

Answers to Frequently Asked Questions on Development Evaluation

Albert-Eneas Gakusi

African Development Bank, Abidjan, Côte d'Ivoire

Email: albertg75011@gmail.com

How to cite this paper: Gakusi, A.-E. (2026). Answers to Frequently Asked Questions on Development Evaluation. *Modern Economy*, 17, 869-908.
<https://doi.org/10.4236/me.2026.177045>

Received: March 11, 2026

Accepted: July 21, 2026

Published: July 24, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper provides answers to frequently asked questions on development evaluation. Drawing on selected literature and the author's experience in independent development evaluation at the African Development Bank, it presents key concepts, principles, and practices that underpin the discipline. The paper discusses the definition and purpose of evaluation, its distinction from related disciplines, the respective roles of self-evaluation and independent evaluation, the main types of evaluation, evaluation processes, methods used to generate credible evidence, mechanisms for quality assurance, and the use of evaluation findings to support learning and accountability. It is intended as a practical reference for evaluators, development practitioners, policymakers, and other stakeholders interested in understanding and applying evaluation in development contexts. As the paper reflects evaluation practice up to 2020, it does not address more recent developments, including the use of artificial intelligence in evaluation, the methodological adaptations prompted by the COVID-19 pandemic, or the recent implications of evolving international development aid policies.

Keywords

Development Evaluation, Independent Evaluation, Evaluation Methods and Approaches, Evaluation Criteria, Evaluation Quality Assurance, Evaluation Use

1. Introduction

Presented in a question-and-answer format, this paper addresses frequently asked questions on development evaluation. It serves both as a general guide to development evaluation and as a practice note informed by the experience of the Independent Development Evaluation (IDEV) function of the African Development

Bank (hereafter referred to as “the Bank”). Drawing on selected publications, including the Good Practice Standards of Multilateral Development Banks and the evaluation criteria and standards of the Development Assistance Committee of the Organisation for Economic Co-operation and Development (OECD/DAC), as well as the author’s professional experience, the paper translates established evaluation principles into practical responses to issues commonly encountered in evaluation practice. It is intended as an accessible reference for evaluators, development practitioners, policymakers, and other stakeholders interested in understanding the principles and practice of development evaluation.

The paper addresses a range of topics, including the definition of evaluation; its relationship with related disciplines; the respective roles of self-evaluation and independent evaluation; the different types of evaluation; evaluation processes; methods for generating credible evidence; quality assurance; and the use of evaluation findings and recommendations.

References are provided at the end of the paper. Annex 1 presents the development evaluation criteria and suggested evaluation questions, Annex 2 contains a selection of commonly used evaluation concepts, and Annex 3 provides a list of acronyms.

2. Methodology to Identify the Frequently Asked Questions on Development Evaluation

The paper is based on a compilation of questions submitted by IDEV staff concerning development evaluation concepts, methods, practices, and institutional arrangements. An initial set of 66 questions was collected and subsequently reviewed to identify overlaps, duplication, and thematic similarities. Questions addressing the same concept or methodological issue were merged, resulting in a consolidated set of 28 questions that form the structure of this paper. The consolidation process sought to preserve the substance of the original submissions while improving clarity, reducing repetition, and ensuring broad coverage of key evaluation topics. While the questions originated from IDEV staff and therefore primarily reflect the perspectives and information needs of evaluation practitioners within a multilateral development bank, the paper subsequently benefited from external review by senior evaluation professionals from the Independent Evaluation Office of the United Nations Development Programme (UNDP), the Independent Office of Evaluation of the International Fund for Agricultural Development (IFAD), and two experienced independent evaluators. Their comments helped strengthen the relevance, clarity, and applicability of the paper for a broader evaluation audience.

3. Frequently Asked Questions

3.1. What Is Evaluation and What Is Meant by Development Evaluation?

Evaluation is a discipline that applies social research theories and methods for

determining the value, merit, worth or significance of entities to be evaluated, i.e. the *evaluands* or more communally the *interventions*. The evaluands might be projects, programs, organizations, policies, strategies, practices, products, processes or services, personnel, proposals, job applications, etc. Typically, evaluation uses the following steps: defining the evaluand and the evaluation purpose, clarifying evaluation questions, developing an appropriate research design, identifying relevant standards of value indicating which criteria to use to evaluate, collecting appropriate data and synthesizing the results of the analysis to respond to the evaluation questions, presenting findings and drawing explicitly evaluative conclusions that says how good, valuable or important the evaluand is (Davidson, 2005; Scriven, 1991). The two fundamental purposes of any evaluation are accountability and learning for improvement. However, Cook notes that current evaluation practice tends to privilege accountability over learning, limiting the ability of evaluations to contribute meaningfully to program improvement (Cook, 2025).

