



OUTCOME-LEVEL
EVALUATION

**A COMPANION GUIDE TO THE
HANDBOOK ON PLANNING
MONITORING AND EVALUATING FOR
DEVELOPMENT RESULTS FOR
PROGRAMME UNITS AND EVALUATORS**

DECEMBER 2011

TABLE OF CONTENTS

Acronyms and Abbreviations	iii
1 Introduction	1
1.1 UNDP Evaluation Policy: focusing on the outcome level	1
1.2 Purpose of the companion guide on outcome-level evaluation	1
1.3 Structure of the companion guide on outcome-level evaluation	2
2 Evaluating Outcomes	3
2.1 What is an outcome?	3
2.2 What is an outcome-level evaluation?	4
2.3 Types of outcome-level evaluations	4
3 Being Ready for Outcome-Level Evaluations	6
3.1 The evaluation plan	6
3.2 Results frameworks	8
3.3 Partnerships	8
3.4 Monitoring	9
3.5 Timing	10
4 Assuring the Quality of Outcome-Level Evaluations	10
4.1 Management	10
4.2 Peer and reference groups	10
4.3 Defining the evaluation approach	11
4.4 Quality assessment system for decentralized evaluation reports	12
5 Designing Outcome-Level Evaluations	13
5.1 Revisiting the results framework	13
5.2 Relating evaluation criteria to the results framework	16
5.3 Relating evaluation criteria to evaluation questions	17
5.4 Making the most of the evaluation matrix	19

6	Conducting Outcome-Level Evaluations	20
6.1	Attribution...	20
6.2	...versus contribution	21
6.3	Interpreting contributions to outcomes	22
6.4	Dealing with outcomes and indicators that are not clearly defined	23
6.5	Assessing the effectiveness of UNDP partnerships in contributing to the achievement of outcomes	24
6.6	Collecting and analysing data to ascertain contribution to outcomes	26
7	Telling the Outcome Story	28
7.1	The red thread	28
7.2	Suggested report structure	29
	Annexes	
	Annex 1. Sample Table of Contents for an Inception Report	31
	Annex 2. Example of an Evaluation Matrix	32
	Annex 3. Semi-Structured Interview Guide	36
	Annex 4. Preparing Interview Summary Sheets	48
	Tables and Figures	
	Table 1. An Example of Mapping UNDP Activities, Outputs and Intermediate Results	24
	Table 2. Examples of Assessment of Effectiveness: Evidence at Project Level	27
	Table 3. Suggested Report Structure	30
	Figure 1. The Intention of the Results Chain	7
	Figure 2. Model of UNDP Contribution to Outcomes and Impact	14
	Figure 3. Effectiveness and Efficiency – Relating Evaluation Criteria to Results Categories	16



ACRONYMS AND ABBREVIATIONS

ADR	Assessment of Development Results
ARE	Annual Report on Evaluation
BDP	Bureau for Development Policy
CPAP	Country Programme Action Plan
CPD	Country Programme Document
PME Handbook	Handbook on Planning, Monitoring and Evaluating for Development Results
POPP	Programme and Operations Policies and Procedures
SU/SSC	Special Unit for South-South Cooperation
UNDAF	United Nations Development Assistance Framework
UNEG	United Nations Evaluation Group

1. INTRODUCTION

1.1 UNDP EVALUATION POLICY: FOCUSING ON THE OUTCOME LEVEL

According to the UNDP Evaluation Policy, programme units – policy, practice and regional bureaux, as well as country offices – ensure that “planned evaluations assess [their] **contribution to development results at the outcome level**”.¹ This implies that, in evaluating their own performance, programme units should be preoccupied with ascertaining whether and how they have helped bring about changes in human development conditions, including in the behaviour of people and/or institutions targeted through UNDP initiatives. Moreover, they should try to understand why particular initiatives have or have not succeeded in a given context. Such initiatives could comprise broad programmes or programme components, clusters of projects or individual projects, and activities such as advocacy or advisory services. They all, in one way or another, aim to make a difference – i.e., to contribute to one or several outcomes.

The purpose of outcome-level evaluation is to find out how UNDP has gone about supporting processes and building capacities that have, indeed, helped make a difference. In doing so, evaluation aims to identify which UNDP approaches have worked well and which have faced challenges, and to use lessons learned to improve future initiatives and generate knowledge for wider use. Evaluation also serves the purpose of holding UNDP accountable for the resources invested in its work.

1.2 PURPOSE OF THE COMPANION GUIDE ON OUTCOME-LEVEL EVALUATION

UNDP has in place a range of guidance on evaluation, which is aligned with the norms and standards of the United Nations Evaluation Group (UNEG),² including:

- The UNDP Evaluation Policy
- The UNDP Handbook on Planning, Monitoring and Evaluating for Development Results (PME Handbook)³
- The UNDP Programme and Operations Policies and Procedures (POPP)⁴
- The UNDP Evaluation Resource Centre (ERC), which contains additional guidance and links to relevant resources.⁵

¹ UNDP, ‘The evaluation policy of UNDP’, DP/2011/3, 10 November 2010, p.10.

² <www.uneval.org>

³ UNDP, ‘Handbook on Planning, Monitoring and Evaluating for Development Results’, New York, 2009.

⁴ Available on the UNDP intranet: <content.undp.org/go/userguide/>.

⁵ Available on the UNDP internet: <erc.undp.org>. This includes the UNDP Evaluation Guidance for GEF-Financed Projects and the UNDP Guidance Note on the Quality Assessment System for Decentralized Evaluation Reports.

This companion guide will seek to keep duplication of information already contained in the above-mentioned guidance to a minimum. Process and management-related issues relating to all UNDP evaluations, even if they are relevant to outcome-level evaluations, will thus not be repeated in any detail. Rather, the purpose of the companion guide is to provide support on approaches and methodologies pertaining to evaluations that are pitched at the outcome level – i.e., evaluations whose aim it is to ascertain that UNDP has contributed to changes in human development conditions. It aims to improve the quality of outcome-level evaluation and, thereby, to strengthen UNDP’s ability to manage for development results.

The primary target audiences of this companion guide are:

- UNDP management and staff of programme units responsible for planning outcome-level evaluations and commissioning evaluators to conduct such evaluations
- Independent evaluators commissioned by UNDP to conduct outcome-level evaluations
- Stakeholders in the evaluation, including governments, civil society entities, United Nations organizations, multilateral and bilateral donors and other development partners.

1.3 STRUCTURE OF THE COMPANION GUIDE ON OUTCOME-LEVEL EVALUATION

Against the background of Chapter 1, Chapter 2 seeks to build an understanding of the intention of outcome-level evaluations. It also outlines how different types of evaluation, e.g., of programmes, projects or themes, can be pitched at the outcome level.

Chapter 3 reiterates critical instructions provided in the PME Handbook, emphasizing issues that are of particular importance in planning for outcome-level evaluations. For example, ensuring that a programme, project or other type of initiative is evaluation-ready requires a robust results framework. Consistent and meaningful results-oriented monitoring, especially of the progress made towards achieving outcomes, is of particular importance, as is the maintenance of strong partnerships with government and other entities that are also working towards the achievement of the same outcome.

Chapter 4 focuses on how quality can be assured in managing outcome-level evaluations and provides suggestions as to how to strengthen the evaluation process. The use of peer and reference groups is discussed, as well as the significance of terms of reference and inception reports as quality assurance milestones.

Chapter 5 provides guidance on methodological concerns in designing outcome-level evaluations. It seeks to clarify how to emphasize an outcome-level focus; how to use the results framework of an initiative as a starting point in defining the evaluation scope, applying evaluation criteria and developing evaluation questions; and how to develop an evaluation matrix.

Chapter 6 provides guidance on methodological concerns in conducting outcome-level evaluations. It discusses the focus of UNDP outcome-level evaluations on assessing contribution to the achievement of outcomes. Key issues in analysing and interpreting quantitative and qualitative data are reviewed, and guidance on evaluating partnerships is provided.

Chapter 7, finally, focuses on the preparation of the evaluation report. An evaluation, while telling a story, needs to build an argument, providing a background, furnishing evidence, drawing conclusions and formulating recommendations.

This guide does not cover use of evaluations, which is described in Chapter 8 of the PME Handbook and the POPP. The relationship between outcome-level evaluations and independent evaluations conducted by the UNDP Evaluation Office is described in Section 5.3 of the PME Handbook.

2. EVALUATION OUTCOME

2.1 WHAT IS AN OUTCOME?

Chapter 2 of the PME Handbook and, in particular, Section 2.4 on Finalizing the Results Framework, defines the concept of an outcome, and describes how outcomes relate to different levels of the results framework:

“Outcomes describe the intended changes in development conditions that result from the interventions of governments and other stakeholders, including international development agencies such as UNDP. They are medium-term development results created through the delivery of outputs and the contributions of various partners and non-partners. Outcomes provide a clear vision of what has changed or will change globally or in a particular region, country or community within a period of time. They normally relate to changes in institutional performance or behaviour among individuals or groups. Outcomes cannot normally be achieved by only one agency and are not under the direct control of a project manager.”⁶

It is this latter point that is key to understanding the nature of outcomes. Outcomes are beyond the managerial responsibility, beyond the immediate control, of UNDP programme or project managers. However, UNDP is responsible for planning and implementing initiatives in such a way that they are most likely to contribute to the achievement of outcomes; UNDP can thus be held accountable for the achievement of results. Outcomes happen as a result of, all the work that has been done by UNDP in cooperation with development partners. Outcomes reflect the developmental momentum that has been gained by primary stakeholders in programme countries, as a consequence of UNDP’s initiatives. Outcomes are what primary stakeholders do under their own steam, upon their own initiative, following UNDP’s delivery of outputs – the services and products generated under a programme or project. Outcomes are not the sum of outputs delivered through UNDP programmes and projects; rather, they occur when outputs are used by primary stakeholders to bring about change.

‘Knowledgeable training workshop participants’ are not an outcome. ‘UNDP advisory reports’ are not an outcome. ‘UNDP procurement of medication’ is not an outcome. An outcome is when men and women use knowledge gained through UNDP training in their day-to-day work and bring about changes. An outcome is when UNDP advisory reports are used by government officials to develop new policies. An outcome is when counterparts use UNDP models and systems to develop transparent and accountable procurement systems of their own.

⁶ UNDP, ‘Handbook on Planning, Monitoring and Evaluating for Development Results’, *ibid*, p.56.

Outcomes are not what UNDP delivers, but the developmental achievement to which UNDP contributes. Outcomes are what UNDP's work is ultimately all about: making a difference.

2.2 WHAT IS AN OUTCOME-LEVEL EVALUATION?

According to the POPP, “decentralized evaluations [i.e., evaluations commissioned by programme units] should always address UNDP's contribution to development results at the outcome level, and address the short-term, medium-term and long-term contributions of UNDP to the attainment of outcomes”. First and foremost, development results are defined as part of national development plans or poverty reduction strategies. UNDP seeks to contribute to the achievement of such national results by formulating outcomes in the context of:

- United Nations Development Assistance Frameworks (UNDAFs);
- The results frameworks of global, regional and South-South programmes;
- The results frameworks of country programmes, as defined in the Country Programme Documents (CPDs) and further elaborated in the Country Programme Action Plans (CPAPs);
- The results frameworks of projects, although projects, increasingly, are designed to contribute explicitly to the achievement of outcomes defined in programmes, and do not necessarily define their own outcomes;
- Cross-cutting themes, such as gender equality, capacity development or environmental sustainability, that may or may not be reflected in distinct results frameworks, but whose expected outcomes can be derived from UNDP policy statements, including the Strategic Plan.

Based on the above, all decentralized evaluations should, in principle, be pitched at the outcome level. In evaluating any kind of UNDP initiative, it should not only be possible to ascertain to what extent it was designed to contribute to the achievement of outcomes; it should also be possible to obtain an indication of how successful the initiative was in, in fact, contributing towards the achievement of outcomes.

While outcome-level evaluations focus on outcomes, this does not mean that other aspects of UNDP initiatives are neglected. In order to understand whether everything has been done to contribute to the achievement of outcomes, evaluations also need to look at how well the initiative was planned, what activities were carried out, what outputs were delivered, how processes were managed, what monitoring systems were put in place, how UNDP interacted with its partners, etc. Outcome-level evaluation does not, therefore, imply an exclusive preoccupation with outcomes; but it does mean that all UNDP initiatives should be evaluated in terms of their overriding intention to bring about change in human development conditions at the outcome level.

Notwithstanding the above, it is conceivable that some decentralized evaluations would focus exclusively on issues relating to processes, implementation or management arrangements, especially in a larger organizational context. Such evaluations may, but would not necessarily, address UNDP's contribution at the outcome level. However, they would be highly specialized evaluations that may be warranted in particular circumstances, e.g., during periods of organizational change.

