



Independent
Evaluation
Office

United Nations Development Programme

IMPACT EVALUATION GUIDANCE: GUIDE FOR PROGRAMME MANAGERS

February 2026

Contents

- 1. Introduction _____ 1
- 2. Why impact evaluation for UNDP? _____ 2
- 3. Defining impact and impact evaluation _____ 2
 - 3.1 Levels of impact evaluation _____ 3
 - 3.2 Causal attribution and counterfactual _____ 4
- 4. Why it is important to address UNDP programme specificities _____ 5
- 5. When not to do an impact evaluation: determining if impact evaluation is the right option _____ 6
- 6. Implementing impact evaluation _____ 8
- Annex I. Key concepts _____ 11
- Annex II. Checklist for impact assessment suitability _____ 14
- Annex III. Checklist for evaluating the need for impact data collection _____ 15

1. Introduction

The United Nations Development Programme (UNDP) recognizes the critical importance of measuring and assessing impact to understand the contributions of its interventions on development outcomes. To support the planning and execution of impact evaluations within UNDP, UNDP's Independent Evaluation Office (IEO) has developed a set of guidance documents that outline how UNDP could best implement decentralized impact evaluation and utilize impact evidence. This guidance document for programme managers addresses the growing need for rigorous evaluations within UNDP, in alignment with its Strategic Plan and knowledge management strategy. Impact evaluation, informed by consultations and best practices, enables programme managers to critically assess initiatives and fosters continuous learning by identifying successes and failures. Insights from impact evaluations can guide decisions on programme improvement, scaling or discontinuation, and demonstrate measurable impacts to stakeholders and donors.

2. Why impact evaluation for UNDP?

Impact evaluation has evolved in international development, shifting from a focus on accountability and transparency to addressing broader knowledge gaps. This shift emphasizes learning and effectiveness, supporting development organizations to understand what strategies are successful and why, leading to improved programme design and implementation (Figure 1). Although UNDP has historically conducted outcome evaluations, there is growing recognition of the importance of understanding the long-term effects of its development initiatives and the initiatives it would like to scale. Below are several reasons why impact evaluations are a valuable tool for UNDP:

- **Bridging knowledge gaps:** Impact evaluations are crucial for UNDP to transform into a learning organization. They address specific knowledge gaps that other monitoring and evaluation activities may overlook. By rigorously attributing changes to UNDP programmes, these evaluations provide unique insights that can inform programme design and implementation. This information is vital for deciding whether to scale, modify or discontinue a programme based on its effectiveness.
- **Identifying scalable programmes:** Impact evaluations help identify successful programmes that can be scaled or replicated in different contexts. For instance, if a community-based conflict prevention initiative shows significant improvements in peace outcomes, it may be adapted for similar regions facing comparable challenges. This adaptability is essential for UNDP, which operates in diverse environments, allowing for the effective application of successful models across various settings.
- **Comparative analysis of strategies:** Impact evaluations facilitate comparing the ability of different methods to achieve the same objectives, enabling UNDP and governments to identify the most effective strategies. For example, if an evaluation reveals that a specific approach to employment generation yields better outcomes for income generation than others, UNDP can refine its programmes based on this evidence. This comparative analysis is critical for addressing complex development issues effectively.
- **Demonstrating effectiveness to stakeholders:** Impact evaluations play a vital role in showcasing the measurable significance and effectiveness of UNDP programmes to stakeholders and donors. By systematically assessing the outcomes of its initiatives, UNDP can identify successful programmes and allocate resources to those that yield the greatest positive impact. This transparency fosters trust and confidence among stakeholders, because they can see concrete evidence of UNDP's commitment to effectiveness and accountability.

Although UNDP’s integrated results and resources framework (IRRF) architecture includes impact and outcome levels, it does not capture the causal attribution required by impact evaluations. The IRRF tracks programme-level outputs but does not attribute specific impacts to UNDP programmes and Strategic Plan outcomes. Measuring impact and impact evaluations go further, enabling UNDP to measure its contribution to the Sustainable Development Goals accurately.

- **Supporting organizational learning and improvement:** Ultimately, impact evaluations foster a cycle of learning and improvement within UNDP and outside of it, through the use and sharing of knowledge of what worked and why. By understanding the effectiveness of its programmes, UNDP can continuously refine its strategies and interventions. This commitment to learning not only enhances programme design but also ensures that UNDP remains responsive to the evolving needs of the communities it serves, thereby maximizing its overall impact in international development.