Development evaluation is part of a broader evaluation landscape and it follows the same steps to promote economic and social development. According to the Development Assistance Committee of the Organization for Economic Co-operation and Development (OECD/DAC): “Evaluation is the systematic and objective assessment of an on-going or completed project, program or policy, its design, implementation and results”. Evaluation aims to determine the relevance and coherence of objectives, development effectiveness, efficiency, impact and sustainability. The coherence criterion, recently adopted to examine the compatibility of an intervention with other policies and actions, is grounded in earlier work on policy coherence for development (OECD, 2002, 2005; Picciotto, 2004). In the post-2015 agenda, coherence has been expanded to include inclusive and sustainable development (OECD, 2015, 2018, 2023a). Recent OECD guidance emphasizes that the DAC criteria should be applied thoughtfully and adapted to context rather than used mechanically, with evaluators expected to justify how each criterion is interpreted and weighted (OECD, 2021).

An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors. Evaluation also refers to the process of determining the worth or significance of an activity, policy or program (OECD/DAC, 1991, 2010a, 2019).

An evaluation is intended to provide credible information about all types of impacts, including intended and unintended, positive and negative, direct and indirect. It needs to demonstrate the *causality* or *attribution* to determine if the program was responsible of any results of development interventions. However, in development evaluation, more often than not, evaluation is only able to assess the *contribution* to an intervention especially when it is financed by multiple partners. Even then, it is often not possible to assess the specific contribution to results by an individual partner. In practice, development interventions often pay more attention to activities and outputs than to outcomes and impacts that reflect real

changes in people's lives and development conditions (Burt, 2006).

In line with the *Principles for Evaluation of Development Assistance* (OECD/DAC, 1991) and the revised evaluation criteria (OECD/DAC, 2019), as well as recent guidance on their application (OECD, 2021), the independent development evaluation of the Bank conducts evaluations to: a) improve future development policies, programs and projects through feedback of lessons learned; b) provide a basis for accountability, including the provision of information to the public; and c) promote an evaluation culture in the Bank and in Africa. This is done by strengthening national monitoring and evaluation systems, establishing regional networks and communities of practice, establishing evaluation platforms, strengthening national evaluation associations, etc. (IDEV, 2013).

In this paper, a distinction is made between attribution and contribution. Attribution refers to establishing that observed results were caused by a specific intervention and would not have occurred in its absence. Such claims generally require rigorous experimental or quasi-experimental designs capable of isolating the effects of the intervention from other influencing factors. Contribution, by contrast, assesses the extent to which an intervention plausibly influenced observed results within a broader context where multiple actors and external factors are also at work. Because most development interventions operate in complex environments, development evaluations often focus on assessing contribution rather than demonstrating direct attribution. Patton (2008a, 2008b) highlights that measuring development assistance requires approaches that generate credible and useful evidence for decision-making, emphasizing utilization-focused evaluation as a way to ensure that findings meaningfully inform policies and programs.

3.2. What Is the Difference between Evaluation, Monitoring, Auditing, and Research?

While evaluation shares common features with monitoring, auditing and research, it is still a fast-growing discipline whose specificities include drawing overall evaluative conclusions and making judgements from multiple lines of evidence about a single evaluand for learning and accountability. Evaluation uses information from monitoring, audit, and research and judges the merit, quality and significance of the entity evaluated. Reciprocally, evaluation produces useful information for those different functions.