2.3 TYPES OF OUTCOME-LEVEL EVALUATIONS

As specified in the POPP,⁷ the following types of evaluations should be pitched at the outcome level:

- UNDAF evaluations, mostly conducted jointly with other UN organizations;
- Global, regional and South-South programme evaluations (looking at the programme as a whole).
While the Evaluation Office is responsible for end-of-cycle independent evaluations of such

⁷ Ibid.

programmes, the Bureau for Development Policy (BDP) (responsible for implementing the global programme), the regional bureaux (responsible for implementing the regional programmes) and the Special Unit for South-South Cooperation (SU/SSC) (responsible for implementing the South-South programme), may commission mid-term evaluations of their respective programmes;

- Country programme evaluations (looking at the programme as a whole). While the Evaluation Office conducts end-of-cycle independent evaluations, known as Assessments of Development Results (ADRs), in some countries, the Evaluation Office does not evaluate all country programmes at the end of their cycle. In the absence of an ADR, a country office may commission an end-of-cycle programme evaluation. It may also, in certain cases, wish to commission a mid-term programme evaluation;
- Evaluations of programme components, for example, of one or several outcomes defined in a programme, or of sets of projects that contribute to the achievement of one or several programme outcomes, as defined in global, regional, country and South-South programmes;
- Individual project evaluations that aim to contribute to outcomes as defined in project documents or broader programme frameworks (e.g., UNDAFs or CPAPs);
- Thematic evaluations covering, for example, UNDP focus areas, cross-cutting issues such as gender and capacity development, partnerships, programmatic approaches, cooperation modalities, or business models.

Other types of evaluation, depending on the needs of programme units, are also conceivable and could be pitched at the outcome level.

3 BEING READY FOR OUTCOME-LEVEL EVALUATIONS

3.1 THE EVALUATION PLAN

The Evaluation Policy, Chapter 3 of the PME Handbook and the POPP provide detailed information on the requirements of programme units to submit evaluation plans to the Executive Board at the time a new programme is approved – whether it is a global, regional, country or South-South programme.

Key to developing evaluation plans is that they should “**address all outcomes in the programme document**”.⁸ Different types of evaluation – i.e., UNDAF, programme, project, outcome-specific, thematic, etc. – could be specified in putting together such an evaluation plan. It is up to the discretion of every programme unit, in coordination with its development partners and in accordance with existing management arrangements, to devise the most “appropriate mix outcome-level...evaluations”.⁹ The Executive Board, subject to its approval of the plan, will hold programme units accountable for the plan’s implementation – bearing in mind the possibility of adjusting the plan during the course of the programme cycle in line with the Evaluation Policy.¹⁰

Programme units that conduct operational activities outside the framework of such programmes, should also plan to commission outcome-level evaluations of their initiatives.

3.2 RESULTS FRAMEWORKS

Chapters 2 and 3 of the PME Handbook, as well as the corresponding sections of the POPP, describe in considerable detail principles and tools to be applied in planning development initiatives. Such principles and tools are to be used in planning an UNDAF, a UNDP programme or project, or any other kind of initiative. A core message is that **plans should be results-oriented**. This results-orientation – i.e. the identification of a causal chain of inter-linked results, whereby UNDP activities lead to the delivery of UNDP outputs, which contribute to the achievement of outcomes and eventually impacts on the part of primary stakeholders (see Figure 1 for an overview of the intention of the results chain) – is reflected in a results framework. Results frameworks tend to be abstractions, reflecting a linear logic of cause and effect, which does not necessarily correspond to the more organic and iterative nature of real development. However, a good results framework can provide an effective blueprint

⁸ UNDP, ‘The evaluation policy of UNDP’, *ibid*.

⁹ *Ibid*.

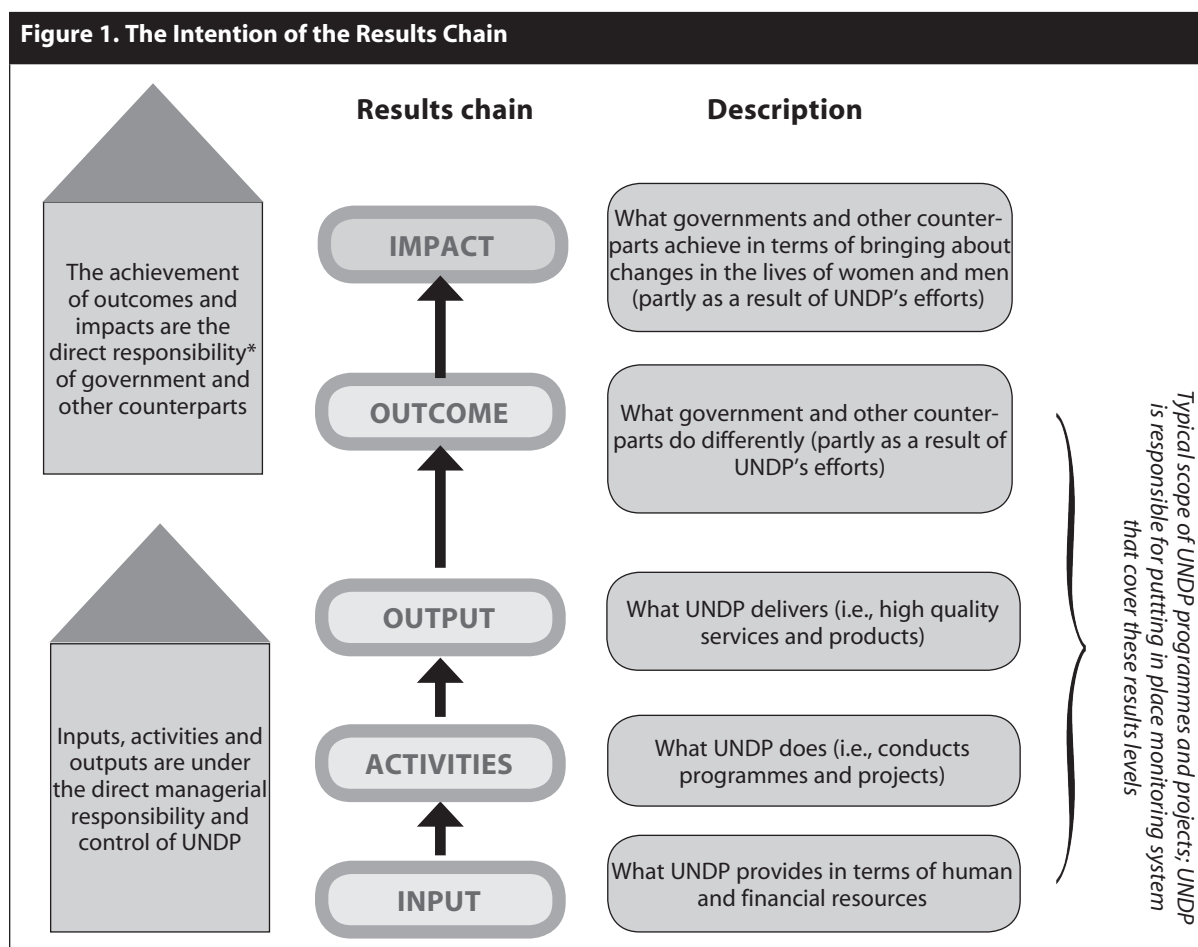
¹⁰ The evaluation plan, formulated by the programme unit, thus effectively becomes the basis for the Executive Board’s assessment of compliance.

for a given initiative, not only defining different levels of results, but also specifying SMART (specific, measurable, achievable, relevant, and time-bound) indicators, baselines and targets. Looking at a results framework should permit a reader to understand what is to be achieved, when and by whom.

The PME Handbook highlights the integrated nature of planning, monitoring and evaluation. While the results framework is a planning tool, it provides the basis for ongoing monitoring, in particular of outputs and the progress made towards the achievement of outcomes. Where they exist, **results frameworks are also the starting point for any evaluation.**

Although the availability of a results framework is not essential for the conduct of an evaluation – i.e., evaluators could also try to define, retrospectively, the intentions of programme or project planners and implementers – it significantly facilitates evaluation; as such, every effort should be made to develop a strong results framework at the planning stage. Results frameworks not only allow evaluators to understand and assess initiatives more easily, they also provide programme managers with a strong basis on which to demonstrate their achievements. **A good results framework thus goes a long way towards ensuring that an initiative is evaluation-ready.**¹¹

In conducting outcome-level evaluations it is also important to understand the interconnected nature of results frameworks, with the UNDAF results frameworks providing higher-level outcomes at the country level, and other programme frameworks – such as those for the global, regional and South-South programmes – linking more immediately into the results defined in the UNDP Strategic Plan.



*UNDP is nonetheless responsible for ensuring that its initiatives are planned and implemented in such a way that they contribute to the achievement of outcomes

¹¹ It should be noted that different UNDP partners may use different terms to describe results categories or may use different definitions to describe the same terms. In such cases, it is important to understand the intention of the respective terms and to align results frameworks accordingly.

3.3 PARTNERSHIPS

As is clear from Figure 1 above, outcomes are achieved by primary stakeholders in government and civil society, and UNDP is rarely alone in contributing towards their achievement. A wide range of influences, as represented by the actions (or lack of actions) of partners and other stakeholders, affect the achievement of planned outcomes. In evaluating whether and how UNDP has contributed to the achievement of outcomes it is, therefore, important to understand how UNDP has defined such influences, what kind of risk management strategy it has put in place to mitigate adverse influences, and how it has engaged with a wide range of partners to ensure the achievement of results. Such partners could include not only government entities, but also national and international non-governmental organizations, faith-based organizations, academic institutions, political parties, regional organizations, bilateral donor agencies, as well as multilateral organizations, including United Nations partners.

In preparing for outcome-level evaluations, it is critical that a thorough stakeholder analysis be conducted.¹² Key partners contributing towards a given outcome should not only be well informed about the evaluation, but should also be closely involved in its design – e.g., through participation in a reference group – and, where possible, participate in its conduct – e.g., through a joint evaluation. This is of particular relevance in evaluating outcomes of country-based programmes and projects that are, typically, planned with the government and other national counterparts in the context of UNDAFs or joint assistance strategies involving contributions by a range of development partners. **Partnership is thus integral to the achievement of outcomes, requiring dedicated analysis by any outcome-level evaluation.** Such evaluations are greatly facilitated not only by well-documented partnerships, but also by the level of preparedness and active engagement of such partners in the evaluation process.

Increasingly, demand is growing for joint outcome-level evaluations, for example, with programme country governments, or with other development partners (e.g., United Nations organizations, multilateral development banks, bilateral donors, non-governmental organizations) contributing towards the achievement of the same outcome, or in the context of a comprehensive UNDAF evaluation. Section 6.3 of the PME Handbook provides specific guidance on the conduct of joint evaluations.

3.4 MONITORING

Chapters 3 and 4 of the PME Handbook and the POPP provide detailed guidance on monitoring for results. **Outcome-level evaluations benefit significantly from the availability of monitoring data** that have been collected consistently, based on monitoring and evaluation systems established to track indicators and targets at the output and, especially, the outcome levels. The availability of baseline data, and the conduct of baseline surveys at the outset of UNDP initiatives where baseline data is not readily available, is of particular importance. The absence of monitoring data not only deprives programme and project managers of a regular means of self-assessment and calibration, it also leaves evaluations without a potential wealth of information on the process of implementation and achievements along the way. Without this information, it is difficult to speak to what works, why and how.

With respect to outcome monitoring, it is particularly important that programme and project partners are closely involved in the establishment and maintenance of monitoring systems. Outcomes, as indicated above, are beyond the direct managerial responsibility of UNDP programme and project managers. They are achieved by primary stakeholders. Where outcome-monitoring systems have not been established by primary stakeholders, UNDP – in coordination with other development partners working towards the same outcomes – should advocate for the establishment of such systems and,

¹² In the context of this Companion Guide, stakeholder analysis is taken to involve the identification of primary partners and other stakeholders, and the determination of their involvement or role in a given programme or project context, prior to the conduct of the evaluation in order to establish their level of participation in preparing for and conducting the evaluation, and how to assess their contribution in methodological terms. A partnership analysis is conducted as part of an outcome-level evaluation, to determine the relevance, effectiveness, efficiency and sustainability of partnership in contributing to the achievement of development results.

where possible, support primary stakeholders in setting them up and maintaining them.

3.5 TIMING

Chapters 3, 5 and 6 of the PME Handbook provide guidance on key concerns in planning for evaluations. In developing an evaluation plan, programme units should also determine when during the programme cycle evaluations specified therein should be conducted. The timing of evaluations should be reassessed during annual programme reviews. Given the outcome-oriented nature of UNDP evaluations, the likelihood of being able to say more about progress made towards achieving outcomes is higher towards the end, or even after the completion of (i.e., through ex-post evaluations), UNDP programmes and projects. However, even mid-term evaluations, or evaluations conducted at the outset of an initiative (i.e. ex-ante evaluations), can say something about progress made towards the achievement of outcomes or in setting up delivery and monitoring systems that are conducive towards the achievement of outcomes. Such evaluations may also provide early indications of the potential sustainability of initiatives and their likelihood of being upscaled. In order to manage the full set of planned evaluations during any given programme cycle, programme units are in fact advised to commission some evaluations during the early stages of their programmes.

