

# ASSURING THE QUALITY OF EVALUATION DESIGN AND METHODOLOGY

## CHAPTER 7

This chapter describes key components of quality evaluation design and elements of quality evaluation reports to help UNDP managers, evaluation managers and partners carry out effective **quality assurance of the evaluation process and products**. It is intended to enhance knowledge about available methods and tools to ensure that key evaluation products—such as the ToR, evaluation design and reports—meet the quality criteria as defined by the governing norms, standards and policies. This chapter also aims to help external evaluators understand the quality standards that are expected of evaluations in UNDP.

While external evaluators are responsible for refining the methodology and carrying out the evaluation, overall design and methodology is largely determined by the information provided in the evaluation ToR. Therefore, those responsible for drafting the ToR can refer to this chapter for information on key elements of the design and the role of stakeholders, defining the context, the evaluation purpose, and focusing the evaluation before drafting and finalizing the ToR. Quality assurance considerations for the evaluation methodology (Section 7.5) are also covered in this chapter.

## 7.1 OVERVIEW

Developing a quality evaluation design involves a thorough understanding of what is being evaluated (**the initiative and its context**) and making decisions about the following key elements and how each will contribute to valid and useful evaluation results:

- The **purpose** of the evaluation
- The **focus** of the evaluation, that is, the **key questions** that the evaluation seeks to answer
- The **sources and methods** for obtaining information that is credible and defensible
- The procedures that will be used to **analyse and interpret data and report results**

- The **standards** that must be reached for the initiative to be considered successful
- The **evidence** that will be used to indicate how the initiative has performed and demonstrate its results (outputs and outcomes)

## ROLE OF STAKEHOLDERS

Stakeholders play an important role in designing and carrying out a quality evaluation. Stakeholders include individuals and groups that have a vested interest in the initiative or the results of the evaluation. Their involvement at all stages of the evaluation—including focusing the evaluation, shaping the questions to be addressed, identifying credible sources of evidence, reviewing findings and assisting in their interpretation—increases the credibility, potential usefulness and sustainability of evaluation results. Typically, stakeholders can be divided into three major categories, which are not mutually exclusive:

- **Those involved in the implementation of the initiative**—for example, donors, collaborators, strategic partners, administrators, managers and staff
- **Those served or affected by the initiative**—for example, intended beneficiaries, relevant organizations and agencies, government officials, advocacy groups, skeptics, opponents and staff of the implementing or competing agencies
- **Primary users of the evaluation**—for example, the specific persons in a position to do or decide something regarding the initiative, such as donors, UNDP programmatic counterparts (programme or outcome board) and partners in joint evaluation

The level of involvement of stakeholders will vary among evaluations. When designing an evaluation, it is important for the commissioning programme unit to identify stakeholders early and draw upon their knowledge as the evaluation design is shaped, starting with their meaningful involvement in developing the ToR. This is particularly critical for joint evaluations, in which case partners involved in the evaluation should be involved in all phases of developing the evaluation design.

## 7.2 DEFINING THE CONTEXT

UNDP evaluations support the UNDP human development focus “to help people build a better life” by generating knowledge about what works, why and under what circumstances. Therefore, quality evaluations not only focus on the attainment of outputs and outcomes but also assess how initiatives adapt to the contexts in which they operate and how and why they contribute to outputs and outcomes.

Evaluations must be conceived and designed with a thorough understanding of the initiative and the context within which it operates. The UNDP commissioning unit and relevant stakeholders who are engaged in drafting of ToRs (see Annex 3) are responsible for articulating necessary information for evaluators to have a good understanding of the initiative, the evaluation context, focus and purpose of the evaluation, and key questions to be addressed in the evaluation.

## UNDERSTANDING THE INITIATIVE

To produce credible information that will be useful for decision makers, evaluations must be designed with a clear understanding of the initiative, how it operates, how it was intended to operate, why it operates the way it does and the results that it produces. It is not enough to know what worked and what did not work (that is, whether intended outcomes or outputs were achieved or not). To inform action,