Impact evaluations also fulfil the knowledge management strategy’s objective to apply insights across development portfolios, boosting programme results and impact. They support the organizational effectiveness framework, which emphasizes ‘knowledge’ and ‘impact measurement’ as crucial components, reinforcing impact evaluations as central to UNDP’s global mission.

Figure 1. Why implement impact evaluations?



3. Defining impact and impact evaluation

Impact evaluation at UNDP is a systematic and empirical approach aimed at assessing the causal effects of UNDP interventions on development outcomes. It seeks to answer the critical question of “So what?”, by determining whether specific interventions lead to significant and transformative changes in development. Unlike other forms of evaluation that may focus on immediate outputs or outcomes, impact evaluation emphasizes both short-term and long-term effects, providing a comprehensive understanding of how UNDP programmes contribute to sustainable development. The definition of ‘long-term effects’ can vary depending on the programme intent, reflecting the diversity and complexity of UNDP programmes.

Impact evaluations are characterized by their rigorous methodology, which includes the use of counterfactuals to establish a comparison between groups that received the intervention (treatment groups) and those that did not (control or comparison groups). This approach enables evaluators to isolate the effects of the intervention and ascertain whether observed changes can be attributed to UNDP’s actions. The evaluations can be conducted at various levels—global, national, institutional, community, household or individual—depending on the specific context and beneficiaries of the programme.

Impact evaluations can be integrated as part of the programme design (ex ante) or can be retrospective (ex post). Integrated impact evaluations are planned alongside programme design, collecting baseline data before implementation for both the treatment and the non-treatment groups. In contrast, retrospective impact evaluations can be conducted after programme completion—ex post—even though not originally planned and without prior data collection; however, it can be more challenging to establish causal relationships in retrospective impact evaluations.

3.1 Levels of impact evaluation

Impact evaluations can be conducted over various time-frames:

- **Ultimate effects (long-term effects):** These evaluations assess the broader, long-term impacts of a programme, such as whether an employment initiative has led to a reduction in poverty and improved livelihoods over time.
- **Intermediate effects (medium-term effects):** Given the challenges of measuring long-term impacts, impact evaluations may focus on intermediate outcomes that serve as proxies for longer-term effects. For instance, assessing whether a programme has increased household income can provide evidence of its potential long-term benefits.
- **Short-term effects:** Impact evaluations may also focus on immediate changes resulting from interventions, particularly those that are expected to catalyse broader changes, such as training programmes for community leaders.

Impact evaluation is relevant and applicable across various thematic areas, regardless of a programme's size, its level of beneficiaries, its global or local scope, or its duration. Impact evaluation can be applied to any programme and, as mentioned above, may include assessment at various levels—for example, global, national, institutional, community, household or individual.

UNDP impact evaluations are categorized into two types: evaluations conducted by the IEO and decentralized evaluations. Independent impact evaluations are conducted and managed by the IEO. Decentralized impact evaluations, on the other hand, are commissioned by country offices, UNDP programme and policy units, or regional and policy bureaux.

Distinction between outcome evaluation and impact evaluation

UNDP outcome evaluations look at different levels of programme results and measure immediate results or the programme's contribution to broader change processes, or they delve deeper into the sustained effects of a programme on the community. Outcome evaluations typically do not require rigorous comparisons to control groups, because they focus on direct results.

In contrast, impact evaluations necessitate a more complex analysis to isolate the programme's effects on long-term development outcomes. The use of counterfactuals is a defining feature of impact evaluations, because they help establish causal relationships and determine what would have happened in the absence of intervention.

3.2 Causal attribution and counterfactual

Preventing erroneous impact evaluation findings: getting the counterfactual right

- The group receiving programme interventions must be comparable to the group not receiving them.
- Inherent differences between the groups can skew observed outcomes, making it unclear if changes are due to the programme or other factors.
- Considerations include different starting points and trends and the possibility that participants may have other influencing factors.
- Not accounting for the effect of such differences can lead to underestimating or overestimating the programme's impact.

To accurately assess the programme's impact, it is crucial to consider a counterfactual scenario: what would have occurred without the programme? A 'counterfactual' represents a hypothetical situation that illustrates what would have happened in the absence of the intervention. In evaluation, it allows for a comparison between the actual outcomes (with the programme) and the hypothetical outcomes (without the programme). For instance, a counterfactual analysis might pose the question, "What would have happened to the target population if the programme had not been implemented?" This approach helps isolate the programme's effects from other external factors that may influence outcomes.