3.2.1. Monitoring and Evaluation

Although closely related, monitoring and evaluation are two distinct processes with different purposes. Monitoring is an internal management responsibility using indicators to systematically measure progress and efficiency of on-going activities, to take remedial actions, and to update plans toward achieving the desired results. Logical frameworks are useful for monitoring as they specify indicators and means of verification. The baseline information gives the reference points for demonstrating change towards the set targets and the level of achievement of objectives. Monitoring needs also to take account of changes in the external envi-

ronment and major assumptions. Evaluation, on the other hand, develops clear chains of reasoning needed to answer questions about quality, merit and pertinence of policies, programs and projects. It relies on multiple sources of evidence including monitoring data to draw conclusions to support decisions about what to do next and to account for intended and unintended results. It is often carried out with the help of thematic or/and sectoral external experts. It can be carried out at any stage of the evaluand cycle: design quality, implementation, completion or ex-post. Evaluation may be carried out by an internal team, but to ensure credibility, ex-post evaluation should be conducted by an independent team that is given the freedom to evaluate without recourse to operational management in coming to conclusions and recommendations (see the question 3.6 below on independent evaluation).

Within the Bank, operations departments are responsible for monitoring and reporting on implementation of operations. They prepare supervision reports, portfolio reviews, project completion reports, country and regional strategy completion reports, and annual development effectiveness reviews. A more comprehensive and proactive monitoring approach was introduced in 2013 with the creation of the Delivery and Performance Management Results Department responsible for measuring progress in portfolio performance across sector divisions, countries and regions. While there have been changes in the organization of mainstreaming performance monitoring over time, the aim remains to highlight areas requiring continued or increased attention through a portfolio dashboard which indicates the performance of operations under execution. The results of the monitoring activities by operations departments including those of the Delivery and Performance Management Results Department are inputs for evaluation. They help identify issues and questions to be addressed, while providing data for evaluation. In turn, evaluation findings can inform improvements in monitoring systems and indicators.

3.2.2. Audit and Evaluation

As already mentioned, evaluation determines the merit, worth or significance of the evaluand including the results obtained and their sustainability. An audit is rather an assessment of the adequacy of management controls in order to ensure the economical and efficient use of resources; the safeguarding of assets; the reliability of financial and other information; the compliance with regulations, rules and established policies; the effectiveness of risk management; and the adequacy of organizational structures, systems and processes (a compliance audit). An audit may extend to an assessment of value-for-money and the achievement of results (a performance or value-for-money audit), which is a complement to evaluation. Audit provides assurance and accountability to stakeholders.

While audit and evaluation differ in purpose, the skills required and methods used, they have some similarities. Both disciplines analyze the data objectively and within defined standards and criteria. In short, auditing tends to focus on compliance with requirements, while evaluation tends to focus on attributing observed

changes to a policy, program, or project (Morra-Imas & Rist, 2009). In the Bank, for accountability reasons and for improvement of future projects and policies, both audit and evaluation are independently conducted and reported to the Board of Directors through the Audit and Finance Committee (AUF) for Audit and the Committee on Operations and Development Effectiveness (CODE) for evaluation. Audit and evaluation can be carried out at any time of the implementation of the activities concerned. Both disciplines are mutually supportive and organizationally close.

3.2.3. Social Science Research and Evaluation

The two activities are at times quite similar. Program evaluation is an example of applied social sciences. It requires knowing the basic principles of the social sciences including quantitative and qualitative research methods to collect data and analyse them. However, evaluation is much more applying social science methods to solve social problems. “The evaluation process identifies relevant values or standards that apply to what is being evaluated, performs empirical investigation using techniques from social sciences” (Scriven, 1991, 2004). The evaluation makes a synthesis of information from multiple sources in order to integrate relevant observations to draw conclusions useful for decision making. Both research and evaluation use empirical inquiry methods and techniques, but often differ in their purpose. Evaluation usually aims to help solve identified development issues through responding to evaluation questions. The results of evaluation are used to improve development strategies and determine how resources will be allocated to different programs.

Evaluation is transdisciplinary. It draws on the methods and theories of diverse disciplines and professions to design and conduct appropriately relevant and rigorous evaluations. Among those disciplines are sociology, political science, economics, psychology, communications, management, information technology, health sciences, education, organizational development, natural sciences; statistics, etc.

Research seeks to understand relationships among variables or describe phenomena in order to develop knowledge that can be generalized and applied. Its emphasis is on hypothesis testing to generate knowledge about how the world works and why it works as it does.

