberate and unintended) during implementation. Best suited to the monitoring and evaluation of initiatives supporting governments to integrate an international standard into legislation, policies and development plans, and to assist governments to implement legislation, policies and development plans based on an international standard.</p>
-------------------------------------	---

²⁸ For more information on Outcome Mapping, see: www.idrc.ca/EN/Resources/Publications/Pages/ArticleDetails.aspx?PublicationID=1004

Outcome harvesting ²⁹	Outcome harvesting can be applied in evaluations of normative work where there are no pre-determined outcomes in the form of a logic model or theory of change and no performance monitoring information against pre-set indicators. The method works in reverse order from standard evaluations: the evaluator begins by ‘harvesting’ outcomes – both positive and negative – from reports, interviews and other sources and then works backwards to determine the extent to which the project or initiative contributed to the outcomes. The harvested outcomes go through a culling process where knowledgeable, independent sources validate and substantiate them. The methodology is suited to highly complex normative environments where cause and effect are unclear. Can be time consuming and requires a high degree of participation. Most applicable for evaluating the 2 nd and 3 rd type of normative work.
Process tracing	Process tracing or process induction is a method aimed at generating and analysing qualitative data on the causal mechanisms or processes, events, actions, expectations, and other intervening variables that link hypothesized causes to observed effects. ³⁰ Oxfam UK has used process tracing successfully to evaluate policy influence in South Africa and Bolivia (Oxfam UK, 2012). Particularly well suited to evaluating normative work supporting governments to integrate an international standard into legislation, policies and development plans.
Systems analysis	Systems analysis is the study of sets of interacting entities and the relationships among them. The method is relatively new to evaluation and can be helpful in analysing complex systems, such as those pertaining to governance, trade, and various social systems where there are multiple variables, interlinking relationships and communication that can effect normative outcomes. Suitable for all three categories of normative work.
Most Significant Change	Most Significant Change (MSC) is a form of participatory monitoring and evaluation that is story-based and qualitative. It can be used in evaluating human rights and democracy assistance interventions where outcomes are difficult to predict and where there may be no predetermined indicators (UNIDO and UNEP, 2008). The process involves the collection of significant change stories emanating from the field, and the systematic selection of the most significant of these stories by panels of designated stakeholders or staff. MSC can identify both positive and negative changes brought about by an intervention. MSC can help stakeholders focus their activities on the drivers of change. Best suited to assessing the outcomes and impact of national government efforts to apply policies, laws and development plans based on international norms at the community level.
Appreciative Inquiry ³¹	A methodology first used in organizational development and now in evaluation that builds on an organization’s assets rather than its shortcomings. Appreciative

²⁹ The Swiss Agency for Development and Cooperation used outcome harvesting for an evaluation of the BioNET Global Programme 2007-2011, which supports a global environmental voluntary network that promotes a standard taxonomy linked to the implementation of the Convention on Biological Diversity. See: www.outcomemapping.ca/resource/resource.php?id=374 for more information on outcome harvesting.

³⁰ See: Collier, D. (2011). ‘Understanding Process Tracing’. *PS: Political Science and Politics*, 44:4. Berkeley: University of California, pp. 823-830. Retrieved from polisci.berkeley.edu/people/faculty/CollierD/Understanding%20Process%20Tracing.pdf

³¹ For more information on AI, see: appreciativeinquiry.case.edu/

	<p>Inquiry (AI) asks participants to interview each other using questions that generate examples, stories and metaphors about positive aspects of organizational life. Participants analyse the results in groups looking for the themes and topics that can become the foundation for positive organizational development. AI can be applied in participatory evaluations that frame evaluation questions in a positive light and that involve participants determining practical action to produce a desired positive state. AI can be used in evaluating all three categories normative work.</p>
<p>Photovoice³²</p>	<p>Photovoice is a method that can be applied in evaluations where the perspectives of youth, marginalized or disempowered groups are important. Participants are asked to represent their point of view by taking photographs, discussing them together, developing narratives to go with their photos, and conducting outreach or other action. This method could, for example, be used to understand the perspectives of the beneficiaries of normative work who may not respond to surveys and interviews. Best suited to evaluating at the community level the outcomes and impact of policies, laws and development plans based on international norms.</p>

79. Many other methods, tools and resources are available to evaluation managers for evaluations of particular types of normative work. For example, evaluation managers could use Participatory Rural Appraisal³³ if called upon to evaluate normative interventions at the community level. For national-level evaluations, some worldwide longitudinal data sets may be helpful, particularly where the normative work involves governance issues. Freedom House, Polity and Transparency International are among the organizations that have such data and the UN itself has many potentially useful indexes, UNDP’s Human Development Index, among them. The caveat in using these is causality: although a country’s performance during the course of the UN support may have improved according to the index, the improvement is unlikely to have been caused by the UN support alone.
80. A number of special methods are available for assessing advocacy initiatives. Among them are media tracking, policy mapping, network mapping, and structured bellwether interviews³⁴. These and others are described in Coffman’s (2009) *User’s Guide to Advocacy Evaluation Planning*. Coffman argues that evaluating advocacy requires a non-traditional mind-set because successful advocacy work is constantly in flux, adjusting to changes in the environment. In this area of normative work, planned outcomes are likely to shift quickly.

³² For more information, see: en.wikipedia.org/wiki/Photovoice

³³ Participatory Rural Appraisal is a participatory methodology with a set of participatory tools often used in rural development to ensure that the knowledge and opinions of local stakeholders are prominent in the planning, management and evaluation of development programmes.

³⁴ Bellwethers are gauges of future trends or predictors of future events. In this context, bellwethers are knowledgeable and innovative thought leaders whose opinions about policy issues carry substantial weight and predictive value. With this methodology bellwethers are informed before the interview about what the interview will generally cover but are not given specific details to ensure that their responses are unprompted.

For Coffman, “The real challenge is assessing what happens along the way and what can be learned from that journey” (Coffman, 2007, p. 32). As noted in Chapter 1, this may call for a nontraditional approach to evaluation, such as developmental evaluation.

81. *Evaluating Democracy Support* (Burnell, 2007) chronicles the many challenges involved in evaluating programmes supporting democratization, rule of law, anticorruption and human rights. Published by the International Institute for Democracy and Electoral Assistance and the Swedish International Development Cooperation Agency, the paper makes a strong case for participatory approaches and for a mix of quantitative and qualitative measures.
82. Evaluating normative products and services such as publications, training courses, databases, knowledge networks and so on may call for special methods in combination with others noted in Tables 9, 10 and 11. For example, peer reviews can be helpful in assessing the technical quality of normative materials. Web analytics are helpful in measuring the use of web-based resources over time. When evaluating normative products, evaluation managers may wish to assess them against internationally recognized practices, such as conducting needs assessments, involving stakeholders in their design, targeting the materials for distinct audiences, field testing the materials before using them, integrating gender equality and human rights principles, and ensuring adequate budgets for their promotion and distribution.³⁵