'Causal attribution' is the process of linking a particular outcome directly to an intervention or action. It focuses on establishing a clear cause-and-effect relationship, demonstrating that changes observed in a target group can be confidently traced back to the intervention rather than to external factors. Causal attribution involves isolating the programme's effects, establishing a causal link, using rigorous methods and eliminating alternative explanations:

- **Isolating the programme's effect:** Evaluators must rule out other influences that could affect outcomes.
- **Establishing a clear causal link:** The relationship between the intervention and the outcome must be well defined.
- **Using rigorous methods:** Employing methodologies such as randomized controlled trials (RCTs) to control external variables.
- **Eliminating alternative explanations:** Demonstrating that observed outcomes are not due to unrelated factors.

4. Why it is important to address UNDP programme specificities

As more UNDP programmes and offices begin to adopt impact evaluations, it is important to factor in unique programme specificities that have hindered their broader integration at UNDP. Although these challenges can be addressed post facto, in the case of new programmes impact measurement considerations should be incorporated into the programme's planning and by adopting a systematic approach to impact data collection.

- **Distinct UNDP programme characteristics:** UNDP's vast operational scale, with programmes in 170 countries and across diverse thematic areas (such as inclusive growth, governance, environment, crisis and resilience), makes it challenging to standardize impact evaluation. Each programme targets various levels (e.g. community, household, institutional) with unique outcomes, requiring tailored approaches for effective impact measurement. For example, while some programmes may focus on measurable outcomes such as poverty reduction, others, particularly in governance, may have more complex objectives that are harder to quantify. The beneficiaries of UNDP programmes also can vary widely, even within the same project. Some initiatives may target entire communities, whereas others focus on specific institutions. This variability complicates the evaluation process, because different groups may experience different impacts from the same intervention. This complexity necessitates a flexible approach to evaluation that can accommodate the specific characteristics of each programme as well as programme-specific evaluation methods to capture accurate, scalable data.
- **Impact data collection:** Impact evaluations rely heavily on comprehensive data collection. UNDP currently lacks a standardized procedure for gathering detailed data on beneficiaries before, during and after programme implementation. This gap makes it difficult to establish control or comparison groups, which are essential for assessing the true impact of interventions. Additionally, the absence of a robust data-collection system can hinder the ability to conduct longitudinal studies, which are necessary for understanding long-term effects. This limitation complicates the creation of counterfactuals and hinders accurate impact analysis. Without reliable data, the credibility and utility of impact evaluations are compromised.
- **Operational limitations:** UNDP should design impact evaluations for programmes that are more conducive to rigorous assessment. Many UNDP initiatives are small scale and short term, making it challenging to evaluate their long-term impacts. Furthermore, when UNDP provides resources but does not manage the implementation, it becomes difficult to attribute outcomes directly to UNDP's involvement. This lack of direct control complicates the assessment of UNDP's contributions to broader socio-economic changes, because the organization may not be able to influence how programmes are executed on the ground.
- **Incorporation into the IRRF:** Impact evaluations are central to the pillars of 'knowledge' and 'impact measurement' within the IRRF. Although the IRRF captures outputs and outcomes, it does not fully address the attribution necessary for impact measurement and evaluations. Therefore, it is essential to revisit the IRRF to incorporate impact measurement parameters to provide deeper insights beyond mere reporting of progress towards Sustainable Development Goals.

5. When not to do an impact evaluation: determining if impact evaluation is the right option

When not to do an impact evaluation

- Avoid conducting impact evaluations solely for accountability purposes. There are more cost-effective methods to establish accountability.
- Refrain from performing an impact evaluation if there are already sufficient studies available to inform programming decisions.
- Do not pursue an impact evaluation if counterfactuals cannot be established.

Although mixed methods can help address the measurement challenges associated with RCTs, which are commonly used in impact evaluations, the real issue lies not just in the methods themselves but in the feasibility of conducting an impact evaluation. Programme managers should focus more on recognizing when impact evaluations are not practical. Some situations in which they are not practical are listed below (see also Figure 2):¹