3.3. What Are the Types of Evaluation?

There are many different types of evaluations used at different times in the cycle of a project, program, or policy (Schwandt, 2015). The most important distinction made in evaluation types is between *prospective*, *formative* and *summative* evaluation. *Prospective evaluations* focus on probable future outcomes. *Formative evaluations* focus on implementation and improvement. *Summative evaluations* focus on results, enabling to make decisions to continue, replicate, scale up, or end a given project, program, or policy (Morra-Imas & Rist, 2009). Both self- and in-

dependent evaluations can be carried out at any stage of a policy, program or project cycle.

3.3.1. Prospective or Ex-Ante Evaluation

This type of evaluation is sometimes called an *ex-ante evaluation*. An ex-ante or a prospective evaluation is concerned with forward-looking, future-oriented questions to assess the likely outcomes of projects, programs, or policies. It is carried out in order to determine what would be the actions to be taken in order to successfully implement development activities and attain development objectives. An ex-ante or prospective evaluation often include program theory reconstruction or assessment and scenario studies as well as summaries of existing research and evaluation to ascertain the empirical support for proposed initiatives. This kind of evaluation was introduced in the Bank for the public sector with the first publication of the *Readiness Review* in 2009 aiming at assessing the quality at entry of all public sector operations before their approval by the Board of Directors. In the same year an *Additionality and Development Outcomes Assessment* (ADOA) was adopted for the private sector for the same purpose and for improving quality at entry in terms of design and selection of private sector operations.

3.3.2. Formative or Process Evaluation

This type of evaluation is conducted during the implementation of a project or program to understand how it is working and to take any remedial measures. A special type of formative evaluation is sometimes called *process evaluations*, which consists of investigating the process of delivering the program, including alternative delivery procedures. A formative evaluation looks into the ways in which a program, policy, or project is implemented. It assesses and documents the degree of fidelity and variability in program implementation, expected or unexpected, planned or unplanned. It concentrates on activities and outputs. With its strong focus on monitoring, the independent evaluation of *Paris Declaration on Aid Effectiveness* is an illustration of a formative evaluation (Wood et al., 2011).

Another type of formative evaluation is a midterm or midpoint evaluation, which is conducted about halfway through a project, program, or change in policy to help identify which features are working well and which are not. Related to this type of evaluation but quite different is *Developmental evaluation* that aims to help complex innovations adjust to complex, nonlinear dynamics in uncertain environments (Patton, 2010). Finally, *real-time evaluation* also takes place during the course of implementation. Its time frame is short relative to that of other types of evaluations. Its use is critical in the context of humanitarian operations with each evaluation typically lasting days, rather than weeks (Herson & Mitchell, 2005).

3.3.3. Summative, Outcome or Impact Evaluation

Summative evaluation is often called an *outcome or impact evaluation*. Relative subtypes are ex-post evaluation, end-evaluation and retrospective evaluation. A

summative evaluation is conducted after the completion of an intervention or on a mature intervention to determine the extent to which anticipated results were realized. A summative evaluation is intended to provide information about the worth and impact of the program. Summative evaluations are used to answer questions such as: What changes were observed in targeted populations, organizations or policies during and at the end of the intervention? To what extent can the observed changes be attributed to it? Were there unintended effects which were not anticipated at the planning stage? Were they positive or negative? Was the intervention efficient or cost effective and are the results sustainable? Such evaluations draw conclusions and lessons learned to help take decisions to continue the intervention with the same conditions or with modifications, to scaling it up or to discontinuing it, or to support future interventions under similar conditions. Most of IDEV's evaluations including thematic, sectoral and country strategy evaluations are summative. For reasons of credibility, a summative evaluation is much more likely to involve external evaluators than do prospective or formative evaluations.

3.4. What Is a Meta-Evaluation?

By definition, a meta-evaluation is the evaluation of an evaluation to determine its quality and/or value (Scriven, 1991). A meta-evaluation assesses the quality of an evaluation through the analysis of its merit, validity, utility, conduct, credibility, and costs. Evaluations should produce valid and justifiable conclusions; be useful to the clients and other relevant audiences; be conducted in an ethical, legal, professional, and otherwise appropriate manner; be credible to relevant audiences; and be as economical, quick, and unobstructive as possible. While evaluating an evaluation report, getting inputs from stakeholders other than the authors of the report is essential to complete a fair-meta-evaluation (Davidson, 2005). The synthesizing of the assessment's results helps to reach a judgment about the merits of the evaluation. The term is also often used to assess a group of evaluations for their quality according to the above characteristics.