**Table 27. Key aspects of the initiative**

Key Aspect	Questions to Ask
<b>Demand</b>	What is the need or demand for the initiative? What problem or development opportunity is the initiative intended to address?
<b>Beneficiaries</b>	Who are the beneficiaries or targets of the initiative? Who are the individuals, groups or organizations, whether targeted or not, that benefit directly or indirectly from the development initiative?
<b>Scope</b>	What is the scope of the initiative in terms of geographic boundaries and number of intended beneficiaries?
<b>Outputs and Outcomes</b>	What changes (outcomes) or tangible products and services (outputs) are anticipated as a result of the initiative? What must the project, programme or strategy accomplish to be considered successful? How do the intended outcomes link to national priorities, UNDAF priorities and corporate Strategic Plan goals?
<b>Activities</b>	What activities, strategies or actions, both planned and unplanned, does the programme take to effect change?
<b>Theory of Change or Results/ Outcome Map</b>	What are the underlying rationales and assumptions or theory that defines the relationships or chain of results that lead initiative strategies to intended outcomes? What are the assumptions, factors or risks inherent in the design that may influence whether the initiative succeeds or fails?
<b>Resources</b>	What time, talent, technology, information and financial resources are allocated to the effort?
<b>Stakeholders and Partnership Strategy</b>	Who are the major actors and partners involved in the programme or project with a vested interest? What are their roles, participation and contributions—including financial resources, in-kind contributions, leadership and advocacy—including UN organizations and others? How was the partnership strategy devised? How does it operate?
<b>Phase of Implementation</b>	How mature is the project or programme, that is, at what stage or year is the implementation? Is the implementation within the planned course of the initiative? Is the programme mainly engaged in planning or implementation activities?
<b>Modifications from Original Design</b>	What, if any, changes in the plans and strategies of the initiative have occurred over time? What are the potential implications for the achievement of intended results?
<b>Evaluability</b>	Can the project or programme as it is defined be evaluated credibly? Are intended results (outputs, outcomes) adequately defined, appropriate and stated in measurable terms, and are the results verifiable? Are monitoring and evaluation systems that will provide valid and reliable data in place?
<b>Cross-cutting Issues</b>	To what extent are key cross-cutting issues and UN values intended to be mainstreamed and addressed in the design, implementation and results?

evaluations must provide credible information about why an initiative produced the results that it did and identify what factors contributed to the results (both positive and negative). Understanding exactly what was implemented and why provides the basis for understanding the relevance or meaning of project or programme results.

Therefore, evaluations should be built on a thorough understanding of the initiative that is being evaluated, including the expected results chain (inputs, outputs and intended outcomes), its implementation strategy, its coverage, and the key assumptions and risks underlying the Results Map or Theory of Change. The questions outlined in Table 27 should be understood by the evaluators in conducting the evaluation.

## THE EVALUATION CONTEXT

The evaluation context concerns two interrelated sets of factors that have bearing on the accuracy, credibility and usefulness of evaluation results:<sup>48</sup>

- **Social, political, economic, demographic and institutional factors, both internal and external**, that have bearing on how and why the initiative produces the results (positive and negative) that it does and the sustainability of results.
- **Social, political, economic, demographic and institutional factors within the environment and time frame of the evaluation** that affect the accuracy, impartiality and credibility of the evaluation results.

Examining the internal and external factors within which a development initiative operates helps explain why the initiative has been implemented the way it has and why certain outputs or outcomes have been achieved and others have not. Assessing the

### Box 38. Guiding questions for defining the context<sup>49</sup>

- What is the operating environment around the project or programme?
- How might factors such as history, geography, politics, social and economic conditions, secular trends and efforts of related or competing organizations affect implementation of the initiative strategy, its outputs or outcomes?
- How might the context within which the evaluation is being conducted (for example, cultural language, institutional setting, community perceptions, etc.) affect the evaluation?
- How does the project or programme collaborate and coordinate with other initiatives and those of other organizations?
- How is the programme funded? Is the funding adequate? Does the project or programme have finances secured for the future?
- What is the surrounding policy and political environment in which the project or programme operates? How might current and emerging policy alternatives influence initiative outputs and outcomes?

48 Evaluation results refer to the end product of the evaluation—the sum of information the evaluation generates for users, including findings, conclusions, recommendations and lessons.

49 Guiding questions within this chapter are illustrative, not exhaustive, of the possible questions that could be considered.

initiative context may also point to factors that impede the attainment of anticipated outputs or outcomes, or make it difficult to measure the attainment of intended outputs or outcomes or the contribution of outputs to outcomes. In addition, understanding the political, cultural and institutional setting of the evaluation can provide essential clues for how best to design and conduct the evaluation to ensure the impartiality, credibility and usefulness of evaluation results.

### 7.3 THE EVALUATION PURPOSE

All evaluations start with a purpose, which sets the direction. Without a clear and complete statement of purpose, an evaluation risks being aimless and lacking credibility and usefulness. Evaluations may fill a number of different needs. The statements of purpose should make clear the following:

- Why the evaluation is being conducted and at that particular point in time
- Who will use the information
- What information is needed
- How the information will be used

The purpose and timing of an evaluation should be determined at the time of developing an evaluation plan (see Chapter 3 for more information). The purpose statement can be further elaborated at the time a ToR for the evaluation is drafted to inform the evaluation design.

**TIP** **Sample Purpose Statement**—“This project evaluation is being conducted at the request of the national government and UNDP to provide information about the status of project implementation to ensure accountability for the expenditures to date and the delivery of outputs and so that managers can make midcourse corrections as appropriate.”