- **Impact evaluation isn't always the right tool for every question:** Some questions about programme effectiveness can be answered through programme monitoring data and outcome evaluations rather than impact evaluations. Outcome data can provide valuable insights into how the programme is being delivered, including participant engagement, adherence to the programme model, and any challenges faced during implementation. These data can also help programme managers identify areas for improvement.
- **Not the right timing:** Conducting an impact evaluation during the early development phase of a programme may be premature, because key elements such as objectives, target population and delivery methods are still being refined. Instead, the focus should be on collecting monitoring data to ensure proper implementation and to inform future impact evaluations.
- **Context feasibility:** UNDP programmes operate in diverse programme contexts. Special programmes operating in rapidly changing or unpredictable contexts, such as crisis responses, face unique challenges when it comes to conducting impact evaluations. In such cases, it would be useful to use outcome evaluations instead of the time-consuming processes involved in rigorous evaluation.
- **Feasibility of measuring indirect effects:** Programmes aim to create indirect effects, such as community-wide benefits stemming from individual interventions. For example, livelihood programmes can improve household nutrition and community health. However, measuring these indirect effects can be complex and may require additional resources, such as longitudinal data

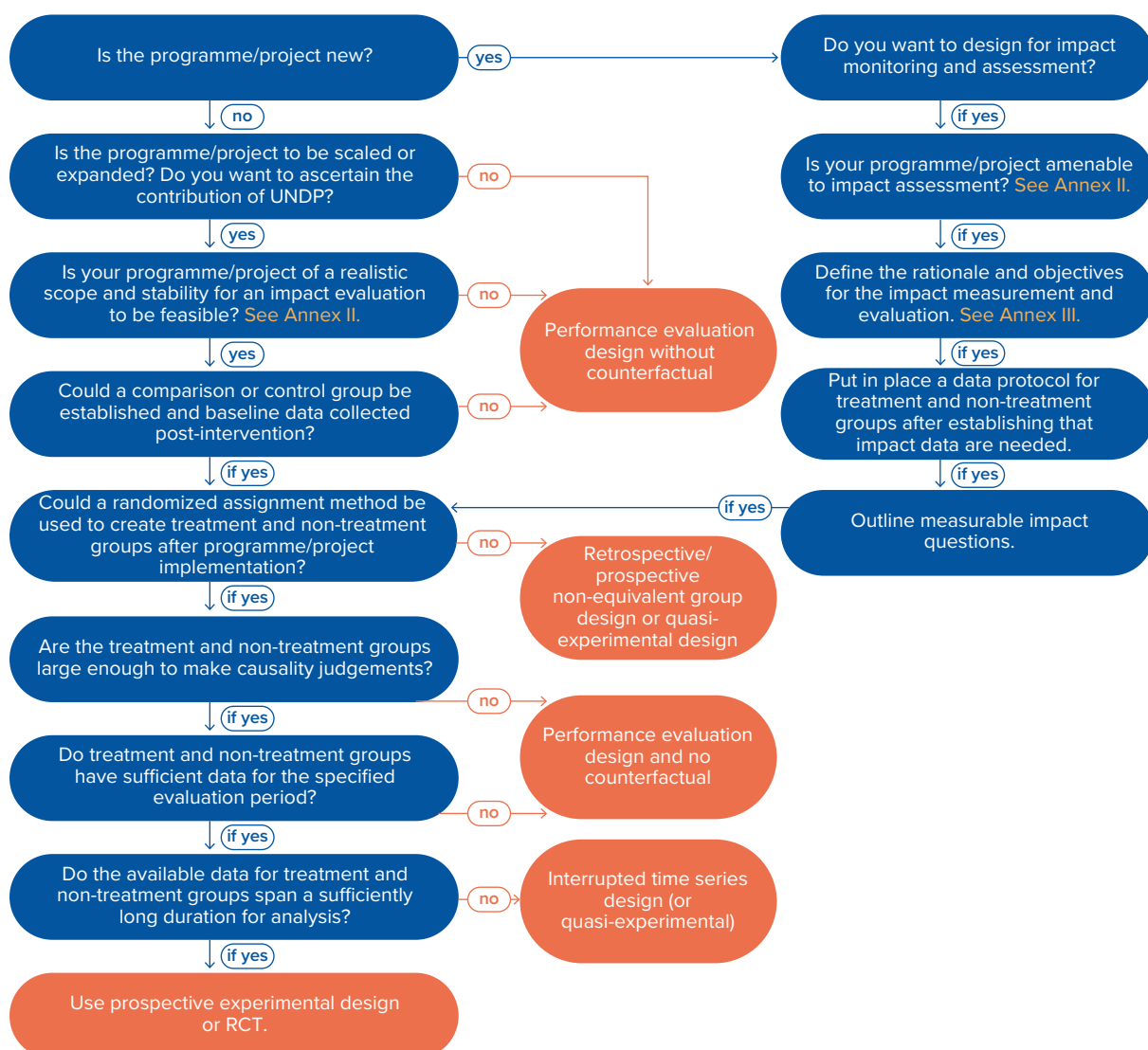
¹ For an excellent discussion on this topic, see Mary Kay Gugerty, and Dean Karlan, “Ten Reasons Not to Measure Impact – and What to Do Instead”, *Stanford Social Innovation Review*, vol. 16, No. 3 (Summer 2018), pp. 41–47. Available at https://prevention-collaborative.org/wp-content/uploads/2021/08/Gugerty_2018_Ten_Reasons_Not_to_Measure_Impact_and_What_to_Do_Instead_1.pdf.