Meta-evaluation can be applied at nearly any stage of an evaluation, from evaluating its planned questions and methods, to a mid-evaluation review, to evaluating the completed report by submitting it to independent consultants or external critics to ensure impartiality and rigor (Shadish, 1998). The above-mentioned evaluation of Paris Declaration contains an audit statement by an independent audit team attesting to the quality of evaluation. The audit positively responded to the questions as to whether the findings can be trusted, whether the evaluation was conducted independently, and whether the evaluation process was rigorous. It has become standard to commission an independent review of major high-stakes evaluations to determine whether the evaluation meets generally accepted international standards of quality. "Meta-evaluation should not be confused with meta-analysis which is an approach to synthesizing quantitative studies on a common topic" (Scriven, 1991).

3.5. What Is Equity Focused Evaluation?

Equity-focused evaluation is still an emerging area of work initiated by UNICEF and aiming at designing and managing evaluations to assess the contribution of policies, programs and projects to equitable development results to put more emphasis on social justice in development interventions. It is a response to increased focus on equity in human development at the international level. It strives to respond to the following questions: a) how can one strengthen the capacity of Governments, organizations and communities to evaluate the effect of interventions on equitable outcomes; b) what are the evaluation questions to ensure interventions are relevant and are having an impact in decreasing inequity, are achieving equitable results, and are efficient and sustainable; and c) what are the methodological implications in designing, conducting, managing and using Equity-focused evaluations?

Equity-focused evaluation puts an emphasis on refining and refocusing existing evaluation approaches and techniques to evaluate to what extent policies, programs and projects achieve equitable development results. The judgement is based on the usual evaluation criteria of relevance, effectiveness, efficiency, impact and sustainability—and, in humanitarian settings, coverage, connectedness and coherence—of the interventions. It involves a rigorous, systematic and objective process in the design, analysis and interpretation of information in order to answer specific questions, including those of concern to worst-off groups. It provides assessments of what works and what does not work to reduce inequity, and it highlights intended and unintended results for worst-off groups as well as the gaps between best-off, average and worst-off groups. It provides strategic lessons to guide decision-makers and to inform stakeholders regarding the steps needed to move toward equitable development (Bamberger & Segone, 2011).

The preceding sections discuss concepts, principles, and approaches that are broadly applicable to development evaluation across organizations and contexts. The following sections draw more specifically on the experience and institutional arrangements of the Independent Development Evaluation function of the African Development Bank. While many of the issues discussed are relevant to evaluation practice more generally, the examples and procedures described reflect the Bank's governance structure, evaluation policies, and operational processes.

3.6. What Is the Organizational Arrangement for the Evaluation Function in the Bank?

As for other multilateral development banks (MDBs), the evaluation function at the Bank is composed of self-evaluation and independent evaluation which are carried out separately from each other and are, in practice, mutually complementary and reinforcing each other through cross-fertilization of knowledge and experience.

3.6.1. Self-Evaluation

Self-evaluation is conducted by operational departments responsible for designing

and implementing sectoral, country and regional integration strategies, programs, projects, and corporate service activities. Like independent evaluation, it is undertaken for two purposes: a) to account for results, and b) to learn from the experience to inform and improve policies, strategies and operations. Also, self-evaluation is undertaken at different stages of project cycle including design, implementation and completion.

At design level, the Bank has introduced *Readiness Reviews* in 2009 to assess the quality of all public sector operations before their approval by the Board of Directors. In the same year, an *Additionality and Development Outcomes Assessment* (ADOA) was instituted for the private sector operations for the same purpose. Both initiatives represent a significant improvement in operations planning in that they involve evaluation as an integral part of every intervention from the beginning and thus better assess the likelihood of success or failure and improve the quality-at-entry. In addition to the readiness review and ADOA, self-evaluation comprises periodic supervision reports, mid-term evaluations, country and regional portfolio reviews, sectoral reviews, annual portfolio performance reviews, project completion reports, annual development effectiveness reports and other analytical works.