### 7.4 FOCUSING THE EVALUATION

#### EVALUATION SCOPE

The evaluation scope narrows the focus of the evaluation by setting the boundaries for what the evaluation will and will not cover in meeting the evaluation purpose. The scope specifies those aspects of the initiative and its context that are within the boundaries of the evaluation. The scope defines, for example:

- The unit of analysis to be covered by the evaluation, such as a system of related programmes, policies or strategies, a single programme involving a cluster of projects, a single project, or a subcomponent or process within a project
- The time period or phase(s) of the implementation that will be covered
- The funds actually expended at the time of the evaluation versus the total amount allocated

- The geographical coverage
- The target groups or beneficiaries to be included

The scope helps focus the selection of evaluation questions to those that fall within the defined boundaries.

## EVALUATION OBJECTIVES AND CRITERIA

Evaluation objectives are statements about what the evaluation will do to fulfil the purpose of the evaluation. Evaluation objectives are based on careful consideration of: the types of decisions evaluation users will make; the issues they will need to consider in making those decisions; and what the evaluation will need to achieve in order to contribute to those decisions. A given evaluation may pursue one or a number of objectives. The important point is that the objectives derive directly from the purpose and serve to focus the evaluation on the decisions that need to be made.

**TIP** Possible project evaluation objectives—“To assess the status of outputs; to assess how project outputs are being achieved; to assess the efficiency with which outputs are being achieved.”

**Evaluation criteria** help focus evaluation objectives by defining the standards against which the initiative will be assessed. UNDP evaluations generally apply the following evaluation criteria to help focus evaluation objectives: relevance, effectiveness, efficiency, sustainability and impact of development efforts.<sup>50</sup>

**Relevance** concerns the extent to which a development initiative and its intended outputs or outcomes are consistent with national and local policies and priorities and the needs of intended beneficiaries. Relevance also considers the extent to which the initiative is responsive to UNDP corporate plan and human development priorities of empowerment and gender equality issues. Relevance concerns the congruency between the perception of what is needed as envisioned by the initiative planners and the reality of what is needed from the perspective of intended beneficiaries. It also incorporates the concept of responsiveness—that is, the extent to which UNDP was able to respond to changing and emerging development priorities and needs in a responsive manner.

An essential sub-category of relevance is the criteria of **appropriateness**, which concerns the cultural acceptance as well as feasibility of the activities or method of delivery of a development initiative. While relevance examines the importance of the initiative relative to the needs and priorities of intended beneficiaries, appropriateness examines whether the initiative as it is operationalized is acceptable and is feasible within the local context. For example, an initiative may be relevant in that it addresses a need that intended beneficiaries perceive to be important, but inappropriate because

50 OECD, ‘DAC Criteria for Evaluating Development Assistance’, Development Assistance Committee. Available at: [http://www.oecd.org/document/22/0,2340,en\\_2649\\_34435\\_2086550\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/22/0,2340,en_2649_34435_2086550_1_1_1_1,00.html).

the method of delivery is incongruent with the culture or not feasible given geographic or other contextual realities. In applying the criterion of relevance, evaluations should explore the extent to which the planning, design and implementation of initiatives takes into account the local context.

**Effectiveness** is a measure of the extent to which the initiative's intended results (outputs or outcomes) have been achieved or the extent to which progress toward outputs or outcomes has been achieved.

Evaluating effectiveness in project evaluations involves an assessment of cause and effect—that is, attributing observed changes to project activities and outputs—for example, the extent to which changes in the number of voters can be **attributed** to a voter education project. Assessing effectiveness in outcome evaluations will more likely examine UNDP contributions toward intended outcomes. For example, an outcome evaluation might explore the extent to which the observed outputs from a voter education project—along with other UNDP outputs and those of other partners, such as professionalizing the electoral administration—contributed towards achieving stated outcomes relating to inclusive participation measured by international observers and other reputable experts.

Assessing effectiveness involves three basic steps:

1. Measuring change in the observed output or outcome
2. Attributing observed changes or progress toward changes to the initiative (project evaluation) or determining UNDP contributions toward observed changes
3. Judging the value of the change (positive or negative)

**Efficiency** measures how economically resources or inputs (such as funds, expertise and time) are converted to results. An initiative is efficient when it uses resources appropriately and economically to produce the desired outputs. Efficiency is important in ensuring that resources have been used appropriately and in highlighting more effective uses of resources.

As the nature and primary purposes of project and outcome evaluations differ, the application of criterion will also differ. For example, in assessing efficiency, a project evaluation might explore the extent to which resources are being used to produce the intended outputs and how resources could be used more efficiently to achieve the intended results. An outcome evaluation may involve estimates of the total UNDP investment (all projects and soft assistance) toward a given development outcome. The application of this criterion, particularly in UNDP outcome evaluations, poses a challenge as the nature of UNDP initiatives (for example, soft assistance), does not always lend itself to conventional efficiency indicators. In such cases, some analysis of delivery rates, the reasons some initiatives are implemented more quickly than others, and overall management ratios at the programme level might be considered. It is also important to assess how the partnership strategy has influenced the efficiency of UNDP initiatives through cost-sharing measures and complementary activities.