Task 2.6: Selecting the evaluation team

83. The selection of the evaluation team is one of the evaluation manager’s most important tasks. Although much depends on the evaluation’s focus, scope and budget, the following general points should be kept in mind.
84. All members of the team should be familiar with the norms, standards or conventions at the centre of the intervention. Subject specialists may be needed depending on the nature of the normative work. If, for example, the work is highly technical, as in intellectual property, international law, marine biology, air safety or trade-related matters, a specialist can be immensely helpful. If the normative work involves advocacy, there should be someone on the team with expertise in policy change, media relations, development communication and, perhaps, social media. Most important, team members, including the team leader, should have strong backgrounds in evaluation and wide-ranging experience. Clear, concise writing is an often overlooked but critically important skill for evaluation. An evaluation can fall apart if the team leader and those contributing to the evaluation report are weak writers. The more complex the evaluation – and evaluations of normative work tend to be complex – the more essential it is to have strong writing and analytical skills.
85. Most evaluations of UN normative work reviewed for this handbook used hybrid teams with internal and external team members. That said, sometimes it may be advantageous to use only internal evaluators. Other times, as with some high-level, agency-wide evaluations of

³⁵ The assessment template in Annex 8 of the evaluation of FAO’s capacity development activities in Africa (FAO, 2010) captures many these assessment criteria.

normative work or when highly specialized expertise is needed, external experts might be required.

86. The evaluation manager must be acutely attuned to gender and human rights, as well as cultural and political nuances, when composing the evaluation team. Teams should ideally include a gender equality and human rights specialist and they should have an appropriate balance of women and men. Local consultants can complement the work of international consultants whenever field missions are conducted. Local knowledge and local language expertise are critically important because most evaluations of normative work in the UN system require an understanding of the intervention's social, organizational and institutional contexts. In circumstances calling for participatory models of evaluation, government and other stakeholders may be involved in gathering and analysing data.
87. Interpreters and translators are indispensable in countries where there are significant language barriers. The best professional interpreters and translators that could be found should be hired, staying clear of resources provided by stakeholder groups so as to avoid undue filtering of information. Once recruited, an interpreter becomes an integral part of the team. The interpreter needs time at the beginning of the evaluation to learn about the normative context and to become familiar with the intervention's technical terminology. In order to facilitate the work of the interpreter, the evaluation manager needs coach other team members in the art and discipline of short-sentence interviewing.
88. The evaluation manager should keep the team to a manageable size in order to avoid the inevitable problems that plague large teams and to keep the cost of the evaluation in check. This can be accomplished by selecting team members with multiple, complementary skills and experience. The role of each team member needs to be well defined, as should the lines of accountability. The evaluation manager should set aside time at the start of the evaluation to ensure team members are briefed on evaluation ethics.

Task 2.7: Preparing the evaluation terms of reference

89. The terms of reference (ToR) for evaluating normative work should follow the same pattern as the ToR for evaluating development initiatives. It should contain, at a minimum, the evaluation purpose, use, scope, focus, key questions, cross-cutting themes, design, methodology, timing and background information on the intervention to be evaluated, along with a theory of change. The ToR should also describe the roles and responsibilities of the main participants in the evaluation. The document should describe the norms, standards or conventions in some detail. Often, evaluation ToRs refer to the UN ethical guidelines and standards to be followed. Some ToRs state the expertise required within the evaluation team. The ToR should briefly explain to readers why the evaluation is taking place, what it will focus on and who will participate in it. It does not usually contain the design matrix, which is the basis of the plan for conducting the evaluation. Ideally, the ToR should be no more than 10 pages and written in plain language for broad distribution, in which case the budget

Team members involved in a FAO country evaluation, which included an assessment of FAO's normative work, received a good deal of light-hearted criticism regarding the evaluation ToR, which stretched to about 30 pages. Some stakeholders asked that the final report be kept shorter than the ToR.

details are excluded.

90. The ToR should remain in draft until the evaluation begins. It is good practice to tweak the questions and methodology in light of new information received as the evaluation gets under way. Sometimes consultants recruited for the evaluation team or members of the reference group bring new perspectives that warrant changes to the ToR.

Box 7. Lessons in Preparing Evaluations of Normative Work

The involvement of internal stakeholders and, to the extent possible, relevant external stakeholders, throughout the evaluation increases the perceived relevance, and stakeholders' ownership of the evaluation (UNEG, 2010).

"It is important to strike an appropriate balance between promoting the ownership of evaluation findings and recommendations without compromising the independence of evaluation, which can be aided by clearly defining the role of such a reference or consultative group prior to commencing work" (UNEG, 2010, p. 4).

Crafting indicators can be challenging, particularly so for some types of normative work. Experience shows that this task is best done in a group which includes subject matter specialists and someone with a strong results-based management background and experience preparing indicators.

Sex ratios are an insufficient indicator for gender equality. It is not numbers, but the substance of those numbers that counts – e.g. representation versus participation (ILO, 2012b).

Focus the evaluation on what is most important and what is feasible. In evaluations of normative work, the higher one climbs the results chain, the more difficult the causal analysis becomes.

Not everything that counts can be counted and not everything that can be counted counts. (Albert Einstein)

Rigorous thinking supersedes rigorous methods. (M. Patton, 2011)

Box 8. Quality Check – Preparing Evaluations of Normative Work

- ✓ Has the purpose of the evaluation of normative work been clearly spelled out and the use of the evaluation data determined?
- ✓ Have several key questions been identified as the focus of the evaluation of normative work?
- ✓ Have the various design options been considered and an appropriate design selected?
- ✓ Has a stakeholder analysis been carried out and an outreach strategy prepared?
- ✓ Have stakeholders (women and men) been identified to participate in the evaluation of normative work and their roles and responsibilities described?
- ✓ Have key data sources been identified?
- ✓ Have various methods been considered and a combination of the most appropriate methods selected in light of the evaluation purpose and context?
- ✓ Have appropriate, gender-sensitive indicators been selected for each outcome and impact?
- ✓ Do the design and methods selected take into account the key evaluation questions?
- ✓ Do the design and methods allow for a thorough analysis of cross-cutting issues, including human rights and gender equality?
- ✓ Have solutions been explored for the challenges anticipated?

Useful resources

Burnell, P. (ed.) (2007). *Evaluating Democracy Support. Methods and Experiences*. Stockholm: International Institute for Democracy and Electoral Assistance and Swedish International Development Cooperation Agency.

Canadian International Development Agency (1997). *Guide to Gender-Sensitive Indicators*. Available at: <www.acdi-cida.gc.ca/>

Coffman, J. (2009). *User's Guide to Advocacy Evaluation Planning*. Cambridge, Massachusetts: Harvard Family Research Project.

Coffman, J. (2007). "What's Different About Evaluating Advocacy and Policy Change?" *The Evaluation Exchange Vol. XIII, No. 1*.

Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. Thousand Oaks, California: Sage.

Reisman, J., Gienapp, A., and Stachowiak, S. (2007). *A Guide to Measuring Advocacy and Policy*. Baltimore: Annie E. Casey Foundation.

Stern, E., Stame, N., Mayne, J., Forss, K, Davies R., and Befani, B. (2012). “Broadening the Range of Designs and Methods for Impact Evaluations.” *Department for International Development Working Paper No. 38*. London: Department for International Development.