In determining the timing of the implementation of the evaluation plan, **it is important to consider the extent to which it would be useful to feed evaluative information into higher-level or broader strategic and programmatic evaluations** that are scheduled, such as UNDAF evaluations or programme evaluations conducted by the Evaluation Office, including ADRs. Depending on the timing of these latter evaluations, other planned evaluations, as reflected in the evaluation plan, should, ideally, be conducted well in advance. The timing of outcome-level evaluations should be determined in coordination with programme and project partners, not least since the primary focus of such evaluations is on determining UNDP's contribution to developmental changes that programme and project partners are primarily responsible for.

4 ASSURING THE QUALITY OF OUTCOME-LEVEL EVALUATIONS

4.1 MANAGEMENT

Chapters 6 and 7 of the PME Handbook and the POPP provide guidance on the evaluation process, and on the role of UNDP staff in planning, commissioning, managing and quality assuring evaluations. Evaluation Office reviews of outcome-level evaluations demonstrate a number of typical shortcomings in managing outcome-level evaluations, including leaving the planning for specific evaluations to the last minute; minimizing the significance of evaluation as a key component of results-based management; avoiding close dialogue with governments and other development partners in planning for, designing and conducting evaluations; overlooking the importance of conducting a thorough stakeholder analysis and involving stakeholders closely in the evaluation planning, design and review process; neglecting to put in place quality assurance mechanisms that will uphold the independence and validity of the evaluation; and allocating insufficient time and resources to the evaluations etc. This range of management challenges will not be specifically addressed in this companion guide. However, the PME Handbook and POPP contain detailed information on the management process, and the ERC provides additional evaluation-related resources. Section 6.3 of the PME Handbook provides specific guidance on the conduct of joint evaluations.

4.2 PEER AND REFERENCE GROUPS

Within the programme unit commissioning an outcome-level evaluation, it is good practice to assign one or two peers from within the programme unit (or even from outside the programme unit) as quality enhancers. These individuals act as ‘critical friends’, as a sounding board, and are available for feedback and advice throughout the evaluation process. They are involved in the review of key evaluation outputs, including terms of reference, the inception report, and successive versions of the draft evaluation report.

Programme units should establish a reference group for each evaluation. The reference group should not be too large and include approximately five individuals. It could comprise direct stakeholders in the evaluation, including the government, as well as less directly affected development partners, including non-governmental organizations, civil society representatives, or donors. It could also include colleagues from other UNDP bureaux or offices that may have conducted similar evaluations in the past. Ideally, the reference group should represent a range of perspectives and interests as a means of strengthening the credibility and eventual use of the evaluation. However, it should also include persons who have some technical expertise and experience with regard to evaluation design, conduct and quality assurance. The reference group members should commit to providing comments on terms of reference, the inception report, at least one early draft of the evaluation report, as well as a mature draft

if significant changes are introduced following the circulation of the first draft. While receiving written comments would, by default, be essential, conducting at least one face-to-face meeting, or conducting a video conference where possible, would be preferable.

The concern of the quality enhancers and reference group would be to ensure that evaluation standards, as provided by UNEG,¹³ are adhered to, which would include safeguarding its transparency and independence, and minimizing conflicts of interest. It would also involve providing advice on the evaluation's relevance, on the appropriateness of evaluation questions and methodology, and on the extent to which conclusions and recommendations are plausible in light of the evidence that is presented and action-oriented.

4.3 DEFINING THE EVALUATION APPROACH

Terms of reference and inception reports constitute key tools in assuring the quality of outcome-level evaluations. While terms of reference reflect a programme unit's intentions and proposed approach for the evaluation, the inception report represents the commissioned evaluation team's interpretation of the terms of reference and outlines the specific methodology to be applied. It is essential that both documents be reviewed by peer and reference groups, and that sufficient time be allocated to the review process. As such, programme units should not treat terms of reference and inception reports as internal. Buy-in and ownership on the part of stakeholders can only be obtained if they are given the opportunity to review and contribute to the terms of reference and inception reports, and are essential in increasing the likelihood of the evaluation being used once conclusions and recommendations are made available. In addition to the references in the PME Handbook indicated below, UNEG has also issued a Quality Checklist for Evaluation Terms of Reference and Inception Reports.¹⁴

TERMS OF REFERENCE

Evaluation Office assessments of outcome-level evaluations demonstrate that many programme units do not yet pay sufficient attention to the formulation of clear and realistic terms of reference. Annex 3 of the PME Handbook provides a detailed, annotated template for terms of reference. While programme units have some leeway in customizing terms of reference for their own needs, all elements should be addressed. In developing terms of reference, programme units should be as clear as possible about the scope of the evaluation, including key evaluation questions that are oriented, in particular, towards ascertaining UNDP's contribution to the achievement of outcomes. Similarly, management arrangements should be specified along with an indication of outputs and a rough timeframe. Terms of reference often err on the side of detail when it comes to defining methodological approaches and timelines. Some flexibility should be left for the evaluation team to propose, in its inception report, how it will go about conducting the evaluation.

INCEPTION REPORTS

Chapter 6 of the PME Handbook discusses the requirement that evaluators should prepare an inception report after having conducted a preliminary round of discussions and assessments. The inception report defines the scope, design and associated technical instruments as well as a specific implementation plan. The purpose of the inception report is to interpret and clarify the terms of reference, and to come to a basic agreement with the commissioning programme unit on how the evaluation will be conducted, how the contribution to the achievement of outcomes will be ascertained, and what the final product will look like.

Evaluation office reviews of the quality of outcome-level evaluation reports indicate that programme units often do not require inception reports to be prepared. In neglecting this step, programme units

¹³ See <www.uneval.org/normsandstandards/index.jsp?doc_cat_source_id=4> (accessed October 2011).

¹⁴ See <www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=608> (accessed May 2011).

are deprived of an important step in assuring the quality of the evaluation. Inception reports should be kept brief and, particularly in the case of smaller project evaluations, can be kept to a bare minimum. However, all programme units should, as a matter of principle, require the preparation of an inception report that, in particular, demonstrates how the evaluation team will assess UNDP's contribution to the achievement of outcomes. Annex 1 contains a proposed outline to be used for inception reports for complex or large evaluations. At its most basic the inception report should clarify the methodology to be used in the evaluation and should include an evaluation matrix (see Section 5.4 below) and interview protocols (Annex 1 suggests a number of minimum requirements).

4.4 QUALITY ASSESSMENT SYSTEM FOR DECENTRALIZED EVALUATION REPORTS

Following the UNDP Executive Board's approval of the new Evaluation Policy, the Evaluation Office has been mandated to conduct regular assessments of all decentralized evaluations conducted by UNDP programme units. Detailed information regarding these assessments is contained in the Guidance Note on the Quality Assessment System for Decentralized Evaluation Reports that is posted in the ERC.¹⁵ The quality of outcome-level evaluations will be assessed, inter alia, on the basis of quality criteria contained in the PME Handbook, the POPP and this Companion Guide. In planning for, designing and conducting outcome-level evaluations, programme units should refer to the Guidance Note to ensure full adherence to stated quality standards.

¹⁵ <erc.undp.org>

5 DESIGNING OUTCOME-LEVEL EVALUATIONS

Chapter 7 of the PME Handbook provides extensive information on important steps in designing an evaluation, and key issues pertaining to evaluation methodology, including defining the evaluation context, purpose, scope, criteria and questions, as well as issues relating to data collection, analysis and synthesis. This information is equally relevant to the design of outcome-level evaluations. The following sections provide additional information to help clarify challenges that require particular attention in designing outcome-level evaluations.

5.1 REVISITING THE RESULTS FRAMEWORK

A key step in preparing for an outcome-level evaluation is to develop an explicit model of how outcomes might – or have actually – occurred. The initial development of an outcome model takes place during programme planning. Examples of this are problem trees or results maps, which are described in Chapter 2 of the PME Handbook. The model is then abstracted and reflected in a results framework as part of a programme or project document (see also Section 3.2 above). Such an outcome model is likely to evolve, and should be iteratively developed in coordination with stakeholders, during the life of the programme or project.

WHAT IS AN OUTCOME MODEL?

An outcome model (also known as results maps, logic models, programme logics, intervention logics, means-ends diagrams, logical frameworks, theories of change, programme theories, outcomes hierarchies and strategy maps, among other names)¹⁶ is a (visual) map of the causal logic of an initiative being evaluated. It shows how certain initiatives (activities, outputs) are conceived as contributing to bringing about desired positive change. The outcome model includes a description of what UNDP contributes in its own right, what it contributes with partners, what partners may do independently, and what non-partners might do.

An outcome model goes beyond the activities and outputs that specific programmes or projects can control and displays the expected full results chain including contributions to outcomes and even impact. An outcome model can come in many forms as it is a visual representation of how change is viewed. What is common to each outcome model is its coverage of the different levels of the results chain – activities, outputs, outcomes (short and long-term) and impacts. Figure 2 illustrates the

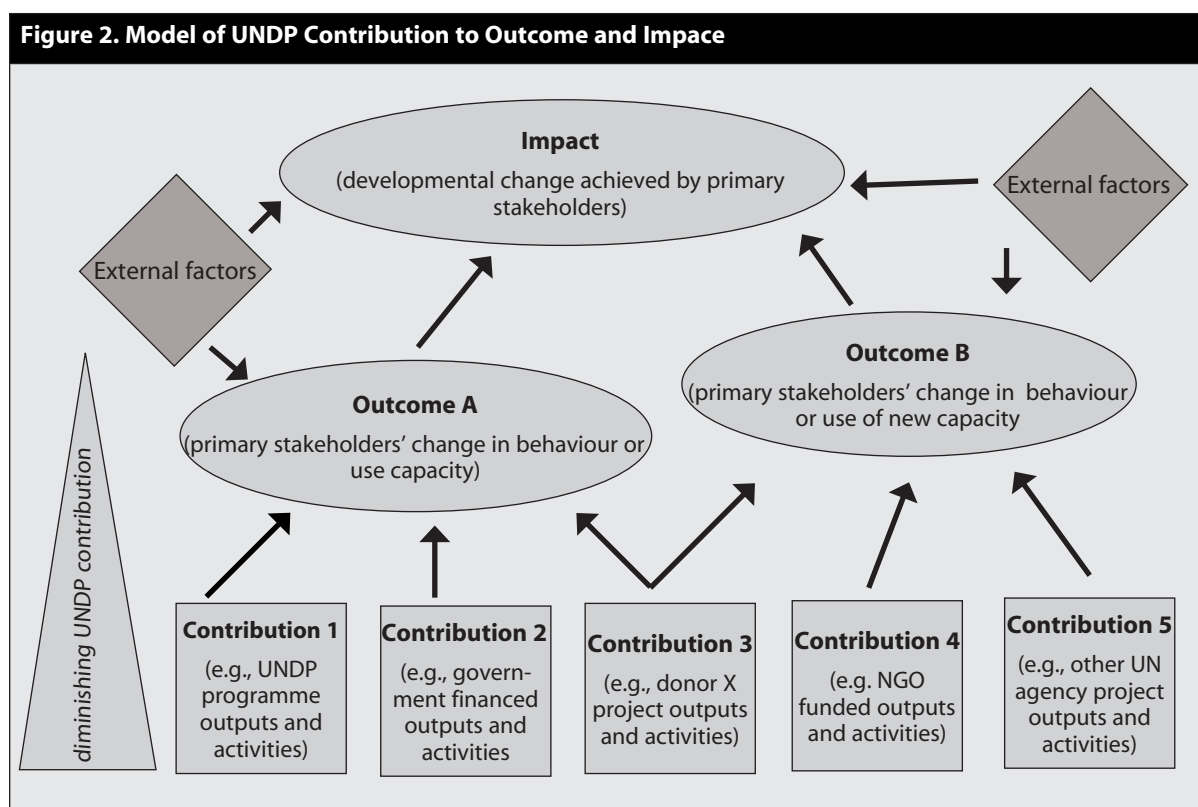
¹⁶ See <knol.google.com/k/what-are-outcomes-models-program-logic-models#> (accessed May 2011). See also, e.g., <outcomesmodels.org/>. It should be noted that different authors and organizations define these terms differently and may differentiate quite clearly between the intention and use of, for example, an outcome model (e.g., focusing on the achievement of a broad range of development results) and a logical framework (e.g., focusing on a narrowly defined causal chain linking one set of results to another).

potential complexity of an outcome model, highlighting UNDP's diminishing contribution to the achievement of outcomes and impact, while taking into account possible external factors that may be crucial in understanding the extent to which outcomes and impact are achieved.