**Sustainability** measures the extent to which benefits of initiatives continue after external development assistance has come to an end. Assessing sustainability involves evaluating the extent to which relevant social, economic, political, institutional and other conditions are present and, based on that assessment, making projections about the national capacity to maintain, manage and ensure the development results in the future.

For example, an assessment of sustainability might explore the extent to which:

- A sustainability strategy, including capacity development of key national stakeholders, has been developed or implemented.
- There are financial and economic mechanisms in place to ensure the ongoing flow of benefits once the assistance ends.
- Suitable organizational (public or private sector) arrangements have been made.
- Policy and regulatory frameworks are in place that will support continuation of benefits.
- The requisite institutional capacity (systems, structures, staff, expertise, etc.) exists.

**Impact** measures changes in human development and people's well-being that are brought about by development initiatives, directly or indirectly, intended or unintended. Many development organizations evaluate impact because it generates useful information for decision making and supports accountability for delivering results. At times, evaluating impact faces challenges: Confirming whether benefits to beneficiaries can be directly attributed to UNDP support can be difficult, especially when UNDP is one of many contributors. However, the impact of UNDP initiatives should be assessed whenever their direct benefits on people are discernible.

In general, applying the following most commonly applied criteria—**relevance, effectiveness, efficiency, sustainability** and **impact**—in combination will help to ensure that the evaluation covers the most critical areas of the initiative. However, not all criteria are applicable, or equally applicable, to every evaluation. Different criteria may need to be applied in unique cases. In determining which criteria to apply, consider the type of evaluation and the contributions of the information to the purpose relative to the cost (use of evaluation resources). For example, evaluations of humanitarian and conflict programming may additionally apply the criteria of connectedness, coherence, coverage and coordination.<sup>51</sup> Box 39 outlines guiding questions to help define evaluation criteria and associated evaluation questions.

### Box 39. Guiding questions for defining evaluation criteria

- To what extent does the criterion inform the purpose of the evaluation?
- How much and what kinds of information do potential users need?
- Should there be equal focus on each of the criteria or will some information be more useful?
- Is this criterion a useful or appropriate measure for the particular evaluation?
- Which criterion will produce the most useful information given available resources?

51 For more detail see: Beck T, 'Evaluating Humanitarian Action Using OECD/DAC Criteria', 2006.

## EVALUATION QUESTIONS

Evaluation questions, when answered, can give users of the evaluation the information they seek in order to make decisions, take action or add to the knowledge base. The evaluation questions refine the focus of the evaluation by making explicit the aspects of the initiative that will be considered when judging its performance.

Evaluation questions reflect the underlying chain of assumptions about how the initiative is expected to operate within its contexts pursuant to the intended outputs and outcomes. The questions chosen for an evaluation should follow from a thorough understanding of the initiative's operations, intentions and context and should be selected for their role in meeting the evaluation purpose, objectives and relevant evaluation criteria.

An indefinite number of questions could be asked for each evaluation criterion. Real world evaluations are limited in terms of time, budget and resources. Therefore, it is important to be strategic in determining what information is needed most and to prioritize evaluation questions. It is better to answer fewer questions robustly than to answer more superficially. A clear and concise set of the most relevant questions ensures that evaluations are focused, manageable, cost efficient and useful.

To ensure that the key questions selected for the evaluation are the most relevant and most likely to yield meaningful information for users, UNDP programme units must solicit input from and negotiate agreement among partners and other stakeholders, including the evaluation team. Commissioning offices should ensure that the evaluation matrix in the evaluation inception report makes clear the linkages among the evaluation criteria, the evaluation questions and the information needs of intended users (see Annex 3 for more details).

## GENDER, EXCLUSION SENSITIVITY AND RIGHTS-BASED APPROACH

Consistent with UNDP development efforts, UNDP evaluations are guided by the principles of gender equality, the rights-based approach and human development.<sup>52</sup> Thus, as appropriate, UNDP evaluations assess the extent to which UNDP initiatives: have addressed the issues of social and gender inclusion, equality and empowerment; contributed to strengthening the application of these principles to various development efforts in a given country; and incorporated the UNDP commitment to rights-based approaches and gender mainstreaming in the initiative design.

Mainstreaming a gender perspective is the process of assessing the implications for women and men of any planned action, including legislation, policies or programmes,

---

52 UNDP, 'The Evaluation Policy of UNDP', Executive Board Document DP/2005/28, May 2006. Available at: <http://www.undp.org/eo/documents/Evaluation-Policy.pdf>.

in all areas and at all levels. It is a strategy for making gender-related concerns and experiences an integral dimension of the design, implementation, monitoring and evaluation of policies and programmes in all political, economic and societal spheres so that women and men benefit equally and inequality is not perpetuated. UNDP evaluations should assess the extent to which UNDP initiatives have considered mainstreaming a gender perspective in the design, implementation and outcome of the initiative and if both women and men can equally access the initiative's benefits to the degree they were intended. Similarly, evaluations should also address the extent to which UNDP has advocated for the principle of equality and inclusive development, and has contributed to empowering and addressing the needs of the disadvantaged and vulnerable populations in a given society.