UNAIDS (2011). *An Introduction to Indicators*. Available at:
<www.unaids.org/en/media/unaids/contentassets/documents/>

UNEG (2005). *Norms and Standards for Evaluation in the UN System*. Available at:
<www.uneval.org>

Wilson-Grau, R. and Britt, H. (2012). *Outcome Harvesting*. Cairo: Ford Foundation, Middle East and North Africa Office. Available at:
<www.outcomemapping.ca/resource/resource.php?id=374>

World Bank (2010). *Cost-benefit Analysis in World Bank Projects: Overview*. New York: Independent Evaluation Group, World Bank. Available at:
<siteresources.worldbank.org/INTOED/Resources/cba_overview.pdf>

International Initiative for Impact Evaluation: Large collection of useful reference materials, tool-kits, downloadable documents and other resources on methodology and best practice for impact evaluation: <www.3ieimpact.org/>

NONIE: A network of over a 100 international evaluation offices, including the UN, The World Bank, other development organizations, developing countries and various regional and global organizations.

The Communication Initiative Network: Comprehensive information on designing, implementing and evaluating communication and advocacy campaigns: <www.comminet.com>

Chapter 3. Collecting and Analysing Data

Task 3.1: Collecting data

91. The handbook turns next to some of the common challenges associated with evaluating normative work, which in most cases will involve much more qualitative than quantitative assessment.

“Most of the advocacy that takes place at private meetings, receptions and dinners away from the limelight is undocumented. Much of the high-level advocacy work is as well. Some events are described briefly in reports and in media coverage, but the details of the tactics used to advocate are not. For example, there is relatively little documentation describing the years of intensive lobbying that led to Pakistan’s Juvenile Justice Ordinance.... The absence of such data impedes organizational learning.”

- *Evaluation of the Government of Pakistan-UNICEF Advocacy and Social Mobilization Strategy* (Stiles, 2002, p. 17)

92. Often the greatest challenge is to find reliable data for the assessment of outcomes (behavioural and institutional changes) derived from normative work. The evaluation team will have to piece together information about the situation before and after the intervention from a document review, interviewing and other methods. Obvious sources of information include the officers responsible for the intervention, the key implementing partners and other key stakeholders. Not to be overlooked are the UN officials who have made formal and informal representations. Some of the best insights – often the golden nuggets of perception – may come from unlikely sources: junior field staff, retired government officials, business leaders, journalists, and even drivers.

93. It is essential to gather the views of both women and men because their needs are distinct and the impact of normative interventions on them is often different. Several UN evaluations (for example, ILO, 2012a) reviewed for this handbook found that reinforcing stereotype roles for women and men in normative work reduces economic opportunities for women. Evaluations have found that normative work which has not included a rigorous gender analysis may have negative outcomes for women and children, even when women and children are targeted as key beneficiaries (UNICEF, 2008).

“External issues should be thoroughly analysed to understand how they interact with a specific project. This is the most difficult component of the evaluation: external influences may be numerous and hard to foresee, particularly in democracy programmes of a political nature. Also, the lack of theoretical consensus among experts, academics and practitioners about the causal relationship between variables and their wider political effects complicates the analysis.”

- Burnell (ed.) (2007), p. 107

94. Analysing the influence of external issues on the outcomes of normative work is essential, but is one of the most difficult aspects of evaluation. Some stakeholders may be reluctant to acknowledge that the UN’s work influenced their decisions to adopt international norms despite evidence to the contrary. The evaluator has to take into account the different perspectives of reality by triangulating data from multiple sources and from multiple lines of inquiry when making judgements as to the merits of an intervention. Occasionally, the evaluator may find a tell-tale sign of influence such as reference to a UN norm in a policy statement or the widespread use of a

UN analytical tool. The evaluators must also look beyond the prescribed outcomes and indicators because the unintended results may be more important than the intended outcomes.

95. People are at the centre of evaluations and dealing with them is often the greatest challenge in normative evaluations of a qualitative nature. Often in the evaluation of normative work, the evaluators must deal with expectations, power relationships, personal politics and fixed notions about the role of evaluation. A review of evaluations that have gone wrong suggests that good communication is an essential ingredient for success (Gervais, 2010). Communicating about international standards is in itself challenging because people have varying interpretations of the standards and at times may resent having to comply with them. Explaining the purpose of the evaluation and how the information will be used is good practice.
96. What should the evaluation manager do if the evaluation team is not receiving answers to the evaluation questions? Although this rarely happens with good planning, the evaluation manager should have a contingency plan. Sometimes the best strategy is to regroup after the first set of interviews or after the first country case study, and to see if any lessons can be learned and if the evaluation team can find a way to refocus the evaluation. If the team cannot, it is best to return to the reference group for guidance and make necessary changes to the evaluation design and methods.³⁶

Task 3.2: Analysing data

97. The nature of the UN's normative work possess challenges for data analysis, as exemplified in a report on an evaluation of UNIFEM programming to facilitate implementation of CEDAW in Southeast Asia:

While our analysis has shown that there is evidence of progress against the related outputs and against the individual outcome level indicators, the question remains whether and how these achievements 'add up' to demonstrating stronger political will and generated/strengthened commitment to CEDAW implementation. (UNIFEM, 2008, p. 9)

98. What does it 'add up to?' That is the evaluator's conundrum. This issue is particularly important in the evaluation of normative work where it is the responsibility of governments as duty bearers to implement international norms, standards and conventions, while UN organizations play a facilitation and/or a supportive role. To resolve the issue the evaluator needs to distinguish between *attribution* and *contribution*, as in the following:

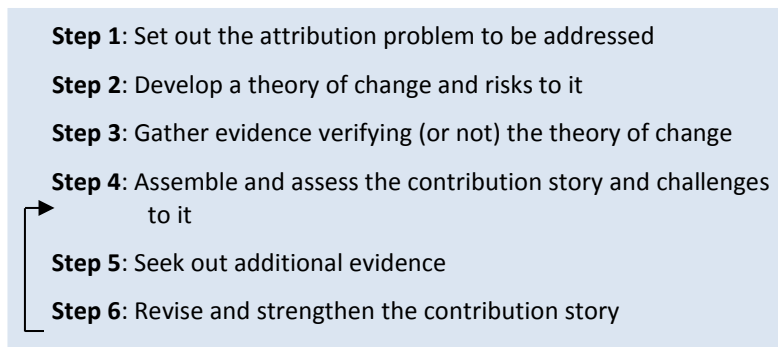
Attribution involves a causal claim about the intervention as the cause of the impact, and measurement of how much of the impact can be linked to the intervention. This contrasts with 'contribution', which makes a causal claim about whether and how an intervention has contributed to an observed impact. This usually involves verifying a theory of change

³⁶ Recent literature on evaluation may provide guidance in dealing with what to do when an evaluation goes wrong. See for example, the special issue of the *Canadian Journal of Program Evaluation*, Vol. 25, No. 3, 2010.

in a way that takes account of other influencing factors and thus reduces uncertainty about the contribution the intervention is making (Stern et al., 2012).