WHY DEVELOP AN OUTCOME MODEL?

There are several reasons for developing or revising an outcome model as an initial step in undertaking an outcome-level evaluation:

- It establishes a map that captures both the codified and tacit knowledge of partners and stakeholders about how an outcome is intended to be achieved;
- It identifies the intended target group of the initiatives at the outcome level and the expected changes that the initiatives will contribute to;
- The process assists in developing key evaluation questions by highlighting information gaps or areas that require further explanation;
- It provides a framework for the evaluative analysis (a performance story framework);
- It spells out the route to impact for UNDP, stakeholders and non-stakeholders;
- It represents the complicated nature of achieving outcomes. Many pieces of a jigsaw need to be put together to create the final picture – reflecting its complex nature. Activities emerge in response to changing needs and opportunities and results emerge in response to the interplay between activities and the changing context.



HOW IS AN OUTCOME MODEL DEVELOPED?

As described in Chapter 2 of the PME Handbook, an outcome model should be developed in close consultation with relevant stakeholders and partners. A key principle for developing an outcome model, particularly at the planning stage, or where no outcome model exists, is working downwards from the outcomes, not upwards from the activities. It is therefore important to begin with a 'rich picture' of

the outcome that explains what it would look like if it were achieved (this could include, for example, a more detailed explanation of different dimensions of the outcome, as well as possible related indicators). While outcomes are formally defined in programme plans, these sometimes do not adequately describe what the outcome is in sufficient depth to plan an outcome evaluation.

A rich picture of success goes beyond simply meeting targets to describe the different aspects of the outcome and to identify specifically who would be doing what differently. When developing the rich picture, it is important to consider what it looks like in terms of the relevant evaluation criteria of relevance, effectiveness, efficiency, and sustainability of the initiative (see Section 7.4 of the PME Handbook for more guidance, as well as Section 5.3).

This rich picture will draw on any previous work that has been done to establish benchmarks or targets, including the findings of previous evaluations, together with input from the stakeholders to answer the following questions:

- If we were successful in achieving this outcome, what would we actually see happening?
- Who would be doing what differently as a result of our initiative?
- The next stage is to identify the projects, programmes, policies and other modalities of technical cooperation (including, e.g., advisory and information services, or advocacy), being undertaken by UNDP, by partners, and by other organizations that may contribute to – or be a barrier to – achievement of the outcome. These should be listed. The most important ones should then be included in a visual display of the outcome model.

Developing a model should bring together key programme or project staff, including relevant specialists (e.g., gender experts or human rights advisers), with key partners and, where appropriate, other stakeholders. Where possible, a monitoring and evaluation specialist could facilitate the session. Experience shows that developing a model can take up to one day.

WHAT IF NO OUTCOME MODEL EXISTED IN THE FIRST PLACE?

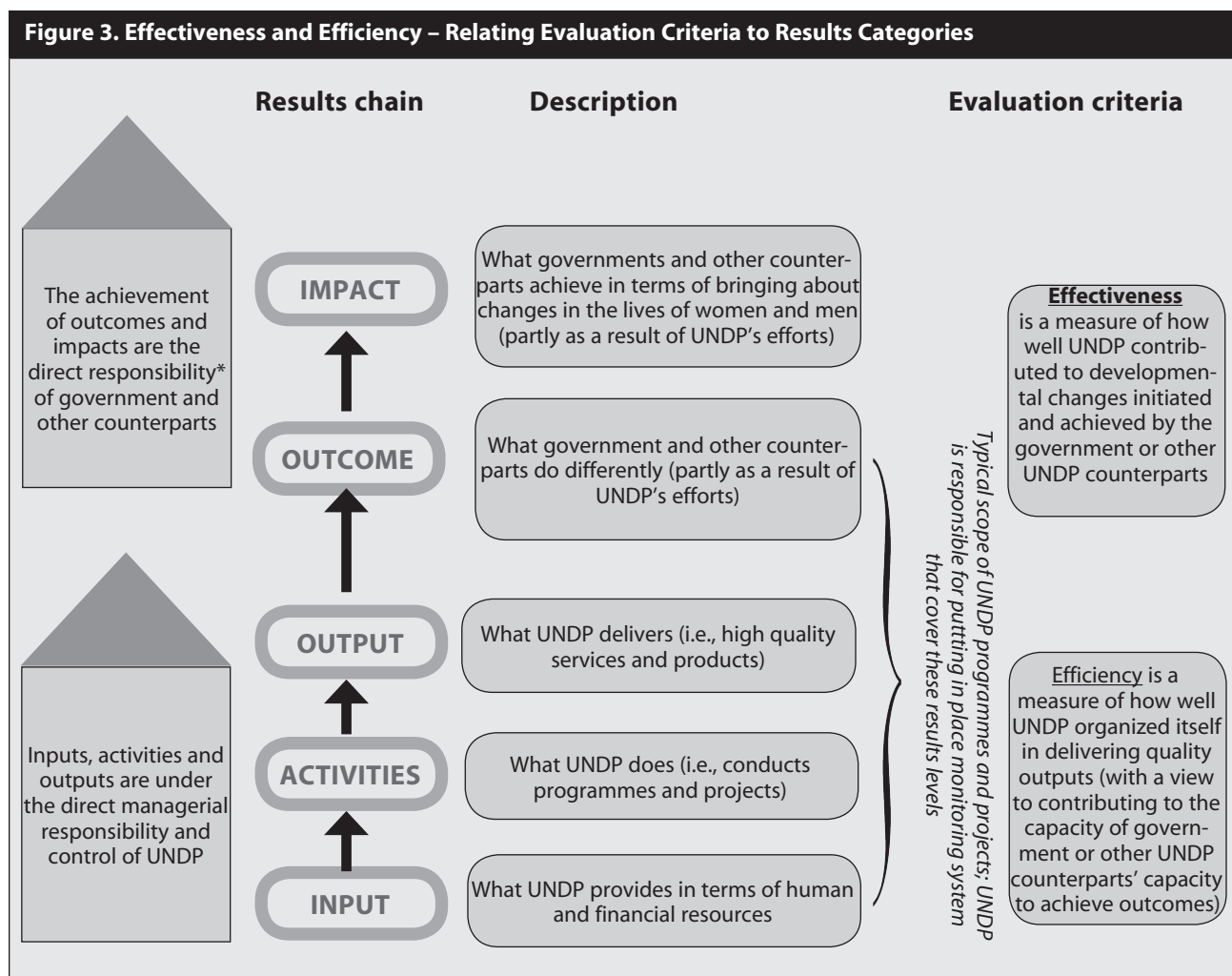
When it comes to commissioning an outcome-level evaluation, the outcome model needs to be updated, or – where one does not exist – may need to be developed for the first time by the commissioning programme unit in close coordination with stakeholders. Thematic evaluations are most likely to face this challenge in so far as cross-cutting concerns, such as human rights or gender equality, may have been integrated across all programme elements in the planning stage. A thematic evaluation focusing, for instance, on gender equality could develop, in cooperation with key stakeholders, an outcome model retrospectively. Such a model would seek to identify outcome-level expectations regarding gender equality at the time the programme was launched, as derived from outcome statements that may specify a differentiated approach towards men and women, or that may highlight issues of gender equality. Subsequently, the outcome model would itemize outputs that would have been required to contribute to the achievement of the outcome, as well as related activities. Such a retrospective outcome model, though often controversial, could be used as a basis on which to design an outcome-level (thematic) evaluation.

HOW IS AN OUTCOME MODEL USED?

The outcome model identifies the outcome that is the focus of the evaluation, the activities of the UNDP, partners and others that may affect it, the influence of major external factors, and possible unintended outcomes. It may also show intermediate outcomes that would be evident before longer-term outcomes, and which might be used as benchmarks that progress is being made towards the achievement of longer-term outcomes. Such intermediate outcomes could, in fact, be used as a basis on which to conduct mid-term evaluations. They may also lend themselves particularly well to ascertaining contribution to the achievement of complex social change and to highlighting progress in

human development, for example, in contributing to greater gender equality and the realization of human rights.

The process of developing the outcome model, for example, through the conduct of inception missions and/or stakeholder workshops, identifies possible gaps in understanding or in the initiative itself, which will help in focusing the evaluation. Outcome models may differ from the original programme or project design as a greater understanding of the issues may have developed over the lifetime of the initiative. Such a reassessment of the feasibility of achieving certain outcomes within given circumstances may be controversial, not least since it may imply deficiencies in planning and/or weaknesses in implementation. However, possible differences between the original programme design and a reassessed outcome model should be noted and incorporated into the terms of reference. The outcome model should be signed off by the evaluation reference group and included as an annex in the terms of reference of the evaluation.



5.2 RELATING EVALUATION CRITERIA TO THE RESULTS FRAMEWORK

Section 7.4 of the PME Handbook provides an overview of how to focus an evaluation, and explains the intention of evaluation criteria. Many evaluations struggle to differentiate clearly between efficiency and effectiveness. Figure 3, based on Figure 1, seeks to demonstrate how the criteria of efficiency and effectiveness can be applied to different results levels.

The efficiency criterion applies primarily to the inputs, activities and outputs that are under the immediate managerial influence of UNDP. While efficiency relates to how economically resources are translated into results, it has to do with how 'well' UNDP has managed its work, from planning

and design, through implementation including the establishment and maintenance of monitoring and evaluation systems and practices, and human resources management, to partnership arrangements with governments, donors, non-governmental organizations, civil society and other stakeholders.

The effectiveness criterion applies primarily to the outcome level. Effectiveness assesses the contribution of UNDP's work – i.e., its use of inputs to conduct activities and deliver outputs – to the achievement of outcomes, as defined in results frameworks.

5.3 RELATING EVALUATION CRITERIA TO EVALUATION QUESTIONS

In designing an evaluation, questions should be formulated in such a way that they address the primary evaluation criteria to be used in the evaluation, as described in Section 7.4 of the PME Handbook. Evaluation questions must be tailored to the scope of the specific evaluation and should take into account how information generated from the questions will be used, the extent to which questions can actually be answered satisfactorily, the adequacy and availability of existing data, and the cost of obtaining primary data.

Typically, in formulating evaluating questions, a number of key questions will emerge. Answering such key questions may require the formulation of additional sub-questions. The illustrative questions below do not differentiate between key and sub-questions.

In the context of UNDP planning, monitoring and evaluation, the level of impact typically corresponds to higher-level results identified in the Millennium Development Goals, national development plans or UNDAFs. Impacts – describing changes in people's lives and development conditions at global, regional and national levels – are usually beyond the scope of UNDP evaluations, partly because UNDP initiatives tend to be relatively modest in reach and are often pitched at the policy level. As such, it is particularly difficult to assess the extent to which UNDP may have contributed to the achievement of impacts on the part of primary stakeholders, bearing in mind the vast array of factors that may have influenced development in an area in which UNDP provides support (see also Figures 2 and 3 above). However, the conduct of impact-level evaluations in UNDP is not ruled out, and larger initiatives in particular, that aim to contribute directly to the human development of men and women, could well be evaluated at the impact level. Impact evaluations tend to require a particular methodological approach that will not be described in this companion guide. Typically, though not in all cases, impact evaluations may build more strongly on quantitative data and use rigorous quasi-experimental designs including counterfactual analyses.

In assessing relevance, the following types of questions may be asked:

- To what extent is the initiative in line with UNDP's mandate, national priorities and the requirements of targeted women and men?
- How did the initiative promote UNDP principles of gender equality, human rights and human development?
- To what extent is UNDP's engagement a reflection of strategic considerations, including UNDP's role in a particular development context and its comparative advantage?
- To what extent was UNDP's selected method of delivery appropriate to the development context?
- To what extent was the theory of change presented in the outcome model a relevant and appropriate vision on which to base the initiatives?

In assessing effectiveness, the following types of questions may be asked:

- To what extent have outcomes been achieved or has progress been made towards their achievement?
- How have corresponding outputs delivered by UNDP affected the outcomes, and in what ways

have they not been effective?

- What has been the contribution of partners and other organizations to the outcome, and how effective have UNDP partnerships been in contributing to achieving the outcome?
- What were the positive or negative, intended or unintended, changes brought about by UNDP's work?
- To what extent did the outcomes achieved benefit women and men equally?

In addressing efficiency, the following types of questions may be asked:

- To what extent have the programme or project outputs resulted from economic use of resources?
- To what extent were quality outputs delivered on time?
- To what extent were partnership modalities conducive to the delivery of outputs?
- To what extent did monitoring systems provide management with a stream of data that allowed it to learn and adjust implementation accordingly?
- How did UNDP promote gender equality, human rights and human development in the delivery of outputs?