The rights-based approach in development efforts entails the need to ensure that development strategies facilitate the claims of rights-holders and the corresponding obligations of duty-bearers. This approach also emphasizes the important need to address the immediate, underlying and structural causes for not realizing such rights. The concept of civic engagement, as a mechanism to claim rights, is an important aspect in the overall framework. When appropriate, evaluations should assess the extent to which the initiative has facilitated the capacity of rights-holders to claim their rights and duty-bearers to fulfill their obligations.

Evaluations should also address other cross-cutting issues, depending on the focus of the evaluation, such as the extent to which UNDP has incorporated and fostered South-South cooperation, knowledge management, volunteerism and UN coordination in its initiative.

## 7.5 EVALUATION METHODOLOGY

The evaluation design must detail a step-by-step plan of work that specifies the methods the evaluation will use to collect the information needed to address the evaluation criteria and answer the evaluation questions, analyse the data, interpret the findings and report the results.

Evaluation methods should be selected for their rigour in producing empirically based evidence to address the evaluation criteria and respond to the evaluation questions. The **evaluation inception report** should contain an **evaluation matrix** that displays for each of the evaluation criteria, the questions and subquestions that the evaluation will answer, and for each question, the data that will be collected to inform that question and the methods that will be used to collect that data (see Box 40). In addition, the inception report should make explicit the underlying theory or assumptions about how each data element will contribute to understanding the development results—*attribution, contribution, process, implementation and so forth*—and the rationale for data collection, analysis and reporting methodologies selected.

## Box 40. Questions for evaluators

The commissioning office should, at a minimum, ensure that the evaluation methods detailed in the evaluators' inception report respond to each of the following questions:

- What evidence is needed to address the evaluation questions?
- What data collection method(s) will be used to address the evaluation criteria and questions? Why were these methods selected? Are allocated resources sufficient?
- Who will collect the data?
- What is the framework for sampling? What is the rationale for the framework?
- How will programme participants and other stakeholders be involved?
- What data management systems will be used? That is, what are the planned logistics, including the procedures, timing, and physical infrastructure that will be used for gathering and handling data?
- How will the information collected be analysed and the findings interpreted and reported?
- What methodological issues need to be considered to ensure quality?

## DATA COLLECTION METHODS

The data to be collected and the methods for collecting the data will be determined by: the evidence needed to address the evaluation questions; the analyses that will be used to translate the data into meaningful findings in response to the evaluation questions; and judgements about what data are feasible to collect given constraints of time and resources. UNDP evaluations draw heavily on data (performance indicators) generated through monitoring during the programme or project implementation cycle. Performance indicators are a simple and reliable means to document changes in development conditions (outcomes), production, or delivery of products and services (outputs) connected to a development initiative (see Chapter 2).

Performance indicators are useful but have limitations. Indicators only indicate; they do not explain. Indicators will not likely address the full range of questions the evaluation seeks to address. For example, indicators provide a measure of what progress has been made. They do not explain why that progress was made or what factors contributed to the progress. UNDP evaluations generally make use of a mix of other data sources, collected through multiple methods, to give meaning to what performance indicators tell us about the initiative.

**Primary data** consists of information evaluators observe or collect directly from stakeholders about their first-hand experience with the initiative. These data generally consist of the reported or observed values, beliefs, attitudes, opinions, behaviours, motivations and knowledge of stakeholders, generally obtained through questionnaires, surveys, interviews, focus groups, key informants, expert panels, direct observation and case studies. These methods allow for more in-depth exploration and yield information that can facilitate deeper understanding of observed changes in outcomes and outputs (both intended and unintended) and the factors that contributed by filling out the operational context for outputs and outcomes.

**Secondary data** is primary data that was collected, compiled and published by someone else. Secondary data can take many forms but usually consists of documentary