collection and sophisticated analytical methods. If these indirect effects are significant and cannot be accurately assessed, it may be more prudent to avoid a formal impact evaluation altogether. Instead, organizations can focus on understanding the direct effects of their interventions and on gathering qualitative data to explore potential indirect benefits.

- **Not worth the effort:** If existing evidence already answers the question of a programme's effectiveness, or if the evaluation will not yield generalizable knowledge, it may not be worth the investment in an impact evaluation.
- **Distraction from improvement:** The push for more impact measurement can sometimes lead to poor studies and wasted money, distracting from collecting data that can actually help improve the performance of an effort.

When substantial evidence from previous studies already indicates the effectiveness of a programme or intervention, conducting another impact evaluation may not yield new insights. Before investing in a new evaluation, careful consideration should be given to whether the existing literature sufficiently addresses the research questions. If similar programmes have been rigorously evaluated in comparable contexts, the findings from those studies may provide valuable guidance for current efforts.

Figure 2. Flow chart for deciding whether to conduct an impact evaluation



6. Implementing impact evaluation

Conducting a robust impact evaluation involves a systematic sequence of steps designed to ensure the reliability of findings and their relevance to future policy and programme decisions. Although these overarching steps are consistent across both global and country-specific evaluations, global programmes often present greater complexity and require meticulous attention in specific areas. Key areas of focus include defining evaluation objectives, formulating evaluation questions, identifying counterfactuals and determining data requirements (see Figure 3). Below are detailed considerations for each step:

- 1. Define the rationale and objectives for the impact evaluation:** This step involves clarifying the reasons for conducting the impact evaluation, as outlined in the checklist in Annex II. Sometimes, the programme may be broad, encompassing multiple initiatives, but the evaluation might focus on a specific component to assess scalability potential. In other cases, a global programme may require a broader review or cross-country comparisons. It is crucial to specify the evaluation's exact objective. What knowledge gap does it aim to address? Is the goal to assess the initial programme phase for scaling up? Is the focus on evaluating effectiveness within a single country or across multiple countries? Or is it to compare the programme's success in achieving its goals in different contexts?
- 2. Specify the programme, project or intervention that will be evaluated for impact:** Clearly articulate the scope and objectives of the programme, project or intervention undergoing impact evaluation. This includes specifying the target population, anticipated outcomes and resources allocated. Encourage country offices to conduct impact evaluations when feasible, because they provide valuable data on what works, for whom and under what conditions. Such evaluations also facilitate better alignment with national priorities, enhancing institutional capacities and fostering local ownership of interventions.
- 3. Build a theory of change:** The theory of change represents the connections between programme activities and their short-term, medium-term and long-term outcomes. It helps programme managers grasp how the programme is intended to function, while outlining the assumptions underlying the theory of change. For evaluators, it provides a framework for determining what needs measurement and testing, ensuring a thorough understanding of how activities are expected to yield desired impacts.
- 4. Identify the timeline, resources and budget for impact evaluations:** Once objectives are defined, the next step involves detailed planning for executing the impact evaluation. Identify internal resources or consider hiring external support. Key considerations include financing staff time, consultancy fees, data-collection resources and any additional support needed during the evaluation. Programme units should clarify who will cover these costs: will they come from the programme budget or other sources?
- 5. Develop measurable impact evaluation questions:** Building on the objectives, programme units should formulate specific, measurable questions that the impact evaluation will address. These questions must align with the programme's theory of change. For global programmes, it is important to differentiate between overarching global evaluation questions and those tailored to country-specific contexts. While some questions capture broad trends and outcomes, others may delve into unique local nuances. Early identification of these questions is critical, because they influence the choice of counterfactuals, data and research methods.
- 6. Define the counterfactual and identify data requirements:** This step closely follows the guidance in Annex III regarding data collection for baseline and post-intervention (follow-up) measures for both the treatment and non-treatment groups. Establishing a proper counterfactual can be particularly challenging for global evaluations. Creating a robust counterfactual that accommodates diverse country contexts while providing valid insights into the global programme's

impact requires careful consideration. For instance, longitudinal studies using methods such as difference-in-differences may be more appropriate for global programmes than RCTs. Additionally, standardizing indicators across countries is vital for ensuring data comparability, consistency and reliability, which are essential for robust global impact evaluations.