In addressing sustainability, the following types of questions may be asked:

- What indications are there that the outcomes will be sustained, e.g., through requisite capacities (systems, structures, staff, etc.)?
- To what extent has a sustainability strategy, including capacity development of key national stakeholders, been developed or implemented?
- To what extent are policy and regulatory frameworks in place that will support the continuation of benefits?
- To what extent have partners committed to providing continuing support?
- How will concerns for gender equality, human rights and human development be taken forward by primary stakeholders?

All UNDP evaluations need to assess the degree to which UNDP initiatives have supported or promoted gender equality, a rights-based approach and human development.¹⁷ In this regard, UNEG guidance on integrating human rights and gender equality in evaluations should be consulted.¹⁸ Key principles comprise:

- Inclusion – i.e., paying attention to which groups benefit and which groups contribute to the initiative under review. Groups need to be disaggregated by relevant categories including disadvantaged and advantaged groups (sex, class, ethnicity, religion, age, location, etc.), duty bearers and rights holders, etc.
- Participation – i.e., stakeholders must be consulted and participate in decisions about what will be evaluated and how the evaluation will be conducted.
- Fair power relations – i.e., bearing in mind power relations between stakeholders and how the initiative affects such relations.

Such principles could either be treated as evaluation criteria in their own right, or as part of the analysis of the main evaluation criteria. In all cases, they should be addressed through evaluation questions. Section 7.4 of the PME Handbook highlights that not all evaluation criteria will be given equal weight

¹⁷ UNDP, 'The evaluation policy of UNDP', *ibid.*

¹⁸ UNEG, 'Handbook for Integrating Human Rights and Gender Equality in Evaluations in the UN System', 2011.

in every evaluation and that additional evaluation criteria may be used. In addition to the principles mentioned above, concepts such as equity, connectedness, coherence or utility could also be used as evaluation criteria. However, if additional criteria are used, they should be clearly defined and differentiated from other criteria. A proliferation of criteria should be avoided. In most cases, additional criteria can be subsumed under the main evaluation criteria. Rather, specific dimensions of broader evaluation criteria could be explored in the context of more narrowly defined evaluation questions.

Key evaluation questions should be defined in the terms of reference of the evaluation. A list of evaluation criteria and corresponding evaluation questions should be elaborated in the evaluation matrix as part of the inception report (see below, and Section 4.3 above).

5.4 MAKING THE MOST OF THE EVALUATION MATRIX

Evaluation Office assessments of outcome-level evaluations indicate that evaluators often neglect to base their data collection efforts on an evaluation matrix. Annex 3 of the PME Handbook refers to the importance of using evaluation matrixes and provides a template.¹⁹ Essentially, the evaluation matrix lists the primary evaluation criteria, the related evaluation questions (and, where needed, sub-questions), the data sources required to answer the questions, and the data collection methods.

Evaluation matrixes can be structured in different ways, depending on the scope and ambition of the evaluation. Annex 2 provides an example of an evaluation matrix that has been filled in prior to the conduct of an evaluation. The matrix should guide the data collection process. As the evaluation proceeds, the matrix is also used to collect and display data obtained from different sources that relate to relevant evaluation criteria and questions.

The evaluation matrix should be developed as part of the inception report. The evaluation matrix is a key tool in systematizing the data collection process as well as in making the evaluation process transparent to the commissioners of the evaluation. To commissioners of evaluations, the process of evaluation often amounts to a 'black-box' exercise, whereby evaluators are asked to set out to find evidence and to produce a report at the end of the process. Requiring the maintenance of an evaluation matrix imposes discipline on the evaluation process, helps identify gaps in evidence and assists evaluators in developing a plausible evaluative argument.

¹⁹ UNDP, 'Handbook on Planning, Monitoring and Evaluating for Development Results', *ibid*, p.199.

6 CONDUCTING OUTCOME-LEVEL EVALUATIONS

6.1 ATTRIBUTION...

Most evaluations come across the issue of attribution, that is establishing a causal linkage between a given development initiative and an observed result. There is a very broad literature on causal attribution which is predominantly tilted to the evaluation of projects or discrete initiatives, and to the use of quantitative methods (inferential statistics and econometrics).²⁰ The traditional approach to attribution is derived from statistical principles and consists in comparing a situation “with project” against another “without project” (also called “control observation”). There is ample literature on the theme, which has seen a recent revival of interest in connection with “randomized control trials”.²¹

In the complex social development context in which UNDP, like many other entities, is engaged, it is often quite difficult to attribute the observed results to the initiative of a single organization, for three reasons:

- Several actors cooperate in UNDP projects and programmes, such as national public institutions, civil society, international organizations, as well as, of course, the primary stakeholders of such projects and programmes. All these actors share a part of the responsibility for achievements and shortcomings of the projects.
- Other exogenous factors may determine certain development results. For example, an economic upturn at the country or regional level or more favourable export prices may bring about higher employment rates and incomes, beyond the achievements of UNDP projects. Alternatively, a major change in the politics or policies of the country, not affected by a UNDP-assisted initiative, may be the key factor that explains a set of development results.
- The very nature of the programme in question may be “complex”, effects and outcomes may be emergent and the country programme may not be amenable to linear logic modelling, partly challenging the classical cause-effect linear thinking.²² This is not an uncommon situation and, for these reasons, some evaluation practitioners prefer the word contribution (see Section 6.2 below), meaning that the results cannot be attributed entirely to one single actor only: each actor plays a

²⁰ Scriven, Michael and Chris L. S. Coryn, ‘The Logic of Research Evaluation’, in Michael Scriven and Chris L. S. Coryn, Editors, *New Directions for Evaluation*, Volume 2008 Issue 118, Pages 89 – 105.

²¹ From the classical Campbell, D. and C. Stanley, ‘Experimental and Quasi-Experimental Designs for Research’, Houghton Mifflin Company, USA, 1966; to more recent Bamberger, M. and H. White, ‘Using Strong Evaluation Designs in Developing Countries: Experience and Challenges’, *Journal of MultiDisciplinary Evaluation*, Vol. 4, N.8, 2007. In this regard, potential ethical challenges in conducting randomized control trials should also be noted.

²² See for example, Rogers, Patricia, ‘Using Programme Theory for Complex and Complicated Programmes’, Royal Melbourne Institute of Technology, EES-UKES Conference London, 2006.

role in the joint initiative and the evaluator's task is to explain how each actor contributed to the observed results.

6.2 ...VERSUS CONTRIBUTION

In most cases, UNDP will adopt a contribution approach to outcome-level evaluations. A research approach called “contribution analysis” stems from the above considerations.²³ Contribution analysis does not expect to firmly establish causality but rather seeks to achieve a plausible association. While a fully-fledged application of contribution analysis methods and approaches may not be realistic in the context of an outcome-level evaluation, some simple principles may be followed. At a minimum, it is recommended that all outcome-level evaluations explicitly acknowledge and address the issue of contribution from the inception phase. As for practical approaches, the following tips can be considered:

- Understand the exact nature of UNDP initiatives (what did it do exactly?). As a specific feature of UNDP initiative: administrative or substantive contribution? How did this relate to activities by partners and other entities?
- Explain whenever possible the outcome model (how the programme is intended to work and what effects it is expected to generate). How did the outcome model take into account activities of partners and other entities?
- Document examples of the ‘value added’ by UNDP initiatives. For example: technical skills and expertise, conceptual frameworks and methods that were not available before. What was the value added of other partners and entities? Some limitations are:
 - Informants may not be fully aware of the factors that have contributed to results;
 - Informants may have a vested interest in attributing results to a particular initiative, partner or organization.
- Apply the ‘before and after’ criterion (understand which situation prevailed before the UNDP initiative and how it has changed to date) but try to check whether alternative explanations could be traced, including the contribution of other partners and entities.
- Contrast rival hypotheses, probe alternative explanations through interviews, document review and a reassessment of the logical chain. Contrast the case observed with other observed in the same country or even in other countries.
- When realistically feasible, consider the counterfactual question: what would have happened without UNDP's initiative? At the most basic level, different stakeholders can be asked this very question which, through further triangulation, may yield a strong perception of whether UNDP made a difference. A more rigorous approach would be to observe a case ‘without UNDP initiative’ that can be compared against one ‘with UNDP initiative’, assuming that it is possible to control for factors that would invalidate the comparison. This may prove to be a very rare case but still possible in some instances.

Examples of UNDP value added may include (but not be restricted to) the following:

- **Facilitating programme or project administration:** UNDP may add value where it facilitates a more effective process, e.g., through reducing procurement times or providing access to international consultants. When there are limitations in the capacity of the public sector to implement programmes, this may be *prima facie* an important contribution of UNDP. On the other hand, this may generate a “moral hazard” problem since it could reduce the incentive for ministries to push for reform of sub-optimal government procedures.
- **Technical:** UNDP may add value through the technical expertise of its consultants and advisers

²³ See, e.g., John Mayne, ‘Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly’, Mimeo, Office of the Auditor General of Canada, 2006.

and by introducing new approaches and methods. This raises the question of the absolute advantage of UNDP in some of these areas: is UNDP the best organization to address them? There is also a distinction between the value added of the project (e.g., through consultants) and that of UNDP (primarily through the country office but also headquarters and regional and sub-regional offices).

- **Knowledge management:** By systematically tracking experience and good practice, in the country or internationally, based on the experience of UNDP or other partners.
- **Neutrality.** UNDP may add value by providing space for different national and international stakeholders to raise issues that may be contentious or to work together in a context where cooperation may otherwise not be possible. It may thus contribute to facilitating dialogue and building consensus.

6.3 INTERPRETING CONTRIBUTIONS TO OUTCOMES

If outcomes appear not to have been achieved

The outcome-level evaluation should investigate four possible alternative explanations for why outcomes appear not to have been achieved, bearing in mind that there may be additional explanations:

- **The programme or project might not have been adequately implemented to produce the outcome.** The evaluation therefore needs to include an assessment of the adequacy of implementation (including fidelity to original design, skills of implementers, and duration and intensity of implementation). The evaluation needs to identify any particular individuals or sites where implementation has indeed been adequately implemented, and present the results from these separately in order to demonstrate possible contribution to outcomes.
- **The programme or project might only be successful in conjunction with other initiatives or factors** – that is, an entire ‘causal package’ might be needed to produce the outcome, consisting of other components (such as outputs from other activities, or favourable context). In the absence of this entire causal package, the initiative might not have achieved the outcome – but it could in other circumstances. The evaluation therefore needs to investigate the ‘causal package’ needed to produce the outcome, identify any particular individuals or sites where the whole causal package has indeed been present, and present the outcomes from these separately in order to demonstrate possible contribution to outcomes. (The ‘causal package’ is about all the sufficient and necessary factors being present to cause a change. For example, a dropped lighted match needs to land on material that burns to cause a fire.)
- **Comparisons between with and without the programme or project might not be appropriate.** The non-participants (whether these are individuals, households, communities or organizations) might have been better off than the programme participants to begin with, for instance, when the support is targeted to the most vulnerable villages or households. A lack of difference between participants and non-participants after the initiative might actually show that the project has improved outcomes for participants.
- **Programme or project outputs might have been delivered too recently to have an effect on the outcome.** In some cases, the time lag between the delivery of goods and services under a UNDP initiative and the use of, for example, newly acquired knowledge and skills, may be considerable. Different political or organizational factors may constrain targeted counterparts in building on new capacities, so an assessment of such constraining factors and their likelihood of being resolved may be needed.

If outcomes appear to have been achieved

Before concluding that the UNDP-supported initiatives have contributed to a set of outcomes, the evaluators should explore other possible explanations for the achievement of outcomes:

- **Alternative outcome pathway** – that is, the outcome may have been achieved but not through

the projected pathway outlined in the outcome model. In some cases, this might be that another agency has undertaken or funded activities which have led to the outcome; in other cases, historical developments will have made the major contribution to the outcome. Using the strategies of causal attribution by key informants and iterative contribution analysis will be useful in this situation.

- **Maturity of outcomes** – that is, the outcomes may have been in the making for a long time and resulted from initiatives conducted by the government, with or without the support of other organizations including UNDP, over a period of time far exceeding that of the UNDP initiative being evaluated. Again, a historical review of other possible contributing factors would need to be undertaken to ascertain the plausibility of UNDP’s contribution to the achievement of outcomes.
- **Comparisons between with and without the initiatives might not appropriate.** The non-participants (whether these are individuals, households, communities or organizations) might have started out at a lower level or had more access to resources. A difference between participants and non-participants at the end of the project might simply reflect pre-existing differences in their capabilities. The evaluation needs to check the comparability of the two groups in terms of important characteristics, some of which (such as motivation) might not be readily measurable.

Some simple, and by no means fool-proof, questions that may assist evaluators in determining the plausibility of UNDP’s contribution to outcomes include:²⁴

- Were outcomes achieved at an appropriate time after UNDP’s efforts began?
- Were the outcomes that were achieved indeed foreseen as part of UNDP initiatives?
- Were outcomes achieved in areas in which UNDP supported the conduct of activities?
- Was the achievement of outcomes most obvious in those areas in which UNDP was most involved?