**Table 28. Summary of common data collection methods used in UNDP evaluations<sup>53</sup>**

Method	Description	Advantages	Challenges
<b>Monitoring and Evaluation Systems</b>	Uses performance indicators to measure progress, particularly actual results against expected results.	Can be a reliable, cost-efficient, objective method to assess progress of outputs and outcomes.	Dependent upon viable monitoring and evaluation systems that have established baseline indicators and targets and have collected reliable data in relation to targets over time, as well as data relating to outcome indicators.
<b>Extant Reports and Documents</b>	Existing documentation, including quantitative and descriptive information about the initiative, its outputs and outcomes, such as documentation from capacity development activities, donor reports, and other evidence.	Cost efficient.	Documentary evidence can be difficult to code and analyse in response to questions.  Difficult to verify reliability and validity of data.
<b>Questionnaires</b>	Provides a standardized approach to obtaining information on a wide range of topics from a large number or diversity of stakeholders (usually employing sampling techniques) to obtain information on their attitudes, beliefs, opinions, perceptions, level of satisfaction, etc. concerning the operations, inputs, outputs and contextual factors of a UNDP initiative.	Good for gathering descriptive data on a wide range of topics quickly at relatively low cost.  Easy to analyse.  Gives anonymity to respondents.	Self-reporting may lead to biased reporting.  Data may provide a general picture but may lack depth.  May not provide adequate information on context.  Subject to sampling bias.
<b>Interviews</b>	Solicit person-to-person responses to predetermined questions designed to obtain in-depth information about a person's impressions or experiences, or to learn more about their answers to questionnaires or surveys.	Facilitates fuller coverage, range and depth of information of a topic.	Can be time consuming.  Can be difficult to analyse.  Can be costly.  Potential for interviewer to bias client's responses.
<b>On-Site Observation</b>	Entails use of a detailed observation form to record accurate information on-site about how a programme operates (ongoing activities, processes, discussions, social interactions and observable results as directly observed during the course of an initiative).	Can see operations of a programme as they are occurring.  Can adapt to events as they occur.	Can be difficult to categorize or interpret observed behaviours.  Can be expensive.  Subject to (site) selection bias.
<b>Group Interviews</b>	A small group (6 to 8 people) are interviewed together to explore in-depth stakeholder opinions, similar or divergent points of view, or judgements about a development initiative or policy, as well as information about their behaviours, understanding and perceptions of an initiative or to collect information around tangible and non-tangible changes resulting from an initiative.	Quick, reliable way to obtain common impressions from diverse stakeholders.  Efficient way to obtain a high degree of range and depth of information in a short time.	Can be hard to analyse responses.  Requires trained facilitator.  May be difficult to schedule.

<sup>53</sup> Methods described are illustrative and not exhaustive of the types of methods that have applicability for UNDP evaluation context.

**Table 28 (cont-d). Summary of common data collection methods used in UNDP evaluations**

Method	Description	Advantages	Challenges
<b>Key Informants</b>	Qualitative in-depth interviews, often one-on-one, with a wide-range of stakeholders who have first-hand knowledge about the initiative operations and context. These community experts can provide particular knowledge and understanding of problems and recommend solutions.	Can provide insight on the nature of problems and give recommendations for solutions.  Can provide different perspectives on a single issue or on several issues.	Subject to sampling bias.  Must have some means to verify or corroborate information.
<b>Expert Panels</b>	A peer review, or reference group, composed of external experts to provide input on technical or other substance topics covered by the evaluation.	Adds credibility.  Can serve as added (expert) source of information that can provide greater depth.  Can verify or substantiate information and results in topic area.	Cost of consultancy and related expenses if any.  Must ensure impartiality and that there are no conflicts of interest.
<b>Case Studies</b>	Involves comprehensive examination through cross comparison of cases to obtain in-depth information with the goal to fully understand the operational dynamics, activities, outputs, outcomes and interactions of a development project or programme.	Useful to fully explore factors that contribute to outputs and outcomes.	Requires considerable time and resources not usually available for commissioned evaluations.  Can be difficult to analyse.

evidence that has direct relevance for the purposes of the evaluation. Sources of documentary evidence include: local, regional or national demographic data; nationally and internationally published reports; social, health and economic indicators; project or programme plans; monitoring reports; previous reviews, evaluations and other records; country strategic plans; and research reports that may have relevance for the evaluation. Documentary evidence is particularly useful when the project or programme lacks baseline indicators and targets for assessing progress toward outputs and outcome measures. Although not a preferred method, secondary data can be used to help recreate baseline data and targets. Secondary information complements and supplements data collected by primary methods but does not replace collecting data from primary sources.

Given the nature and context of UNDP evaluations at the decentralized level, including limitations of time and resources, evaluators are often likely to use a mix of methods, including performance indicators, supplemented relevant documentary evidence from secondary sources, and qualitative data collected by a variety of means.

Table 28 presents brief descriptions of data collection methods that are most commonly applied in evaluations in UNDP for both project and outcome evaluations.

Commissioning offices need to ensure that the methods and the instruments (questions, surveys, protocols, checklists) used to collect or record data are: consistent

with quality standards of validity and reliability,<sup>54</sup> culturally sensitive and appropriate for the populations concerned, and valid and appropriate for the types of information sought and the evaluation questions being answered. In conflict-affected settings, factors such as security concerns, lack of infrastructure, limited access to people with information and sensitivities, and ethical considerations in terms of working with vulnerable people should be considered in determining appropriate data collection methods.