99. As one moves up the logic chain to impact, it is often more accurate to speak of contribution or influence rather than of direct cause and effect (UNEG, 2012). Where UN normative work has influenced policy or legislation, “the best that can be expected is plausible interpretation rather than firm ‘proof’” (Stern et al., 2012, p. 34).
100. Mayne’s (2011) work on contribution analysis provides evaluators of the UN’s normative work with a practical approach to making causal claims and reducing the uncertainty of causality where data are limited and where experimental or quasi-experimental designs are not possible. The contribution analysis is a six-step process (see Box 9) of verifying, and revising where needed, the intervention’s theory of change, paying close attention to factors that influence the outcomes and impact.

Box 9. Steps in a Contribution Analysis



Source: Adapted from Mayne, 2011.

101. Contribution analysis is based on the premise that an intervention is part of a broader causal package. When the intervention is working, it is an essential element of the package, and when the package is working, it is sufficient to bring about the intended change. It follows that the intervention on its own is insufficient to bring about the change, but a necessary component of the causal package, in which case it is a *contributory* cause. The intervention is deemed to be strongest when it is the ‘trigger’ for the change. Where there is sufficient evidence to conclude this, the evaluator can state that comprehensive causal package was sufficient to cause the collective results and the intervention actions were the trigger that made it happen (Mayne, 2011).
102. Contribution analysis forces the evaluator to seek out the direct and indirect factors that influence change in complex programming environments common to the UN’s normative work and to make the underlying assumptions explicit. Not only is contribution analysis useful, it is sufficiently rigorous to instil confidence in the evaluators’ conclusions regarding the effectiveness of normative interventions.
103. Evaluators need to be vigilant about applying human rights and gender equality lenses to their data analysis. This applies to the evaluation of all normative work, not just the work that

has an explicit human rights or gender equality focus. The analysis needs to be applied at three levels: the original design of the intervention to see whether human rights and gender equality concerns had been considered; the extent to which the intervention then actually applied human rights and gender equality principles and/or a human rights-based approach; and analysis of the outcomes and impact of an intervention to determine the extent to which those outcomes and impact are aligned with human rights and principles and how they may affect women and men differently. Assistance from human rights and gender specialists can be invaluable in making these determinations. For more detailed information, UNEG's *Integrating Human Rights and Gender Equality in Evaluation*, is a helpful guide. Additional resources are listed at the end of this chapter.

104. A cautionary note about targets: some evaluation reports reviewed for this handbook based their assessments largely on the extent to which targets were met. None of these took the trouble to assess whether the targets were realistic in the first place or if other, perhaps more important, priorities emerged during the course of the normative work. Targets at best tell only part of the performance story and, particularly in normative work, targets are subject to change at the drop of a hat. Although targets should be adjusted frequently, managers and donors are often reluctant to change them once set. For these reasons, the evaluator should explore them carefully before jumping to conclusions.
105. The use of matrices to categorize information can assist data analysis.³⁷ A reflective workshop can also be a useful participatory way to analyse data. Some evaluations of normative work have held a workshop immediately after collecting data wherein members of the evaluation team return to the ToR and examine each evaluation question in detail. The workshop needs to be well facilitated to ensure full participation and critical reflection. The use of index cards to help identify and organize key points visually is helpful. This method is particularly useful in sorting through massive amounts of data to identify what is really important. It is also a good way to prepare the presentation of the team's preliminary findings to the reference group or to stakeholders at the end of an evaluation mission.

Task 3.3: Identifying lessons and best-fit practices

106. In the course of analysing their data, members of the evaluation team should take the time to identify lessons and best-fit practices. The terms 'lessons' and 'best-fit practices' are used advisedly. Lessons do not become 'lessons learned' until they are actually implemented. Some organizations repeat the same mistakes despite claiming to have learned lessons. Best-fit practices are more appropriate than 'best practices' because not all practices, however effective in one setting, fit all others. These nuances are particularly applicable to the evaluation of normative work because of the varied contexts in which the work takes place. It follows that the lessons at the end of each chapter in this handbook may not apply in every evaluation of normative work.

³⁷ An independent evaluation of the UNIDO and UNEP Cleaner Production Programme in 2008 used matrices to rate performance in relation to policy advice across 18 countries. The evaluators organized data on outputs, outcomes and impacts in three categories: scale of results, type of evidence and strength of evidence (UNIDO and UNEP, 2008, pp. 92-93).

107. Identifying lessons and best-fit practices is often challenging: what turns up in some evaluation reports is either insufficiently grounded in evidence or written in the form of recommendations. Lessons usually come from qualitative inquiry that asks: “Why did this happen and what can be learned from it?” Best-fit practices also come largely from qualitative inquiry that asks: “Why did this work well and can it be applied in similar normative work under like conditions?” The task of identifying lessons and best-fit practices should be shared within the evaluation team. It is important to set aside at least a full day for this because it often requires considerable reflection and discussion. Input from key stakeholders can be helpful. Those most closely associated with the intervention could be asked already during the scoping stage of the evaluation to self-identify lessons and best-fit practices, which the evaluation team could then look at more closely during the data collection phase (and not stumble into them mid-process). At the very least, stakeholders should be given the opportunity to review and comment on the lessons and best-fit practices once drafted.
108. Since most lessons are learned from trial and error, the evaluators need to probe for this information during interviews because managers are likely to be hesitant in reporting work that goes awry. Some lessons and best-fit practices may come from literature reviews. For example, there are invaluable lessons and examples of best-fit practices in the literature on consensus building, advocacy, behavioural change communication, networking, capacity development and policy influence.

Box 10. Lessons in Collecting and Analysing Data

Attention to potential unexpected effects is particularly important for evaluations of normative work, such as where a Convention has been in place for some time with limited follow-up, and given that in complex environments almost anything can happen that may not be directly under the control of the agency spearheading the normative undertakings.

One should also identify multiple and sometimes simultaneous casual pathways. For example, advocacy may involve direct engagement with government, and frequently at the high political level simultaneously with developing support within the administration. Often it involves community mobilization. It can involve direct and/or indirect action, both formal – a media campaign – and informal, including discussions with high-level officials.

Cooperation from partners is needed for data collection.

“Obtaining information from both women and men may increase the cost and time of data collection. This needs to be considered but is often neglected during the planning and budgeting of the M&E [monitoring and evaluation] exercise.” (ILO, 2012b, p. 9)

“Most IE [Impact Evaluation] questions about the effects of an intervention are best answered through a ‘contributory’ lens: successful programmes are part of a sufficient ‘causal package’ – they do not work in isolation.” (Stern *et al.*, 2012, p. 36)

Box 11. Quality Check – Collecting and Analysing Data

- ✓ Have baseline data been identified or recreated?
- ✓ Has contextual information been thoroughly examined?
- ✓ Have the data been recorded carefully and in ways that minimize bias?
- ✓ Have the evaluation team members participated in reviewing and analysing the data collected?
- ✓ Have the evaluators used triangulation in their data analysis?
- ✓ Has a rigorous contribution analysis been carried out that identifies the causal package, assesses whether the normative work was sufficient and necessary to bring about the change identified and whether the normative work triggered the change?
- ✓ Have the key findings been presented to the key stakeholders and validated by them?