In all cases, an analysis of why outcomes have or have not been achieved should be of particular interest to the evaluation. Only in clearly understanding the broader context, the multitude of factors involved and the distinct effects of UNDP-supported initiatives, will the evaluators be in a position to identify lessons and formulate recommendations on how to strengthen programme planning and implementation in the future.

6.4 DEALING WITH OUTCOMES AND INDICATORS THAT ARE NOT CLEARLY DEFINED

Not all stated outcomes are equally easy to evaluate. Consider, as an example, outcomes defined as: “Reduced discrimination of judicial authorities against indigenous people” or “Equitable and efficient administration of justice”. There may be two major problems with these definitions. First, their outcome is described in a vague manner. Second, they are long-term ideal outcomes and the evaluation may lack an adequate time-frame to assess the degree of achievement of the outcome.

Even when outcomes and indicators are over-ambitious or not precisely defined in official documents, it is possible to clarify the initial expectations of the key actors and partners through individual interviews or focus groups. This may shed light on the expected goals of the programme and on the occasionally divergent expectations from each actor.

One possibility for the team is to revisit the outcome model, working upwards from activities to outcomes, highlighting: (i) the activities supported by UNDP (this could group several projects and non-project activities), (ii) their immediate expected outputs, (iii) a set of expected intermediate expected outcomes and change processes that have been generated. In addition, (iv) a set of exogenous factors that may facilitate or hinder the process of achieving the outcomes. In doing so, the outcome model is fleshed out further, bridging the gaps in the causal logic of the original model. Ensuring

²⁴ See John Mayne, *ibid.*

adequate buy-in by different stakeholders would be key to ensuring the validity of the resulting outcome model.

Identifying the intermediate expected results may solve the time-frame issue (having final long-term outcomes without an adequate time window to observe them). Eliciting expectations from the main actor and partners may partly solve the problem of vague definition of outcomes in the official documents. Contrasting expected outputs and intermediate results with changes against observed outputs may help assess the progress made towards longer-term goals.

An example is shown in Table 1 for a hypothetical outcome: “Reducing discrimination of judicial authorities against indigenous people”. While mapping results in the context of a fully-fledged outcome model in written form may not be necessary in all cases, the conceptual scheme may be useful. In any case, eliciting the views and explanation of country office staff and other key stakeholders (e.g., government, civil society organizations) will be important, whether results are mapped out in written form or not.

Table 1. An Example of Mapping UNDP Activities, Outputs and Intermediate Results

Activities (projects and other activities supported by UNDP)	Expected outputs	Expected intermediate results	External factors	Expected progress towards achieving the outcome
Project 1. Training of magistrates	Magistrates are better informed on the legal framework on discrimination. Their attitude has changed	Implications of laws A and B discussed within justice departments	Justice departments understaffed	The legal framework has been significantly improved, judicial and public authorities are better prepared to handle cases, endowed with adequate resources It is easier for private citizens to present formal complaints (individually or through civil society organizations) More information is available through the media or activist groups The issue of discrimination entered the political discourse
Project 2. Law against discrimination	Advisory reports on the law against discrimination are available	The draft law includes new features that facilitate definition or sanctioning of discriminatory actions. It provides clear attribution of jurisdiction, budget, etc.	No budget/ special authority for enforcing law	
Project 3. Strengthening of civil society action groups	X civil society organizations have the necessary skills to prepare or file formal complaints to police or judicial authority	Civil society organizations are applying new techniques, knowledge, creating networks and increasing outreach	Action of civil society organizations and report dissemination and use has been restricted by authorities or by rival lobbying groups	
Non-project activities: e.g., round tables, workshops, consultation, advocacy	Round tables X Y and workshop Z reports are available that reflect discussions and proposed action points related to findings of the National Human Development Report	Knowledge and disseminated findings used as a basis for policy making by the government Civil society organizations, NGOs are re-orienting their objectives and projects taking into account these consultations		

Source: Adapted from ADR Team (2009)

6.5 ASSESSING THE EFFECTIVENESS OF UNDP PARTNERSHIPS IN CONTRIBUTING TO THE ACHIEVEMENT OF OUTCOMES

Figure 3 highlights how different known partners contribute to results that UNDP is working towards and how additional influences affect the achievement of results. Some entities may be engaged in a formal partnership with UNDP while other entities – irrespective of whether they might have

contributed significantly to the achievement of the result – may not be. Assessment of the effectiveness of the UNDP partnership should include considering whether it is likely that more progress would have been made towards the outcome if these other entities had been in formal partnership.

Any evaluation of partnerships hinges on the identification, description and, if feasible, attribution of the benefits or otherwise that accrue from this management arrangement. As with other components of outcome evaluations, the assessment of the partnership needs to be driven by key evaluation questions. These could include:

- What would have happened if a programme or project had not been delivered in partnership?
- In what ways did the outcomes benefit from the partnership?
- How did the partnership combine complementary expertise, knowledge and experience? How did this influence the outcome?
- What outcomes have occurred as a result of the following facets of successful partnerships being present or, conversely, not present?²⁵
 - A shared vision and purpose
 - Interdependence and a clear division of labour
 - Trust building and capacity development
 - Equitable distribution of costs and benefits
 - Equality and empowerment of weaker partners
 - Mutual adjustment and learning

The values that individual partners hold play a key part in the success or otherwise of the partnership. When evaluating the effectiveness of any partnership these have to be taken into account. When partners come together to plan a programme they come with varying expectations and have different ideas about the merit of a programme and their role in such an arrangement. Inevitably, they may need to compromise something in order for the partnership to form and the programme to go forward. It is unlikely that everyone will be happy. These differences can become problematic²⁶ as people are replaced in the partnership or as the understanding of the partnership changes over time.

Understanding and monitoring the value base of a partnership, therefore, is critical to an evaluation. As with most other facets of evaluation, the earlier the issue is tackled the better. A number of approaches have been employed to address this issue, including outcome mapping.²⁷ Moreover, concept mapping²⁸ and facilitation processes used at the beginning of a programme capture what participants value about the programme. Repeating this exercise occasionally throughout will track changes. When used at the beginning, the results of a concept mapping exercise can become the rich picture of the partnership. It will show differences within a group, what their expectations are, what outcomes they value or where there are areas of strong/weak agreement. The “most significant change” tool, an approach that uses stories of specific, typically individual, change, can also be used to track changes in values and identify areas where significant change has occurred,²⁹ giving indications of outcomes including unplanned and unintended outcomes.

²⁵ Elliot Stern, *Evaluating Partnerships*, in Andrés Liebenhal, Osvaldo Néstor Feinstein, Gregory K. Ingram, *Evaluation & Development: The Partnership Dimension*, World Bank, Washington D.C., 2005.

²⁶ Jacques Toulemonde, Celine Fontaine, Emile Laudren and Philippe Vincke, ‘Evaluation in Partnership: Practical Suggestions for Improving their Quality’, *Evaluation* Vol 4(2):171-188, Sage Publishing, London, 1998.

²⁷ IDRC, *Concept Mapping*, <www.idrc.ca/en/ev-26586-201-1-DO_TOPIC.html> (accessed May 2011).

²⁸ Refer Trochim’s *Concept Mapping* at <www.socialresearchmethods.net/mapping/mapping.htm> (accessed May 2011).

²⁹ Refer Davies and Dart on the *Most Significant Change* at <www.clearhorizon.com.au/wp-content/uploads/2008/12/dd-2005-msc_user_guide.pdf> (accessed May 2011).

6.6 COLLECTING AND ANALYSING DATA TO ASCERTAIN CONTRIBUTION TO OUTCOMES

Section 7.5 of the PME Handbook provides an overview of a range of methodological issues such as, inter alia, data collection and analysis. It highlights the importance of obtaining both primary and secondary data (involving, for example, the conduct of desk-based document research) and identifies different data collection methods, such as the conduct of interviews and various types of group reviews (including, for example, the conduct of focus groups discussions).

While obtaining data on efficiency-related issues is important and helps in understanding how or whether a UNDP initiative has contributed to the achievement of results, the primary focus of the data collection effort of an outcome-level evaluation should be on outcomes. In this regard, the availability of clear results frameworks and indicators that define outcomes and how they are to be measured is extremely helpful. Moreover, the availability of a corresponding monitoring system that tracks progress towards the achievement of outcomes is an obvious advantage. Where such systems have not been put in place, or are weak, evaluations may struggle to obtain adequate data, for example, through the use of proxy indicators, or may need to rely more heavily on anecdotal evidence.

Even where data on the achievement of outcomes does not readily exist, it is possible to ascertain changes in the behaviour of individuals or the conduct of business of an institution that could provide an indication of progress made towards the achievement of outcomes (e.g., improved quality of draft policy documents incorporating new concepts; institutionalized engagement with non-governmental organizations; etc.). Where such information is not quantifiable, which often is not in the case of UNDP initiatives, it is essential to triangulate data through different information sources – e.g., asking the same questions to a government official, a civil society representative, a party official, another development partner, etc. In this regard, it is important to bear in mind the biases of interviewees and to try to understand their interests and perspectives. However, beyond documenting the perceptions of interviewees, evaluators should always seek to obtain additional evidence in the form of documents, meeting minutes, logs, draft policy papers, photographs, etc. that would corroborate findings. Obtaining outcome-related evidence remains one of the most frequent shortcomings in evaluation. In designing and conducting outcome-level evaluations, every effort should be made to prioritize the collection of outcome-related data.

QUANTITATIVE DATA

Although most outcome-level evaluations tend to rely on qualitative data stemming from open-ended interviews, some quantitative data may also be available either through previous surveys conducted by UNDP (e.g., staff survey, partner survey), or through primary data collection. As an example, an evaluation team may decide to conduct an opinion survey focusing on an outcome or on a theme. The results may be presented in the report through either simple statistical tables or graphic charts.

Basic statistics may include (arithmetic) mean, median, mode, standard deviation and percentiles. Some inferential statistics on means difference may also be performed if required. Given the type of evaluation questions, outcome-level evaluations are unlikely to use more sophisticated statistical or econometric tools.³⁰ Selected graphs and shorter tables may be more effective in the main report: more complicated statistical tables may be better displayed in the annexes and key results shown in the main text.

It should be noted that qualitative data can also be used for quantitative analyses. For example, the results of a survey requesting respondents to assess results on the basis of qualitative criteria could be used to conduct quantitative analyses to rank data and determine the comparative statistical significance

³⁰ An easily accessible introduction is: Linda G. Morra Imas and Ray C. Rist, *The Road to Results Designing and Conducting Effective Development Evaluations*, The World Bank, Washington DC (USA), 2009.

of different responses.

QUALITATIVE DATA

Specific literature and dedicated software tools are available for qualitative data analysis. Simply put, working with qualitative data consists of three processes: data reduction, data display, and deriving conclusions.³¹ The three processes are not necessarily sequential: they may, and often do, overlap. Data reduction is the process of assembling, classifying, processing, simplifying the mass of data that is gathered through interactions with informants as well as from documents, photographic evidence and other qualitative sources. Data display is a graphic organization of the by-product of data reduction, so that patterns, recurrent observations, clusters, associations can be highlighted. Deriving conclusions consists of formulating preliminary findings on the meaning and implications of data display and continuous verification of the same with the available evidence so that statements can be produced that are corroborated by evidence and highlight implications of evidence and findings for the subject being evaluated.

Fully applying qualitative data collection and analysis tools can prove to be a labour-intensive task for outcome-level evaluations. However, discipline in data gathering and organization is a realistic goal for most outcome-level evaluations, and the following instruments and practices are recommended:

- Preparing an evaluation matrix linking criteria and sub-criteria to questions and sources: this can be used ex ante as a planning tool and ex post as a means to verify that enough evidence is available (see Annex 2);
- Preparing interview checklists, highlighting criteria and sub-criteria that are under review and operational questions. While providing flexibility for the interviewers to adapt to the context, checklists help ensure that key issues are not forgotten (see Annex 3);
- Preparing interview result sheets, so that key findings can be systematically recorded and tracked by team members (see Annex 4).

Data gathered and organized through the above instruments may be coded and findings organized through diagrams, double-entry tables, concept maps, flow charts or any other graphic support (An example of a table showing efficiency findings at project level is shown in Table 2).