7. Identify the appropriate methodology for conducting the evaluation: This step should proceed in parallel with step 6, because the availability of data and the counterfactual design will inform the research method and the chosen method will influence the data and counterfactual design. Methodologies may include experimental designs (such as RCTs), quasi-experimental designs (such as difference-in-differences or matching) or non-experimental methods (such as case studies), depending on context and feasibility. For further details, refer to 'Impact Evaluation Guidance: Approaches and Methods'.²

8. Dissemination and reporting: The final step involves planning how to communicate the impact evaluation findings to relevant stakeholders. Dissemination strategies may include presentations, policy briefs, academic publications and workshops, to ensure that dissemination products are tailored to the needs of different audiences. This tailoring maximizes the likelihood that evaluation findings will inform decision-making and drive programmatic improvements.

Besides ethical considerations, the decision to undertake an impact evaluation should be guided by whether the expected benefits—such as improved programme design, accountability or learning—justify the investment of time and resources required for the evaluation. While implementing an impact evaluation, it is important to consider the following points.

- **Is it worth the cost?**

- Allocate fewer evaluation resources to questions for which we already possess reliable evidence regarding outcomes and to designs that do not effectively address the relevant questions.
- Allocate more evaluation resources to questions that have substantial knowledge gaps and partnerships that incorporate a clear pathway to action.

- **Who will act on the evaluation findings?**

- Ascertain who may take different actions based on the findings. It may be the same implementer/ programme or a future one. Regardless, what is the plan to translate the evidence into action?

- **Are ethical considerations addressed?**

- **Ensure that claims about programme effectiveness are accurate and transparent:** Accurate claims should be supported by robust evidence, including quantitative and qualitative data, and should be presented in a way that is accessible and understandable to all relevant audiences.
- **Respect local contexts and autonomy:** Respecting local contexts and autonomy will foster ownership of the evaluation and acting on the findings. Addressing this consideration involves engaging with local stakeholders, ensuring their perspectives are captured through appropriate evaluation design and implementation.
- **Guide data-collection practices to prioritize privacy and confidentiality:** Data-collection practices must prioritize the privacy and confidentiality of participants, to protect their personal information and maintain their trust. Evaluation participants should be informed about how their data will be used and have the option to provide informed consent.

² See <https://erc.undp.org/methods-center/guidelines/impact-evaluation-guidelines/impact-evaluation-approaches-methods>.

Figure 3. Key steps in designing a robust impact evaluation



Annexes

Annex I. Key concepts

Term	Definition
Activity	Activities are the strategies and actions that will be employed to implement the project. They can include workshops, meetings, counselling, facilitations, product development, recruitment, training, intakes, case management, referral, tracking, etc. Activities are typically described rather than measured.
Baseline data	Data collected before an intervention for the purpose of being compared to data collected after the intervention.
Causal	A type of evaluation question that seeks to determine the impact of an intervention—that is, whether one or more factors causes or affects the outcome, rather than simply being associated with it.
Causal attribution	<p>This process involves linking specific outcomes directly to an intervention, establishing a clear cause-and-effect relationship. To achieve this, evaluators must do the following:</p> <ul style="list-style-type: none">• Isolate the programme’s effect by ruling out other influencing factors.• Establish a clear causal link between the intervention and the observed outcomes.• Use rigorous methodologies, such as RCTs or quasi-experimental designs, to control for external variables.• Eliminate alternative explanations for the observed changes through data collection and statistical analysis.
Causal inference	This refers to determining whether a specific UNDP intervention is responsible for a particular outcome. It addresses the question of whether the intervention caused the observed changes and to what extent. The challenge lies in ruling out alternative explanations, ensuring that the outcomes are indeed attributable to the intervention.
Comparison group	<p>In a quasi-experimental research design, this is the group of research participants that, for the sake of comparison, does not receive the treatment or intervention given to the intervention group. Comparison groups are used to estimate the counterfactual.</p> <p>A valid comparison group will have the same characteristics, on average, as the group of beneficiaries of the programme (treatment group), except for the fact that the units in the comparison group do not benefit from the programme. Comparison group subjects are typically not randomly assigned to their condition, as would be true of control group subjects in an experimental design study.</p> <p>See also ‘control group’.</p>