Useful Resources

Equitas – International Centre for Human Rights Education and the Office of the United Nations High Commissioner for Human Rights. (2011). *Evaluating Human Rights Training Activities A Handbook for Human Rights Educators*. Professional Training Series No. 18. Available at: www.ohchr.org/Documents/Publications/EvaluationHandbookPT18.pdf

Burnell, P. (ed) (2007). *Evaluating Democracy Support. Methods and Experiences*. Stockholm: International Institute for Democracy and Electoral Assistance and Swedish International Development Cooperation Agency. Available at: www.idea.int/publications/evaluating_democracy_support/index.cfm

International Council on Human Rights Policy (2012). “No Perfect Measure: Rethinking Evaluation and Assessment of Human Rights Work. A Report of a Workshop.” Geneva: ICHRP.

International Institute for Democracy and Electoral Assistance (2002). *Handbook on Democracy Assessment*. The Hague: Kluwer Law. Available at: www.idea.int/publications/sod/democracy_assessment.cfm

International Labour Organization (2012). *Integrating Gender Equality in Monitoring and Evaluation of Projects. Guidance Note 4*. Geneva: International Labour Organization, Evaluation Unit.

OECD/DAC. 2010. *Quality Standards for Development Evaluation*. Available at: www.oecd.org/dataoecd/55/0/44798177.pdf

Patton, M. (2001). *Qualitative Research & Evaluation Methods, 3rd edition*. Thousand Oaks, California: Sage.

Mayne, J. (ed) (2012). “Contribution Analysis.” *Evaluation*, Special Issue, 18(3).

Mayne, J. (2011) “Contribution Analysis: Addressing Cause and Effect” in *Evaluating the Complex*, K. Forss, M. Marra and R. Schwartz (Eds.). Piscataway, New Jersey: Transaction Publishers.

White, H. and Phillips, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework*. New Delhi: International Initiative for Impact Evaluation.

UNWomen Gender Equality Portal: <http://genderevaluation.unwomen.org/Reports/Default.aspx>

Chapter 4. Preparing the Evaluation Report

Task 4.1: Preparing the evaluation report

109. A well-written final report is the jewel in the crown of a successful evaluation. The evaluation findings and recommendations are more likely to be taken seriously and acted upon if the report is well structured and pleasurable to read. Many useful guides are available to aid report writing, some of which are listed at the end of this chapter. The quality checklist at the chapter end is a handy reminder of key points from many of these guides.
110. Since no report-writing manuals have been written exclusively for the evaluation of normative work, the following are a few suggestions based on a review of evaluation reports for this handbook.
111. It is essential when describing normative work to provide context, such as the socio-political and cultural background and the history of the work. It is also important to note the contributions of other organizations, both local and international. Readers need to understand the environment within which the normative work took place, including its opportunities, challenges and major elements influencing the outcomes. They also need to know the driving forces behind any normative changes. This information will shed light on the contribution analysis and aid in forming recommendations. Excellent examples of contextual coverage can be found in UNICEF's DevInfo evaluation (UNICEF, 2009) and the 2010 Overall Performance Study of GEF.
112. The evaluation manager should ensure that the report captures not only the expected outcomes but also the unexpected ones, both positive and negative. Given the non-linear nature of some normative work, such as seeking agreement on a standard, advocacy, policy dialogue, capacity development and confidence building, there are bound to be a few unexpected changes, some of which may be more significant than those that were planned.
113. The conclusions and recommendations of an evaluation are critically important. Take the time to craft them carefully because if written poorly, they can become a flash point for criticism that could overshadow all that is good about the rest of the report. Remember that recommendations need to be directed at the people who have the responsibility for implementing them. In most evaluations of the UN's normative work that would be UN managers and their government counterparts and partners. When in doubt, test out the recommendations with members of the reference group or with other stakeholders before the report is completed. When it comes to formulating recommendations, do suggest ways to strengthen normative initiatives, but do not forget to recommend, where warranted, measures to improve the monitoring and evaluation of future normative work. Most of the ILO evaluations reviewed for the handbook, for example, do a good job of crafting and directing recommendations.
114. Not all evaluations need recommendations. For example, the impact evaluation of UNIDO's work on standards, metrology, testing and quality in Sri Lanka (UNIDO, 2010) contained no recommendations, but rather focused on a series of lessons to guide future work.

115. It is also important to acknowledge the limitations of the evaluation, particularly when baseline data are unavailable and monitoring information is limited. Constraints faced in carrying out the evaluation should appear in the methodology section of the report along with a description of what was done to mitigate the limitations.

Box 12. Lessons – Preparing the Evaluation Report

A report that is written in plain language and that is free of jargon goes a long way towards ensuring that it is read.

The best way to ensure that a report is read is to leave out the parts that few people will want to read.

“I often reach the right level of spontaneity in my writing by the fifth draft.” — John Kenneth Galbraith

“Brevity is the soul of wit.” — William Shakespeare, *Hamlet*

Box 13. Quality Check – Preparing the Evaluation Report

- ✓ Is there sufficient evidence (from multiple sources and multiple lines of inquiry) to support each major finding?
- ✓ Are the conclusions well substantiated by the evidence presented and logically connected to evaluation findings?
- ✓ Are the limitations of the evaluation and its methodology described?
- ✓ Are the evaluative judgements reasonable and justified given the evidence presented and the normative context?
- ✓ Are the judgements in the report balanced and fair?
- ✓ Are the recommendations warranted, targeted and actionable?
- ✓ Do the recommendations reflect an understanding of the organization and potential constraints for follow up?
- ✓ Does the report explain how gender equality and a human rights-based approach were integrated in the evaluation and are these cross-cutting themes reflected in the findings, conclusions and recommendations?
- ✓ Does the report use gender sensitive and human rights-based language throughout, including data disaggregated by sex, age and disability?
- ✓ Does the report respect privacy?
- ✓ Is there an executive summary that succinctly encapsulates the rationale for the evaluation, its methodology and its major findings, conclusions and recommendations?