Table 2. Examples of Assessment of Effectiveness: Evidence at Project Level		
Project	Progress towards the expected results	Sources
Sembrando	(+) NGO project catering to 10,000 families a year with a package of integrated social and productive activities. Able to reach rural towns up to an altitude of 2500 metres not covered by any State programme. (-) Limited coverage	PRODOC 2007 Participation of seeding specialist in discussion group Interview with the president of the Institute of Labour and Family (First Lady) and Vice-President; UNDP Official
Microfinance Services	(+) Setting up of mutual groups (UNICAS), development of small-grant and networking skills. The 170 networks have managed to capitalize 1 million soles. (-) Very localized achievements and experience and difficulties in expanding to other areas of the country. Enterprise Development Centre still not off the ground.	PRODOC 2004 Participation of two specialists from the COFIDE Inclusive Rural Enterprise Development Programme in discussion group Interview with the director-general of the Small-Grants Services Programme

³¹ Matthew B. Miles and A. M. Huberman, *Qualitative Data Analysis: An Expanded Sourcebook*, Sage Publications, Thousand Oaks, CA (USA), 1994.

7 TELLING THE OUTCOME STORY

Section 4.4. refers to the Guidance Note on the Quality Assessment System for Decentralized Evaluation Reports. Prior to launching outcome-level evaluations, and – at the very latest – prior to drafting outcome-level evaluation reports, programme units and commissioned evaluators should familiarize themselves with the quality criteria contained in the guidance related to the Quality Assessment.

7.1 THE RED THREAD

Evaluation Office reviews of the quality of outcome-level evaluation reports reveal similar challenges. In preparing reports, the following should be kept in mind:

- Build and tell a story. Sufficient attention needs to be given to explaining the background of the evaluation and the expected contribution to outcomes. The evaluation report should advance steadily from providing descriptive overviews, laying out the facts, to analysing UNDP's contribution based on evaluation criteria, to providing conclusions and recommendations.
- Keep apart analysis, judgement (conclusions) and recommendations. Different chapters can end up covering similar ground, leaving the reader unclear as to whether an issue is being described, analysed or judged. A clear report structure that separates out these elements can help in strengthening the outcome story. While some evaluators prefer a looser structure in order to retain the flexibility of telling a more readable, if not entertaining story, the evaluation's rigour tends to be enhanced by a clear report structure that differentiates between description, analysis, conclusions and recommendations (see Section 7.2).
- Differentiate clearly between relevance, effectiveness, efficiency and sustainability. Figure 3 in Section 5.2 explains how evaluation criteria, in particular effectiveness and efficiency, relate to different results levels. In telling the outcome story it is particularly important to make a clear distinction between outputs (UNDP's managerial responsibility – usually related to the efficiency criterion) and outcomes (the responsibility of primary stakeholders – usually related to the effectiveness criterion).
- Provide ample evidence, particularly at the outcome level. Obtaining evidence on UNDP's contribution to outcomes (let alone impact) is the most challenging aspect of outcome-level evaluations. This is also precisely where the greatest omissions tend to occur in outcome-level evaluation reports. To some extent, this is a function of the data collection process, which may not have focused sufficiently on obtaining solid data on outcomes (see Section 5.4). However, also in drafting reports, there is a tendency of evaluators to focus on aspects of the story that are easier to tell – i.e., accounts of how busy UNDP and its partners have been in implementing promising activities. What is more important in telling the outcome story is whether the promise of the

activities was fulfilled. What did UNDP activities and outputs lead to? What change did they help bring about? What evidence is there for such change?

7.2 SUGGESTED REPORT STRUCTURE

Annex 7 of the PME Handbook provides an annotated outline for evaluation reports (additional guidance is also available in UNEG’s Quality Checklist for Evaluation Reports).³² However, the structure of evaluation reports can vary depending on the nature of the analysis and the subject under review. Programme units, together with the evaluators they commission to conduct the evaluation, should decide on the most appropriate way in which to structure and present the report. The report structure should already be specified in the inception report. Outcome-level evaluation reports could be structured as proposed in Table 3.

Table 3. Suggested Report Structure	
Indicative sections	Description and comments (see also Annex 7 of the PME Handbook for more details)
Title and opening pages	
Table of contents	
List of acronyms and abbreviations	
Executive summary	This should be an extremely short chapter, highlighting the evaluation mandate, approach, key findings, conclusions and recommendations. Often, readers will only look at the executive summary. It should be prepared after the main text has been reviewed and agreed, and should not be circulated with draft reports.
Chapter 1: Introduction	Introduce the rationale for the evaluation, including mandate, purpose and objectives, outline the main evaluation issues including the expected contribution at the outcome level, address evaluability and describe the methodology to be used. Refer to the outcome model and evaluation matrix, to be attached as annexes.
Chapter 2: The development challenge	In addition to providing a general overview of historical trends and development challenges, specifically address the evaluation theme. Explain how the theme is addressed by government(s), and how it is reflected in national policies and strategies. Also provide information on the activities of other development partners in the area.
Chapter 3: UNDP response and challenges	Against the background of Chapter 2, explain what UNDP has done in this area (purely descriptive, not analytical). Provide the overarching outcome model, specifying the results frameworks for the programme, programme area or projects (if available), as well descriptions of some of the main UNDP activities, especially if they are going to be assessed later.
Chapter 4: Contribution to results	<p>Against the background of Chapters 2-3, analyse findings without repeating information already provided. Also, minimize the need to mention additional factual information regarding projects and programmes (these should be described in Chapter 3).</p> <p>Focus on providing and analysing evidence relating to the evaluation criteria. Preferably, structure the analysis on the basis of the main evaluation criteria:</p> <ul style="list-style-type: none"> ■ Relevance (of UNDP’s involvement and its approach) ■ Effectiveness (in contributing to the achievement of outcomes). Pay particular attention to this criterion, demonstrating how UNDP initiatives have, or have not, contributed to the achievement of outcomes. ■ Efficiency (in delivering outputs) ■ Sustainability (of the outcomes) <p>In addressing the evaluation criteria, the narrative should respond to the corresponding questions identified in the evaluation matrix and provide a summary analysis of the findings. Partnerships play a key role in ensuring that primary stakeholders achieve outcomes. As such, all evaluation criteria should cover relevant aspects of partnership – i.e., how were they relevant; how effective were they in contributing to the achievement of outcomes; how efficiently were they managed; and how sustainable are they?</p> <p>Where appropriate, discuss cross-cutting themes separately using the main evaluation criteria.</p> <p>Do not allow the discussion to drift into conclusions and recommendations.</p>

³² See <www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=607> (accessed May 2011).

Chapter 5: Conclusions and recommendations	<p>Conclusions are judgements based on evidence provided in Chapter 4. They are pitched at a higher level and are informed by an overall, comparative understanding of all relevant issues, options and opportunities.</p> <p>Do not provide new evidence or repeat evidence contained in earlier chapters.</p> <p>Recommendations should be derived from the evidence contained in Chapter 4. They may also, but need not necessarily, relate to conclusions. In line with the nature of the evaluation, some recommendations may be more strategic in nature while others may be more action-oriented. Recommendations should be important and succinct. Typically, do not provide more than five to ten.</p>
Annexes	

SAMPLE TABLE OF CONTENTS FOR AN INCEPTION REPORT

CONTENTS

1. INTRODUCTION*

- 1.1. Objective of the evaluation
- 1.2. Background and context
- 1.3. Scope of the evaluation

2. METHODOLOGY

- 2.1. Evaluation criteria and questions*
- 2.2. Conceptual framework
- 2.3. Evaluability
- 2.4. Data collection methods*
- 2.5. Analytical approaches
- 2.6. Risks and potential shortcomings*

3. PROGRAMME OF WORK*

- 3.1. Phases of work
- 3.2. Team composition and responsibilities
- 3.3. Management and logistic support
- 3.4. Calendar of work

ANNEXES

- 1. Terms of reference of the evaluation*
- 2. Evaluation matrix*
- 3. Stakeholder map*
- 4. Tentative outline of the main report*
- 5. Interview checklists/protocols*
- 6. Outcome model
- 7. Detailed responsibilities of evaluation team members
- 8. Reference documents
- 9. Document map
- 10. Project list
- 11. Project mapping
- 12. Detailed work plan

*As indicated in Section 4.3, the structure of inception reports may be adjusted depending on the scope of the evaluation. As a minimum, it is suggested that those elements marked with an asterisk be included in all inception reports.

EXAMPLE OF AN EVALUATION MATRIX

SAMPLE EVALUATION MATRIX				
Criteria/ Sub-criteria	(Examples of) questions to be addressed by outcome-level evaluation³³	What to look for	Data sources	Data collection methods
Relevance	<ul style="list-style-type: none"> ▪ Is the initiative aligned with national strategies? ▪ Is it consistent with human development needs and the specific development challenges in the country? 	<ul style="list-style-type: none"> ▪ How does the project align with national strategies (in specific thematic area)? ▪ How does the project address the human development needs of intended beneficiaries (poor, women, disadvantaged groups)? 	<ul style="list-style-type: none"> ▪ UNDP programme/project documents ▪ UNDP programme/project Annual Work Plans ▪ Programmes/projects/thematic areas evaluation reports ▪ Government's national planning documents ▪ Human Development Reports ▪ MDG progress reports ▪ Government partners progress reports ▪ Interviews with beneficiaries 	<ul style="list-style-type: none"> ▪ Desk reviews of secondary data ▪ Interviews with government partners ▪ Interviews with NGOs partners/service providers ▪ Interviews with funding agencies and other UNCT ▪ Interview with civil societies in the concerned sector ▪ Interviews with political parties leader ▪ Interviews with related parliamentary committees ▪ Related Constitutional bodies such as Human Rights, Women Rights, etc. ▪ Field visits to selected projects
	<ul style="list-style-type: none"> ▪ Are UNDP approaches, resources, models, conceptual framework relevant to achieve the planned outcome? ▪ Are they sufficiently sensitive to the conflict-post-conflict environment in the country? ▪ To what extent has UNDP adopted participatory approaches in planning and delivery of the initiative and what has been feasible in the country context? 	<ul style="list-style-type: none"> ▪ What analysis was done in designing the project? ▪ To what extent have indigenous peoples, women, conflict-displaced peoples, and other stakeholders been involved in project design? ▪ Are the resources allocated sufficient to achieve the objectives of the project? 	<ul style="list-style-type: none"> ▪ UNDP staff ▪ Development partners (UNICEF, WFP, IFAD, UNV, UN Women, Norwegian Embassy, CIDA, Danish Embassy, Finish Embassy, DFID, GIZ etc.) ▪ Government partners involved in specific results/thematic areas ▪ Concerned civil society partners ▪ Concerned associations and federations 	<ul style="list-style-type: none"> ▪ Interviews with UNDP staff, development partners and government partners, civil society partners, associations, and federations

³³ Note: These questions should, in practice, be covered in the protocols used to guide interviews (see Annex 3)

SAMPLE EVALUATION MATRIX

Criteria/ Sub- criteria	(Examples of) ques- tions to be addressed by outcome-level evaluation	What to look for	Data sources	Data collection methods
Effectiveness	<ul style="list-style-type: none"> ▪ Did the project or programme implementation contribute towards the stated outcome? Did it at least set dynamic changes and processes that move towards the long-term outcomes? ▪ How does UNDP measure its progress towards expected results/outcomes in a context of flux? 	<ul style="list-style-type: none"> ▪ What outcomes does the project intend to achieve? ▪ What outputs has the project achieved? ▪ What percentage of the project results at the output level has been achieved? ▪ What changes can be observed as a result of these outputs? ▪ In addition to UNDP initiatives, what other factors may have affected the results? ▪ What were the unintended results (+ or -) of UNDP initiatives? 	<ul style="list-style-type: none"> ▪ Project/programme/thematic areas evaluation reports ▪ Progress reports on projects ▪ UNDP staff ▪ Development partners ▪ Government partners ▪ Beneficiaries 	<ul style="list-style-type: none"> ▪ Desk reviews of secondary data ▪ Interviews with government partners, development partners, UNDP staff, civil society partners, associations, and federations ▪ Field visits to selected projects
	<ul style="list-style-type: none"> ▪ How broad are the outcomes (e.g., local community, district, regional, national)? ▪ Are UNDP's efforts concentrated in regions/districts of greatest need? 	<ul style="list-style-type: none"> ▪ Are the results of the project intended to reach local community, district, regional or national level? 	<ul style="list-style-type: none"> ▪ Evaluation reports ▪ Progress reports on projects 	<ul style="list-style-type: none"> ▪ Desk reviews of secondary data
	<ul style="list-style-type: none"> ▪ Who are the main beneficiaries? ▪ To what extent do the poor, indigenous groups, women, Dalits, and other disadvantaged and marginalized groups benefit? 	<ul style="list-style-type: none"> ▪ Who are the target beneficiaries and to what extent have they been reached by the project? ▪ How have the particular needs of disadvantaged groups been taken into account in the design and implementation, benefit sharing, monitoring and evaluation of the project/programme? ▪ How far has social inclusion been taken into account in the project/programme? ▪ How far has the regional context (least developed region) been taken into consideration while selecting the project/programme? 	<ul style="list-style-type: none"> ▪ Programme documents ▪ Annual Work Plans ▪ Evaluation reports ▪ MDG progress reports ▪ Human Development Reports 	<ul style="list-style-type: none"> ▪ Desk reviews of secondary data