Term	Definition
Control group	<p>Participants in an experimental research study/evaluation who do not receive the treatment/intervention. By not receiving the intervention, the control group provides the basis for estimating the counterfactual under conditions of random assignment.</p> <p>A valid control group is created through randomization, ensuring that, on average, its characteristics are statistically equivalent to those of the treatment group at baseline. This random assignment minimizes selection bias and strengthens causal inference, making experimental designs the gold standard for impact evaluation. Control group participants are typically subject to the same data-collection and observation protocols as the treatment group, except that they do not benefit from the intervention.</p> <p>See also ‘comparison group’.</p>
Counterfactual	<p>A theoretical alternative scenario used to evaluate the impact of a programme by comparing actual outcomes with those that would have occurred without the intervention—for example, assessing what would have happened to a target population if a specific UNDP programme had not been implemented.</p>
Endline data	<p>Data collected at the conclusion of an intervention to measure final outcomes and assess change relative to baseline conditions.</p>
Evaluation	<p>An objective assessment of the results accomplished by a programme. The UNDP Evaluation Policy follows the United Nations Evaluation Group definition of evaluation as “an assessment, as systematic and impartial as possible, of an activity, project, programme, strategy, policy, topic, theme, sector, operational area or institutional performance”.⁹</p>
Experimental group	<p>Participants deliberately assigned to receive the treatment within a controlled experimental research design, used to test causal effects by comparison with a control group. Assignment to the experimental group is typically done through randomization, ensuring that any observed differences in outcomes can be attributed to the intervention rather than pre-existing characteristics.</p> <p>Participants in the experimental group follow the same data-collection and observation protocols as the control group, except that they receive the programme inputs or activities being evaluated. This design minimizes bias and strengthens internal validity, making experimental groups a cornerstone of RCTs.</p> <p>See also ‘treatment group’.</p>
Impact	<p>Positive and negative, primary and secondary, intended or unintended effects that are partly or exclusively attributable to a UNDP development intervention, project or programme. To causally attribute an impact to the programme or activities, two factors must be considered. First, compare the programme with what would happen in the absence of the programme, such as comparing an outcome within a treatment group to that within a control group. Second, if possible, isolate the impacts of the programme from other external influential factors, such as political environment, competing or supporting initiatives from other government agencies, socio-economic and demographic factors, and so on.</p>
Impact evaluation	<p>An empirical evaluation that seeks to quantify the causal effects of a programme or project on outcomes of interest, typically by comparing observed results with an estimate of what would have happened in the absence of the intervention. It is used to assess attribution—whether the intervention itself is responsible for the observed changes.</p>

Term	Definition
Impact measurement	The purpose of impact measurement is to assess the long-term effects and changes that can be attributed to a programme. It aims to determine the overall impact of the programme on the target population or community, including unintended consequences.
Input	The inputs to a programme or project are the resources needed to implement the activities set out. These resources can be financial, tangible materials or the amount of time that is required to carry out certain activities during the lifetime of the project. Examples include funding for certain UNDP programmes and policies, or additional full-time staff members required to carry out the programme.
Outcome	The overall result of the programme, such as a change or improvement in knowledge, behaviour, skills, decision-making or awareness, or changes to policies, practices, partnerships, alliances, coalitions or networks.
Output	An output can be considered a programme- or project-level achievement and represents a tangible, intermediate result that must occur to produce the desired result (the outcome). However, the output is not the final purpose of a programme. Outputs can be tangible products, services, capacities or deliverables that result from the activities. Examples include the number of training sessions conducted, the number of participants, the amount of training material delivered, the number of staff members trained, the amount of money spent, the number of service referrals and the number of assessments administered.
Performance monitoring	The primary purpose of performance monitoring is to track the ongoing implementation and operational aspects of a programme. It focuses on assessing whether the programme is being delivered as planned and whether it is achieving its immediate outputs and outcomes.
Treatment group	<p>A participant group that receives the programme inputs or participates in programme activities, regardless of whether the study design is experimental or quasi-experimental. However, in quasi-experimental designs, the treatment group is not randomly assigned; instead, it is defined by programme participation, and its outcomes are compared to those of a comparison group.</p> <p>The treatment group serves as the basis for estimating the programme's impact by contrasting its outcomes with those of a group that did not receive the intervention. While the term 'experimental group' emphasizes the controlled nature of assignment, 'treatment group' is broader and applies to any evaluation context where participants receive the intervention.</p> <p>See also 'experimental group'.</p>

Source: ^a United Nations Evaluation Group, *Norms and Standards for Evaluation* (New York, 2016).