Useful Resources

Art and Architecture of Writing Evaluation Reports: Treasury Board of Canada Secretariat (2004). Available at: www.tbs-sct.gc.ca/cee/career-carriere/workshops-ateliers/aawer-amrre-eng.pdf

UNEG Quality Checklist for Evaluation Reports (2010). Available at: www.unevaluation.org/papersandpubs/documentdetail.jsp?doc_id=607

UNEG Standards for Evaluation in the UN System (2005). Available at: www.unevaluation.org/uneegstandards

Chapter 5. Following Up on the Evaluation

Task 5.1: Disseminating the evaluation report

116. Although each UN organization has its own policy on disseminating evaluation reports, the evaluation manager can take steps to enhance dissemination and use. Dissemination and use need to be thought through when planning the evaluation, not left until the end. Ideally, they should be addressed in the evaluation's outreach strategy as noted in Chapter 2. Early planning is particularly important for evaluations of normative work that have learning as a primary objective. Dissemination should not be confused with use; dissemination can facilitate use, but is not the same as making decisions, improving programming, or generating knowledge (Patton, 2008). It is also important to keep in mind the various audiences for the evaluation report. Although UN managers are often the primary audience, evaluations of normative work to assist governments to implement legislation, policies and development plans based on international norms and standards may have recommendations that are directed at government representatives, in which case the dissemination strategy should include a follow-up session with the government and perhaps some ongoing monitoring.
117. As noted in earlier chapters, by the time the report reaches senior managers, ideally they should already know what is in it through their involvement in the evaluation reference group and/or other debriefing modalities. It is good strategy to avoid the element of surprise in evaluation reports, particularly when the content may be controversial.
118. UNEG standards recommend that organizations have an explicit disclosure policy that ensures "the transparent dissemination of evaluation results, including making reports broadly available to the Governing Bodies and the public, except in those cases where the reasonable protection and confidentiality of some stakeholders is required" (UNEG, 2005a, p. 4).
119. It is a sign of respect and consistent with recommended UNEG practice to provide those who participated in the evaluation with a copy of the report once it is approved. In the spirit of transparency, those who have taken time from their busy schedules to provide information ought to have an opportunity to see the results. Posting the final report on the organization's website is now common practice. Copies of the report can also be made available through UN regional and national offices. Where there are language barriers, the evaluation manager should translate the executive summary for wide distribution.
120. In circumstances where the evaluation report contains recommendations for improving normative work that is ongoing, and where local stakeholders are receptive to improvements, the evaluation manager should consider a follow-up mission to discuss the recommendations and their implementation. Although this is not standard practice, it has the potential to strengthen local ownership, learning and uptake of the evaluators' recommendations.

Task 5.2: Preparing a management response

121. UNEG encourages member organizations to have “appropriate evaluation follow-up mechanisms” to ensure that “evaluation recommendations are properly utilized and implemented in a timely fashion and that evaluation findings are linked to future activities” (UNEG, 2005a, p. 4). UNEG norms call for the Governing Body and management to respond to the evaluation recommendations, and for member organizations to have “systematic follow-up on the implementation of the evaluation recommendations that management and/or the Governing Bodies have accepted...” and “a periodic report on the status of implementation...” (UNEG, 2005b, p. 10).

Task 5.3: Using the evaluation

122. The formal mechanisms of the management response and follow up are aimed at ensuring that the recommendations are acted on, but they are no guarantee. Additional action is often needed to take full advantage of the knowledge gained in carrying out evaluations and the lessons generated in the process.
123. Given the likelihood of increased demand for evaluations of normative work in the UN system, evaluation managers should consider sharing their evaluation reports and developing a community of practice within the UN family. While information can initially be shared informally, if interest grows, it could expand into a knowledge network supported by an electronic repository. Some of what is learned could be incorporated in professional development programmes for evaluation managers in future. Sharing a database of consultants who are skilled in the evaluation of normative work could be another constructive offshoot.
124. Evaluations of normative work would be of interest to professional evaluation societies, of which there are now many. UN evaluation offices could link with them to strengthen professional practice.

Task 5.4: Evaluating the evaluation

125. This task, which is carried out infrequently, has considerable potential value. The evaluation manager could learn from seeking feedback on the evaluation some months or more afterwards. Were the findings used? What did stakeholders find valuable about the evaluation? What worked well and not so well regarding the evaluation design, methodology and implementation process? What can be learned from the experience to improve the evaluation of normative work in the future? Some of these questions could be answered quickly with information on file; others could be explored in a teleconference, a short survey, e-mail exchanges or in meetings. Such follow-up need not be expensive. Where warranted, evaluation managers could undertake peer reviews of evaluations of normative work as a way to increase learning and strengthen the way in which they are undertaken throughout the UN system.

Box 14. Lessons – Evaluation Follow-Up

Set the stage for use during the planning of the evaluation.

Identify a champion of change at senior level who can carry forward the lessons and recommendations from the evaluation of normative work.

Useful Resources

Costa, P. (2012). *Study on Communicating Evaluation Results*. Paris: OECD DevCom Network. Available at: www.oecd.org/dac/evaluation/StudyonCommunicatingEvaluationResults_FINAL_091112.pdf

UNEG (2005). *Norms for Evaluation in the UN System*. Available at: www.uneval.org/normsandstandards/index.jsp

UNEG (2005). *Standards for Evaluation in the UN System*. Available at: www.uneval.org/normsandstandards/index.jsp

Documents Reviewed

- Bamberger, M. and Rugh, J. (2008). "A Framework for Assessing the Quality, Conclusions, Validity and Utility of Evaluations." Paper presented at American Evaluation Association Conference, Baltimore, 2007.
- Baser, H. and Morgan, P. (2008). *Capacity, Change and Performance*. Maastricht: European Centre for Development Policy Management.
- Bolger, J. and Stiles, M. (2011). *End-of-project Evaluation of the Southeast Asia Regional Cooperation in Human Development (SEARCH) Project*. Gatineau: Southeast Asia Regional Programme, Canadian International Development Agency.
- Burnell, P. (ed.) (2007). *Evaluating Democracy Support. Methods and Experiences*. Stockholm: International Institute for Democracy and Electoral Assistance and Swedish International Development Cooperation Agency. Available at: www.idea.int/publications/evaluating_democracy_support/index.cfm
- Coffman, J. (2009). *User's Guide to Advocacy Evaluation Planning*. Cambridge, Massachusetts: Harvard Family Research Project.
- Coffman, J. (2007). "What's Different About Evaluating Advocacy and Policy Change?" *The Evaluation Exchange* Vol. XIII, No. 1.
- Collier, D. (2011). "Understanding Process Tracing." *PS: Political Science and Politics*, 44:4. Berkeley: University of California, pp. 823 -830. Available at: polisci.berkeley.edu/people/faculty/CollierD/Understanding%20Process%20Tracing.pdf
- Costa, P. (2012). *Study on Communicating Evaluation Results*. Paris: OECD DevCom Network.
- Coulibaly, K. (2010). *Better Work Baselineing and Impact Measurement Tools and Methodologies*. Geneva: International Labour Organization.
- Dozois, E., Langois, M. and Blanchet-Cohen, N. (2010). *DE.201.A Practitioner's Guide to Developmental Evaluation*. Montreal: The J. W. McConnell Family Foundation.
- El-Zanati and Associates (2009). *Djibouti Communication For Development Survey (C4DS)*. Djibouti: United Nations Children's Fund.
- Equitas – International Centre for Human Rights Education and the Office of the United Nations High Commissioner for Human Rights. (2011). *Evaluating Human Rights Training Activities A Handbook for Human Rights Educators*. Professional Training Series No. 18.
- Fetterman, D. (2001). *Foundations of Empowerment Evaluation*. Thousand Oaks, CA: Sage.
- Food and Agriculture Organization of the United Nations (2010). *Evaluation of FAO's Activities on Capacity Development in Africa*. PC104/5. Rome: Office of Evaluation.