SAMPLE EVALUATION MATRIX

Criteria/Sub-criteria	(Examples of) questions to be addressed by outcome-level evaluation	What to look for	Data sources	Data collection methods
Efficiency	<ul style="list-style-type: none"> Has the project or programme been implemented within deadline and cost estimates? Have UNDP and its partners taken prompt actions to solve implementation issues? What impact has political instability had on delivery timelines? 	<ul style="list-style-type: none"> Have there been time extensions on the project? What were the circumstances giving rise to the need for time extension? Has there been over-expenditure or under-expenditure on the project? What mechanisms does UNDP have in place to monitor implementation? Are these effective? 	<ul style="list-style-type: none"> Programme documents Annual Work Plans Evaluation reports ATLAS reports Government partners Development partners UNDP staff (Programme Implementation Support Unit) 	<ul style="list-style-type: none"> Desk reviews of secondary data Interviews with government partners and development partners
	<ul style="list-style-type: none"> Were UNDP resources focused on the set of activities that were expected to produce significant results? Was there any identified synergy between UNDP initiatives that contributed to reducing costs while supporting results? How has the existence of the Project Implementation Support Unit assisted the efficiency of programme delivery 	<ul style="list-style-type: none"> Are resources concentrated on the most important initiatives or are they scattered/spread thinly across initiatives? 	<ul style="list-style-type: none"> Programme documents Annual Work Plans Evaluation reports ATLAS reports Government partners Development partners UNDP staff (Programme Implementation Support Unit) 	<ul style="list-style-type: none"> Desk reviews of secondary data Interviews with government partners and development partners
Sustainability	<ul style="list-style-type: none"> Were initiatives designed to have sustainable results given the identifiable risks? Did they include an exit strategy? How does UNDP propose to exit from projects that have run for several years? 	<ul style="list-style-type: none"> Does/did the project have an exit strategy? To what extent does the exit strategy take into account the following: <ul style="list-style-type: none"> Political factors (support from national authorities) Financial factors (available budgets) Technical factors (skills and expertise needed) Environmental factors (environmental appraisal) 	<ul style="list-style-type: none"> Programme documents Annual Work Plans Evaluation reports 	<ul style="list-style-type: none"> Desk reviews of secondary data
	<ul style="list-style-type: none"> What issues emerged during implementation as a threat to sustainability? What corrective measures were adopted? How has UNDP addressed the challenge of building national capacity in the face of high turnover of government officials? 	<ul style="list-style-type: none"> What unanticipated sustainability threats emerged during implementation? What corrective measures did UNDP take? 	<ul style="list-style-type: none"> Evaluation reports Progress reports UNDP programme staff 	<ul style="list-style-type: none"> Desk reviews of secondary data Interview UNDP programme staff
	<ul style="list-style-type: none"> How has UNDP approached the scaling up of successful pilot initiatives and catalytic projects? Has the government taken on these initiatives? Have donors stepped in to scale up initiatives? 	<ul style="list-style-type: none"> What actions have been taken to scale up the project if it is a pilot initiative? 	<ul style="list-style-type: none"> Evaluation reports Progress reports UNDP programme staff 	<ul style="list-style-type: none"> Desk reviews of secondary data Interview UNDP programme staff

SAMPLE EVALUATION MATRIX				
Criteria/Sub-criteria	(Examples of) questions to be addressed by outcome-level evaluation	What to look for	Data sources	Data collection methods
SAMPLE QUESTIONS RELATING TO THE PROMOTION OF UN VALUES FROM A HUMAN DEVELOPMENT PERSPECTIVE				
Supporting policy dialogue on human development issues	<ul style="list-style-type: none"> ▪ To what extent did the initiative support the government in monitoring achievement of MDGs? 	<ul style="list-style-type: none"> ▪ What assistance has the initiative provided supported the government in promoting human development approach and monitoring MDGs? Comment on how effective this support has been. 	<ul style="list-style-type: none"> ▪ Project documents ▪ Evaluation reports ▪ HDR reports ▪ MDG reports ▪ National Planning Commission ▪ Ministry of Finance 	<ul style="list-style-type: none"> ▪ Desk review of secondary data ▪ Interviews with government partners
Contribution to gender equality	<ul style="list-style-type: none"> ▪ To what extent was the UNDP initiative designed to appropriately incorporate in each outcome area contributions to attainment of gender equality? ▪ To what extent did UNDP support positive changes in terms of gender equality and were there any unintended effects? 	<ul style="list-style-type: none"> ▪ Provide example(s) of how the initiative contributes to gender equality. ▪ Can results of the programme be disaggregated by sex? 	<ul style="list-style-type: none"> ▪ Project documents ▪ Evaluation reports ▪ UNDP staff ▪ Government partners ▪ Beneficiaries 	<ul style="list-style-type: none"> ▪ Desk review of secondary data ▪ Interviews with UNDP staff and government partners ▪ Observations from field visits
Addressing equity issues (social inclusion)	<ul style="list-style-type: none"> ▪ How did the UNDP initiative take into account the plight and needs of vulnerable and disadvantaged to promote social equity, for example, women, youth, disabled persons? 	<ul style="list-style-type: none"> ▪ Provide example(s) of how the initiative takes into account the needs of vulnerable and disadvantaged groups, for example, women, youth, disabled persons. ▪ How has UNDP programmed social inclusion into the initiative? 	<ul style="list-style-type: none"> ▪ Project documents ▪ Evaluation reports ▪ UNDP staff ▪ Government partners ▪ Beneficiaries 	<ul style="list-style-type: none"> ▪ Desk review of secondary data ▪ Interviews with UNDP staff and government partners ▪ Observations from field visits

SEMI-STRUCTURED INTERVIEW GUIDE

Interviews for outcome-level evaluations can be structured or semi-structured. Structured interviews are usually administered as formal questionnaires. In this case, data collection is carefully organized around pre-structured topics for investigation (such as the criteria and sub-criteria found in the evaluation framework), with corresponding ranking scales that can later be summarized or quantified if necessary.

The majority of interviews conducted for outcome-level evaluations are usually semi-structured. In order to conduct quality semi-structured interviews, some form of checklist, guide or prompt sheet is necessary, although it is intended to be flexible. The evaluation criteria and sub-criteria can help organize these types of interviews, but the questions asked are more detailed and specific than the general ones found in the evaluation framework. Also, additional interview management skills, including a sensitive and ethical approach, are needed by the interviewers so that key topics are covered in a wide-ranging conversation. The interview guides can also be used to help write up data on interview summary sheets (see below). Where time allows, the list of specific questions from semi-structured interview guides can be provided to participants in advance.

The time spent on constructing a clear, detailed interview guide in advance of the main data collection mission is important in allowing the evaluation team to organize the raw data and conduct qualitative data analysis. Below are examples of questions that may be asked. Questions should be adapted selectively depending on the informants to be interviewed and the themes to be discussed.

QUESTIONS RELATING TO THE MAIN EVALUATION CRITERIA

RELEVANCE

- How has UNDP's initiative supported or contributed to relevant national policies or strategies? In which areas? Via which types of projects or other forms of advice? Can you provide specific examples of good contributions?
- Has UNDP followed good practices in its development work? Why or why not? Can you provide specific examples of where UNDP approaches were appropriate, well-needed and fit with national efforts? Where there were problems or challenges?
- Did the UNDP project/programme support the government's development goals and strategies?
- Is UNDP's project/programme aligned with government plans, procedures, and policies?
- Did UNDP design the right project/programme to meet the needs of the stakeholders? Why or why not? What could have been done differently?
- Were there obvious or critical gaps that the UNDP project/programme did not address? What were they?
- Did the UNDP project/programme respond to significant changes happening in the local/country/regional/global context? In what ways did adaptation take place? What trade-offs were there (that you know of) between short-term response and support for longer-term initiatives? What could have been done differently?
- Was the project/programme adequately adapted to changes in local conditions? Provide examples.

EFFECTIVENESS

- What activities have been undertaken under the UNDP project(s) you are familiar with? What short-term outputs have been produced? What longer-term effects were produced?

QUESTIONS RELATING TO THE MAIN EVALUATION CRITERIA

- Was the project linked to government activities or activities of other agencies? How well were they coordinated?
- Were there significant unexpected results or achievements that you know of? What were they, at different levels?
- What has been the scope or reach of the projects and their benefits? Who has been affected (either positively or negatively)? To what extent were men and women affected differently?
- Has the UNDP project made a difference via this project? Within a limited area or in this thematic area or sector overall? To whom? In what way?
- Did the project/programme have a capacity development objective? Were needs identified? Were some left out?
- Has the project/programme been effective in developing capacities of the men and women involved?
- Who have been the main beneficiaries of UNDP's work in the project you are familiar with? To what extent did men and women benefit differently? At what level (ministry-wide, specific departments or units, others for whom services or benefits were indirectly provided)?
- Have any benefits been realized via this project for the poor, disadvantaged groups, rural communities, women, or others with specialized needs in the country?
- Has any significant event occurred affecting project/programme outcomes? How well did UNDP adapt to these circumstances or changes?

EFFICIENCY

- To your knowledge, how well did UNDP use its human and financial resources? Were resources used well? Were funds received on time? Why or why not? Were projects approved and launched in a timely fashion? Why or why not? Please provide specific examples.
- Are UNDP procedures and processes easy to understand? What types of reporting were required, and were they submitted on a regular basis? Why or why not? Did the plans and reports required from UNDP add to the burden of implementing partners or beneficiaries in any way? Please provide examples.
- Are you familiar with the monitoring and evaluation arrangements for UNDP's project/programme? How well did M&E work (in your opinion) and what effects did they have on the project in which you were involved?
- How would you describe UNDP's cooperation with other partners, including other Country Team partners and bilateral or multilateral donors, that were important to this initiative? What went well? What could have been done better?

SUSTAINABILITY

- Were the project/programme achievements maintained and expanded over time?
- What was learned from the UNDP-assisted project/programme? Have any knowledge and lessons been used?
- Would you say there is a high degree of national/local ownership of UNDP-assisted projects/programmes? Why or why not? How could national ownership be improved?
- What indications are there that the government, civil society entities or other partners will continue to support, or even upscale, this or similar initiatives?

MDGs

- How did UNDP contribute to the achievement of the MDGs in the country? What specific initiatives, projects, or advice was UNDP able to offer towards fulfilling MDG aims? How has this made a difference to the country's overall development and/or commitment to the MDGs?

Gender

- Was the project or programme based on a gender analysis, targets and resources? What effects were realized in terms of gender equality, if any (provide examples)?
- Were women and men distinguished in terms of participation and benefits within specific projects? Were there clear gender strategies provided and/or technical advice on gender mainstreaming issues?

Equity

- Were specific vulnerable groups helped by UNDP's initiative? If so, how (provide examples)?
- Were the rights of indigenous peoples addressed in the project/programme? If so, how?

PREPARING INTERVIEW SUMMARY SHEETS

An interview summary sheet is a standardized approach to creating a short, typed summary of each interview conducted. It can be used by team members during and after data collection to help them compile and organize data prior to or at the same time as coding. The data summary sheets can also be kept together in a centralized place so that the original data can be rechecked if necessary later on in the evaluation process.

Interview summary sheets can be organized in advance so that data is actually coded to some extent as it is entered into the sheet. The evaluation team would review and sort/code their raw notes, and then select which portions of the raw data would provide information or evidence around the key criteria and sub-criteria. This process is made easier if an orderly interview guide is used which roughly follows the same order of topics or themes. The summary sheets can also be coded further if there is information that does not fall neatly into the pre-arranged categories for the data.

Interview summary sheets should be brief, so that just a few relevant statements or points are extracted under the main topics. Direct quotes can be used as well. Information may not be obtained from every interview on each interview question, so there may be gaps in the summary sheet. The team can assess over time if any gaps exist and what additional information may need to be gathered on specific questions (see also the data accounting sheet below).

Basic contact data (from the interview – enter on the summary sheet)

Institution	Department	Interviewee/Position
Date	Time	Location
Other persons present	Team members present	Notes by
Project/programme focus		

Interview Summary Sheet	
Relevance <ul style="list-style-type: none"> ▪ (itemized key questions/issues) ▪ ▪ 	Interviewer memos/notes
Effectiveness <ul style="list-style-type: none"> ▪ (itemized key questions/issues) ▪ ▪ 	Interviewer memos/notes
Efficiency <ul style="list-style-type: none"> ▪ (itemized key questions/issues) ▪ ▪ 	Interviewer memos/notes
Sustainability <ul style="list-style-type: none"> ▪ (itemized key questions/issues) ▪ ▪ 	Interviewer memos/notes
Promotion of UN Values <ul style="list-style-type: none"> ▪ (itemized key questions/issues) ▪ ▪ 	Interviewer memos/notes
Other Observations	