Available at https://www.unevaluation.org/unevaluation_publications/unevaluation-norms-and-standards-evaluation-un-system.

Annex II. Checklist for impact assessment suitability

Step 1: Determine the rationale for impact evaluation

This step assesses where there is a potential value in conducting an impact evaluation. A checklist categorizes questions into strong and medium qualifiers. A valid rationale is established if at least one strong qualifier or two medium qualifiers are answered affirmatively. Key questions include the following:

- Is there a knowledge gap about the intervention's effectiveness? (strong qualifier)
- Will the evaluation influence project design or resource allocation? (strong qualifier)
- Is there a plan to scale up the intervention? (medium qualifier)
- Is there a strong demand for accountability regarding the programme? (medium qualifier)

Step 2: Assess suitability for impact evaluation

In this step, UNDP staff and evaluators determine if an impact evaluation is the appropriate method for the evaluation question. As in Step 1, questions are categorized as strong and medium qualifiers. A valid rationale requires at least one strong qualifier or two medium qualifiers to be answered positively. Important questions include the following:

- Can the impact of the intervention be defined and measured? (strong qualifier)
- Is there a clear theory of change with measurable indicators? (medium qualifier)
- Does the evaluation question focus on causality? (medium qualifier)
- Is there interest in understanding cost-effectiveness or subgroup impacts? (strong qualifier)

Step 3: Determine feasibility of impact evaluation

Once the rationale and suitability are established, this step evaluates the feasibility of conducting a rigorous impact evaluation. This process includes considerations of resource, data, programmatic and method feasibility. Unlike previous steps, all questions in this step are crucial, and a negative response to many may indicate challenges in implementing the evaluation. Key areas of focus include the following.

Resource feasibility:

- Is there adequate expertise and budget for the evaluation?

Data feasibility:

- Can a proper counterfactual be identified?
- Are baseline data available or can they be collected?
- Are the data of good quality, and will the evaluation team have access to necessary data from government counterparts?

Programmatic feasibility:

- Does the programme's scope support impact evaluation?
- Is the impact evaluation plan integrated into the programme design?
- Will the intervention remain consistent throughout the evaluation period?
- Does the evaluation comply with ethical standards?

Method feasibility:

- Can randomization occur before the intervention for RCTs?
- If randomization is not possible, can quasi-experimental methods be implemented?
- Are there sufficient historical data to analyse the intervention's impact?
- For qualitative methods, is there enough information to follow a robust scientific process?

This structured approach ensures that impact evaluations are justified, suitable and feasible, ultimately leading to informed decision-making and effective programme planning.

Annex III. Checklist for evaluating the need for impact data collection

1. Do outcome data fulfil programme management requirements? (Yes/No)
2. Is prioritizing measuring impact over actionable performance monitoring data more important for the programme? (Yes/No)
3. Is your project/programme amenable to impact assessment? (Yes/No)
4. Do you want to ascertain the contribution of UNDP? (Yes/No)
5. Could a control or comparison group be established and baseline data collected post-intervention? (Yes/No)
6. Is the programme to be scaled or expanded? (Yes/No)
7. Do impact data improve the overall effectiveness of the programme? (Yes/No)
8. Does the absence of an impact data system hinder programme performance? (Yes/No)
9. Can prioritizing measuring impact benefit the programme? (Yes/No)
10. Will there be mechanisms within the programme for data utilization? (Yes/No)
11. Will there be attention to building capacities for data collection and use? (Yes/No)
12. Do the anticipated benefits of implementing an impact data system justify the financial, human and time resources required to establish and maintain it? (Yes/No)

If the response is 'Yes' to 10 of the above 12 questions, develop a right-fit impact evidence system, in which the data collected are practical, relevant and aligned with the organization's goals, enabling effective decision-making and programme enhancement.

Independent Evaluation Office
United Nations Development Programme
One UN Plaza, DC1-20th Floor
New York, NY 10017, USA
Tel. +1 (646) 781 4200

 [/www.undp.org/evaluation](http://www.undp.org/evaluation)

 [/UNDP_Evaluation](https://twitter.com/UNDP_Evaluation)

 [/evaluationoffice](https://www.youtube.com/evaluationoffice)

 [/company/undp-ieo](https://www.linkedin.com/company/undp-ieo)