- Food and Agriculture Organization of the United Nations (2009). *Independent Evaluation of FAO Corporate Objective B-1: “International instruments concerning food, agriculture, fisheries and forestry, and the production, safe use and fair exchange of agricultural, fishery and forestry goods.”* Rome: Office of Evaluation, FAO.
- Food and Agriculture Organization of the United Nations (2008). *Independent Evaluation of FAO’s Role and Work in Statistics.* Rome: Food and Agricultural Organization of the United Nations.
- Food and Agricultural Organization of the United Nations (n.d.). “Normative and Operational Activities.” PowerPoint Presentation. Rome: FAO.
- Funnell, S. and Rogers, P. (2011). *Purposeful Program Theory.* San Francisco: Jossey-Bass.
- Gardner, A. and Geierstanger, S. (2007). “Working With Logic Models to Evaluate Policy and Advocacy Program.” *The Evaluation Exchange XIII (1).*
- Gervais, M. (2010). “As I Recall—Or How to Take Advantage of Less-Than-Successful Evaluation Experiences.” *The Canadian Journal of Program Evaluation, 25(3).*
- Gill, A. and Gilbert, N., (2011). *Written Evidence to the House of Commons Committee on Science and Technology and references therein.*
- Global Environment Facility (2010). *OPS4: Progress Toward Impact.* Fourth Overall Performance Study of the GEF. Executive Version. Washington, DC: Global Environmental Facility Evaluation Office.
- GrantCraft (2006). *Mapping Change: Using Theory of Change to Guide Planning and Evaluation.*
- Guidance Note on Framing and Conducting Impact Evaluations of Normative Work by the UN System.
- International Council on Human Rights Policy (2012). “No Perfect Measure: Rethinking Evaluation and Assessment of Human Rights Work. A Report of a Workshop.” Geneva: ICHRP.
- International Institute for Democracy and Electoral Assistance (2002). *Handbook on Democracy Assessment.* The Hague: Kluwer Law.
- International Labour Organization (2012a). *Time-Bound Programme - Urban Informal Economic Programme on Prevention and Elimination of the Worst Forms of Child Labour in the Urban Informal Economy of Dhaka Metropolitan Area. Final evaluation.* Geneva: International Labour Office, International Programme on the Elimination of Child Labour.
- International Labour Organization (2012b). *Integrating Gender Equality in Monitoring and Evaluation of Projects. Guidance Note 4.* Geneva: International Labour Organization, Evaluation Unit.

- International Labour Organization (2012c). *Promoting freedom of association and collective bargaining rights in the rural and export processing sectors: August 2009-December 2011*. Independent Final Evaluation Report. Geneva: International Labour Office.
- International Labour Organization (2012d). *Support to the Indonesian Time-Bound Programme of the Elimination of Worst Forms of Child Labour-Phase II. Final Evaluation*. Geneva: International Labour Office, International Programme on the Elimination of Child Labour.
- International Labour Organization (2012e). *Independent Final Evaluation Report of Project: Combating Forced Labour and Trafficking of Indonesian Migrant Workers, Phase II*. Geneva: Evaluation Unit.
- International Labour Organization (2011). *Consolidating the Legal and Institutional Foundations of Social Dialogue in the Countries of Western Balkans and Moldova*. Geneva: Evaluation Unit.
- International Labour Organization (2008). *Independent Evaluation of the ILO's Strategy to Support Member States to Improve the Impact of International Labour Standards*. Geneva: International Labour Office.
- Klugman, B. (2010). *Evaluating Social Justice Advocacy. A Values-Based Approach*. Washington DC: Centre for Evaluation Innovation.
- Leeuw, F. and Vaessen, J. (2009). *Impact Evaluations and Development. Nonie Guidance on Impact Evaluation*. Washington, DC: The World Bank.
- Levine, C. and Chapoy, C. (2013). "Building on What Works: Commitment to Evaluation (c2e) Indicator." *Working Paper 19*. New Delhi: International Initiative for Impact Evaluation.
- Louie, J. and Guthrie, K. (2007). "Strategies for Assessing Policy Change Efforts: A Prospective Approach." *The Evaluation Exchange XIII (1)*.
- Mackie, J. (1974). *The Cement of the Universe*. Oxford: Oxford University Press.
- Mayne, J. (ed.) (2012). "Contribution Analysis", *Evaluation*, Special Issue, 18(3).
- Mayne, J. (2011) "Contribution Analysis: Addressing Cause and Effect" in *Evaluating the Complex*, K. Forss, M. Marra and R. Schwartz (Eds.). Piscataway, New Jersey: Transaction Publishers.
- Morra-Imas, L. and Rist, R. (2009). *The Road to Results. Designing and Conducting Effective Development Evaluations*. Washington DC: The World Bank.
- Muhweezi, A., Msham, H. and Mourshed, S. (2010). *Nile Transboundary Environment Action Project. Terminal Evaluation Report*. Entebbe: Nile Basin Initiative Secretariat.
- Nexus Associates, Inc. (2012). *Better Work Haiti. Mid-term Evaluation*. Geneva: International Labour Organization.

- Organization for Economic Cooperation and Development/Development Assistance Committee (2012). *Development and Aid Effectiveness. Glossary of Key Terms in Evaluation and Results-based Management*. Paris: OECD/DAC.
- Organization for Economic Cooperation and Development/Development Assistance Committee (2010). *Quality Standards for Development Evaluation*. Available at: www.oecd.org/dataoecd/55/0/44798177.pdf
- Organization for Economic Cooperation and Development/Development Assistance Committee (2002). *Glossary of Key Terms in Evaluation and Results Based Management*. Paris: OECD, DAC Working Party on Aid Evaluation.
- Ortiz, E. and Inomata, T. (2009). *Assessment of the Global Mechanism of the United Nations Convention to Combat Desertification (JIU/REP/2009/4)*. Geneva: United Nations, Joint Inspection Unit.
- Oxfam UK (2012). *Process Tracing. Draft Protocol*. Guidance on Evaluating Policy Influencing Work. Available at: www.oxfam.org.uk
- Patton, M. (2011a). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford Press.
- Patton, M. (2011b). "Improving Rigor in Service-Learning Research." In Hatcher, J. and Bringle, R. *Understanding Service-Learning and Community Engagement: Crossing Boundaries Through Research*. New Orleans: International Association for Service-Learning and Community Engagement.
- Patton, M. (2008). *Utilization-focused Evaluation: The New Century Text 4th edition*. Thousand Oaks, CA: Sage Publications.
- Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. Thousand Oaks and London: Sage.
- Quinn, M. and Clark, K. (2006). *Progressing the Implementation of CEDAW in the Pacific. An Independent Evaluation of UNIFEM's Programme*. New York: UNIFEM.
- Reisman, J., Gienapp, A., and Stachowiak, S. (2007). *A Guide to Measuring Advocacy and Policy*. Baltimore: Annie E. Casey Foundation.
- Rog, D. (1985). "A Methodological Assessment of Evaluability Assessments." Ph.D. dissertation, Vanderbilt University, Nashville, Tennessee.
- Roman-Morey, E. and Zahran, M. (2011). *Evaluation of the Scope, Organization, Effectiveness and Approach of the Work of the United Nations in Mine Action (JIU/REP 2011/11)*. Geneva: United Nations, Joint Inspection Unit.
- Scriven, M. (1993). *Hard-won lessons in program evaluation. New Directions for Program Evaluation, Series No. 58*. San Francisco: Jossey Bass.