## ISSUES OF DATA QUALITY

UNDP commissioning offices must ensure that the evaluation collects data that relates to evaluation purposes and employs data collection methodologies and procedures that are methodologically rigorous and defensible and produces empirically verified evidence that is valid, reliable and credible.

Reliability and validity are important aspects of quality in an evaluation. **Reliability** refers to consistency of measurement—for example, ensuring that a particular data collection instrument, such as a questionnaire, will elicit the same or similar response if administered under similar conditions. **Validity** refers to accuracy in measurement—for example, ensuring that a particular data collection instrument actually measures what it was intended to measure. It also refers to the extent to which inferences or conclusions drawn from data are reasonable and justifiable. **Credibility** concerns the extent to which the evaluation evidence and the results are perceived to be valid, reliable and impartial by the stakeholders, particularly the users of evaluation results. There are three broad strategies to improve reliability and validity that a good evaluation should address:

- Improve the quality of sampling
- Improve the quality of data gathering
- Use mixed methods of collecting data and building in strategies (for example, **triangulating** or multiple sources of data) to verify or cross-check data using several pieces of evidence rather than relying only on one

### Improve sampling quality

UNDP evaluations often gather evidence from a sample of people or locations. If this sample is unrepresentative of a portion of the population, then wrong conclusions can be drawn about the population. For example, if a group interview only includes those from the city who can readily access the venue, the concerns and experiences of those in outlying areas may not be adequately documented. The sample must be selected on the basis of a rationale or purpose that is directly related to the evaluation purposes and is intended to ensure accuracy in the interpretation of findings and usefulness of evaluation results. Commissioning offices should ensure that the evaluation design makes clear the characteristics of the sample, how it will be selected, the rationale for

---

54 See discussion of validity and reliability in the Issues of Data Quality section of this chapter.

the selection, and the limitations of the sample for interpreting evaluation results. If a sample is not used, the rationale for not sampling and the implications for the evaluation should be discussed.

### **Ensure consistency of data gathering**

Whether using questionnaires, interview schedules, observation protocols or other data gathering tools, the evaluation team should test the data collection tools and make sure they gather evidence that is both accurate and consistent. Some ways of addressing this would be:

- Train data collectors in using observation protocols to ensure they record observations in the same way as each other
- Check the meaning of key words used in questionnaires and interview schedules, especially if they have been translated, to make sure respondents understand exactly what is being asked
- Consider how the characteristics of interviewers (especially age, gender and whether they are known to the informants) might improve or reduce the accuracy of the information provided

### **'Triangulate' data to verify accuracy: Use multiple data sources**

Good evaluation evidence is both consistent and accurate. Building in strategies to verify data will enhance the reliability and ensure valid results.

- Use a mix of methods to collect data rather than relying on one source or one piece of evidence. For example, triangulate the evidence from one source (such as the group interview) with other evidence about the experiences of those in rural areas. (This might be documentary evidence from reports or key informant interviews with people who are credible and well-informed about the situation.)
- Use experts to review and validate evidence.

The challenge for UNDP evaluations is to employ rigorous evaluation design methods that will produce useful information based on credible evidence that is defensible in the face of challenges to the accuracy of the evidence and the validity of the inferences made about the evidence.

## **ETHICAL CONSIDERATIONS**

Evaluations should be designed and conducted to respect and protect the rights and welfare of people and the communities of which they are members, in accordance with the UN Universal Declaration of Human Rights<sup>55</sup> and other human rights conventions. Evaluators should respect the dignity and diversity of evaluation participants

---

55 United Nations, 'Universal Declaration of Human Rights'. Available at: <http://www.un.org/en/documents/udhr/>.

when planning, carrying out and reporting on evaluations, in part by using evaluation instruments appropriate to the cultural setting. Further, prospective evaluation participants should be treated as autonomous, be given the time and information to decide whether or not they wish to participate, and be able to make an independent decision without any pressure. Evaluation managers and evaluators should be aware of implications for doing evaluations in conflict zones. In particular, evaluators should know that the way they act, including explicit and implicit messages they transmit, may affect the situation and expose those with whom the evaluators interact to greater risks.<sup>56</sup> When evaluators need to interview vulnerable groups, evaluators should make sure interviewees are aware of the potential implications of their participation in the evaluation exercise and that they are given sufficient information to make a decision about their participation. All evaluators commissioned by UNDP programme units should agree and sign the Code of Conduct for Evaluators in the UN System.<sup>57</sup> For more information on ethics in evaluation, please refer to the 'UNEG Ethical Guidelines for Evaluation'.<sup>58</sup>

#### **Box 41. Human rights and gender equality perspective in evaluation design**

Evaluations in UNDP are guided by the principles of human rights and gender equality. This has implications for evaluation design and conduct, and requires shared understanding of these principles and explicit attention on the part of evaluators, evaluation managers and evaluation stakeholders. For example, in collecting data, evaluators need to ensure that women and disadvantaged groups are adequately represented. In order to make excluded or disadvantaged groups visible, data should be disaggregated by gender, age, disability, ethnicity, wealth and other relevant differences where possible.