- Stern, E., Stame, N., Mayne, J., Forss, K, Davies R., and Befani, B. (2012). “Broadening the Range of Designs and Methods for Impact Evaluations.” Department for International Development Working Paper No. 38. London: Department for International Development.
- Stiles, M. (2011). *End of Project Evaluation of the Migrant Labour Rights Project (China)*. Gatineau, Quebec: Canadian International Development Agency.
- Stiles, M. (2009). *Evaluation of CIDA-funded Human Rights Cooperation between the University of Ottawa and Peking University*. Gatineau, Quebec: Canadian International Development Agency.
- Stiles, M. (2008). *An Assessment of CIDA’s Contribution to Legal and Judicial Reform in China*. Gatineau: China Programme, Canadian International Development Agency.
- Stiles, M. (2007). *An Evaluation of ESCAP’s Approach to Capacity Development*. Bangkok: United Nations Economic and Social Commission for Asia and the Pacific.
- Stiles, M. (2002). *Evaluation of the Government of Pakistan-UNICEF Advocacy and Social Mobilization Strategy, January 1999-June 2001*. Kathmandu: United Nations Children’s Fund, Regional Office for South Asia.
- Stuart, J. (2007). “Necessity Leads to Innovative Evaluation Approach and Practice.” *The Evaluation Exchange XIII (1)*.
- Tickner, V. (2006). *Impact Assessment of the FAO/MIC Marketing Management Assistance Project in Mozambique, financed by the EC, over the period 2001 – 2005*. Rome: Food and Agricultural Organization of the United Nations.
- UNIFEM (2008). UNIFEM Programme. *Facilitating CEDAW Implementation in Southeast Asia. Evaluation Report*. New York: UNIFEM Evaluation Unit.
- UNIFEM (2004). *Assessment of From Global to Local: A Convention Monitoring and Implementation Project*. New York: UNIFEM, Innovation and Learning Unit.
- United Nations Children’s Fund (2010a). *Advocacy Toolkit: A guide to influencing decisions that improve children’s lives*. New York: UNICEF.
- United Nations Children’s Fund (2010b). Full Terms of Reference: Consultancy Services for the Global Evaluation of the Application of a Human Rights Based Approach in UNICEF Programming.
- United Nations Children’s Fund (2010c). *In Practice Newsletter: Refocusing on Equity*. UNICEF.
- United Nations Children’s Fund (2010d). *In Practice Newsletter: Communication for Development*. UNICEF.

- United Nations Children's Fund (2010e). *Qualitative Study on Essential Family Practices for the prevention and integrated management of child illness and nutrition*. Bamako, Mali: UNICEF.
- United Nations Children's Fund (2009). *Global Evaluation of DevInfo*. New York: UNICEF Evaluation Office.
- United Nations Children's Fund (2008). *Evaluation of Gender Policy Implementation in UNICEF*. New York: UNICEF Evaluation Office.
- United Nations Children's Fund (n.d.). *Reintegration Programme for Children Associated With Armed Forces and Armed Groups Including Children Affected by Conflict in Nepal*. Briefing Note.
- United Nations Economic and Social Commission for Western Asia (2011). *Annual Report of Evaluations 2010*. Beirut: ESCWA.
- United Nations Economic and Social Council (2011). *Programme Evaluation of the Department of Economic and Social Affairs. Report of the Office of Internal Oversight Services, E/AC.51/2011/2*. New York: United Nations Economic and Social Council.
- United Nations Economic and Social Council (2007). *Report of the Office of Internal Oversight Services on the in-depth evaluation of political affairs: Security Council affairs, E/AC.51/2007/2/Add.2*. New York: United Nations Economic and Social Council.
- United Nations Educational, Science and Cultural Organization (2013). *Evaluation of UNESCO's Standard-Setting Work of the Culture Sector (Part I) – 2003 Convention for the Safeguarding of the Intangible Cultural Heritage*. Paris: UNESCO.
- United Nations Educational, Science and Cultural Organization (2012). *Evaluation of the Pilot Phase of the International Fund for Cultural Diversity*. Paris: UNESCO.
- United Nations Educational, Science and Cultural Organization (2012b). *Review of Findings on UNESCO's Normative Work in Culture Captured by Previous Evaluations*. Paris: UNESCO.
- United Nations Educational, Science and Cultural Organization (2010). *Evaluation of UNESCO Strategic Programme Objective (SPO) 11. Sustainably Protecting and Enhancing Cultural Heritage*. Paris: UNESCO.
- United Nations Evaluation Group (2013). *Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management*. New York: UNEG.
- United Nations Evaluation Group (2012). *Terms of Reference. Evaluation of Normative Work in the UN System*. Unpublished document.
- United Nations Evaluation Group, (2012). *UNEG Task Force on Evaluation of Normative Work - Survey Analysis*. New York: UNEG.

- United Nations Evaluation Group (2011). *Integrating Human Rights and Gender Equality in Evaluation – Towards UNEG Guidance*. New York: UNEG.
- United Nations Evaluation Group (2010a). *UNEG Good Practice Guidelines for Follow-up to Evaluations*. UNEG/G(2010)3. New York: UNEG.
- United Nations Evaluation Group (2010b). *UNEG Quality Checklist for Evaluation Reports*, UNEG/G(2010)/2. New York: UNEG.
- United Nations Evaluation Group (2005a). *Standards for Evaluation in the UN System*. New York: UNEG. Available at: www.unevaluation.org/unevaluationstandards
- United Nations Evaluation Group (2005b). *Norms for Evaluation in the UN System*. New York: UNEG. Available at: www.uneval.org
- United Nations Industrial Development Organization (2010a). *Evaluation UNIDO SMTQ [Standards, Metrology, Testing and Quality] Projects in Sri Lanka*. Vienna: UNIDO Evaluation Group.
- United Nations Industrial Development Organization (2010b). *Thematic Evaluation Report. UNIDO activities in the areas of Standards, Metrology, Testing and Quality (SMTQ)*. Vienna: UNIDO Evaluation Group.
- United Nations Industrial Development Organization and United Nations Environment Programme (2008). *Independent Evaluation UNIDO-UNEP Cleaner Production Programme*. Vienna: UNIDO.
- Universalia Management Group (2012). *Global Evaluation of the Application of a Human Rights Based Approach to UNICEF Programming*. New York: UNICEF.
- US Nuclear Regulatory Commission, (2005). *Assessment of the quality of the NRC Research Projects: Memorandum*.
- World Bank (2010). *Cost Benefit Analysis in World Bank Projects*. Washington, DC: World Bank.
- World Bank (2004). *Monitoring & Evaluation: Some Tools, Methods & Approaches*. Washington, DC: Operations Evaluation Department, World Bank.
- White, H. and Phillips, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework*. New Delhi: International Initiative for Impact Evaluation.
- Wilson-Grau, R. and Britt, H. (2012). *Outcome Harvesting*. Cairo: Ford Foundation, Middle East and North Africa Office.