Further, data should be analysed whenever possible through multiple lenses, including sex, socio-economic grouping, ethnicity and disability. Marginalized groups are often subject to multiple forms of discrimination, and it is important to understand how these different forms of discrimination intersect to deny rights holders their rights.

## **ANALYSIS AND SYNTHESIS OF DATA**

Data collection involves administering questionnaires, conducting interviews, observing programme operations, and reviewing or entering data from existing data sources. Data analysis is a systematic process that involves organizing and classifying the information collected, tabulating it, summarizing it, and comparing the results with other appropriate information to extract useful information that responds to the evaluation questions and fulfils the purposes of the evaluation. It is the process of deciphering facts from a body of evidence by systematically coding and collating the data collected, ensuring its accuracy, conducting any statistical analyses, and translating the data into usable formats or units of analysis related to each evaluation question.

---

56 OECD Guidance on Conflict Prevention and Peacebuilding Activities, Working Draft for Application Period, 2008.

57 UNEG, 'Code of Conduct', June 2008. Available at: [http://www.uneval.org/papersandpubs/documentdetail.jsp?doc\\_id=100](http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=100).

58 UNEG, 'Ethical Guidelines for Evaluation', June 2008. Available at <http://www.uneval.org/search/index.jsp?q=ethical+guidelines>.

Data analysis seeks to detect patterns in evidence, either by isolating important findings (analysis) or by combining sources of information to reach a larger understanding (synthesis). Mixed method evaluations require the separate analysis of each element of evidence and a synthesis of all sources in order to examine patterns of agreement, convergence or complexity.

### **Analysis plan**

Data analysis and synthesis must proceed from an analysis plan that should be built into the evaluation design and work plan detailed in the inception report. The analysis plan is an essential evaluation tool that maps how the information collected will be organized, classified, inter-related, compared and displayed relative to the evaluation questions, including what will be done to integrate multiple sources, especially those that provide data in narrative form, and any statistical methods that will be used to integrate or present the data (calculations, sums, percentages, cost analysis and so forth). Possible challenges and limitations of the data analysis should be described. The analysis plan should be written in conjunction with selecting data collection methods and instruments rather than afterward.

### **Interpreting the findings**

Interpreting findings is the process of giving meaning to the evaluation findings derived from the analysis. It extracts from the summation and synthesis of information derived from facts, statements, opinions, and documents and turns findings from the data into judgements about development results (conclusions). On the basis of those conclusions, recommendations for future actions will be made. Interpretation is the effort of figuring out what the findings mean—making sense of the evidence gathered in an evaluation and its practical applications towards development effectiveness.

### **Drawing conclusions**

A conclusion is a reasoned judgement based on a synthesis of empirical findings or factual statements corresponding to specific circumstances. Conclusions are not findings; they are interpretations that give meaning to the findings. Conclusions are considered valid and credible when they are directly linked to the evidence and can be justified on the basis of appropriate methods of analysis and synthesis to summarize findings. Conclusions should:

- Consider alternative ways to compare results (for example, compared with programme objectives, a comparison group, national norms, past performance or needs)
- Generate alternative explanations for findings and indicate why these explanations should be discounted
- Form the basis for recommending actions or decisions that are consistent with the conclusions
- Be limited to situations, time periods, persons, contexts and purposes for which the findings are applicable<sup>59</sup>

---

59 Based on Joint Committee on Standards for Educational Evaluation, 'Programme Evaluation Standards: How to Assess Evaluations of Educational Programmes', 1994, 2nd ed, Sage Publications, Thousand Oaks, CA.

## **Making recommendations**

Recommendations are evidence-based proposals for action aimed at evaluation users. Recommendations should be based on conclusions. However, forming recommendations is a distinct element of evaluation that requires information beyond what is necessary to form conclusions. Developing recommendations involves weighing effective alternatives, policy, funding priorities and so forth within a broader context. It requires in-depth contextual knowledge, particularly about the organizational context within which policy and programmatic decisions will be made and the political, social and economic context in which the initiative will operate.

Recommendations should be formulated in a way that will facilitate the development of a management response (see Chapter 6 and Annex 6 on Management Response System). Recommendations must be realistic and reflect an understanding of the commissioning organization and potential constraints to follow up. Each recommendation should clearly identify its target group and stipulate the recommended action and rationale.

## **Lessons learned**

The lessons learned from an evaluation comprise the new knowledge gained from the particular circumstance (initiative, context outcomes and even evaluation methods) that is applicable to and useful in other similar contexts. Frequently, lessons highlight strengths or weaknesses in preparation, design and implementation that affect performance, outcome and impact.