For all Bank projects, a supervision mission is undertaken at least once per year to monitor all conditions that could affect implementation progress and achievement of project objectives. Those conditions include compliance with loan conditions and covenants, procurement performance, financial performance, progress on activities and physical works, project management, and expected impact on development. A supervision report is prepared at the end of each supervision mission and at the end of each project, the last supervision report recapitulates the implementation problems and the measures taken to resolve them. This report will be the basis for the preparation of a Project Completion Report (PCR). All those self-evaluation reports are the foundation of the work to be carried out by IDEV for the independent evaluations. In principle, PCRs are subject to an independent review by IDEV to attest to their quality.

3.6.2. Independent Evaluation

Independent evaluation at the Bank is grounded in the Independent Evaluation Policy (IDEV, 2016a) amended in July 2019 (IDEV, 2019) and is aligned with OECD-DAC Principles for Evaluation of Development Assistance and Good Practice Standards on independent evaluation (ECG, 2010). It is also guided by the Quality Standards for Development Evaluation (OECD/DAC, 2010b), which outline the essential elements of credible, impartial, and methodologically sound evaluations across all stages of the evaluation process. It has the following dimensions:

- *Organizational independence*: Evaluation staff are free from the control of those responsible for designing and implementing the activities to be evaluated. They have unconstrained and timely access to the information they

need to fulfill their mandate within ethical and legal constraints. IDEV is located outside the management line functions and reports to the Board of Directors through CODE;

- *Behavioral independence*: Reflects the capacity and willingness of IDEV to design its work program, to produce high-quality and uncompromising reports, and to disclose its findings to the Boards of Directors and the public without any management-imposed restrictions;
- *Freedom from outside influence and pressure*: This keeps the independent evaluation function free to set its own priorities, design its processes, craft its products, reach its judgments, and administer its human and financial resources without the interference of the management;
- *Avoidance of conflicts of interest*: Ensures that current, immediate future, and prior professional and personal relationships and considerations do not influence evaluators' judgments or create the appearance of a lack of objectivity (IDEV, 2016a, 2019).

Like for self-evaluation, independent evaluation is undertaken at different stages of the evaluand cycle including design, implementation, and completion. Independent evaluation brings necessary credibility given all the above dimensions of independence. These allow evaluators to be impartial and free from pressure throughout the evaluation process.

To deliver on its mandate, in addition to the front office of the Evaluator General, IDEV is organized in three divisions respectively in charge of the evaluation of infrastructure and the private sector; the evaluation of agriculture, governance, human development, country, regional and corporate activities; and a division in charge of knowledge management, outreach, and strengthening evaluation capacity in regional member countries. The three divisions work in a collaborative way and complement each other.

Through CODE, the Board of Directors oversees and holds accountable the independent evaluation function. It is responsible for safeguarding the independence of the evaluation function in the Bank and for providing oversight and guidance to IDEV. It approves the work program and the budget. CODE discusses evaluation reports and the management's responses to identify general policy or strategic issues for consideration by the Board of Executive Directors. CODE also monitors the implementation of evaluation recommendations (IDEV, 2016a).

3.7. How Are Independent Evaluations Carried Out at the Bank?

The conduct of the evaluation follows a transparent process detailed in *Steps and Instructions for Completing an Evaluation* (IDEV, 2020) and illustrated in *Process Mapping Flowchart* available in IDEV. It starts with internal discussion between the management and the task manager of the evaluation in IDEV. Then the department engages with the operations department in charge of the operations to be evaluated in order to get a better understanding of the issues to resolve and to define the evaluation needs and questions accordingly. This step leads to the

preparation of an *approach paper* which is shared with all stakeholders involved in the evaluation for comments which are addressed in the final version.

The task manager of the evaluation then prepares the terms of reference of the evaluation for the consultant and the department launches an international competitive bidding most of the time. Once the consultant is recruited, he/she is invited to the headquarters to discuss details of the assignment with IDEV management and the task manager before starting interviews with managers and task managers of the operations to be evaluated. On the basis of documentation review, stakeholder interviews and sometimes scoping missions, the consultant prepares an *inception report*, which, if successful, is followed by a field work under the leadership of the task manager. It is worth noting that an increased number of inception reports are being prepared in-house by task managers. The field mission lasts about two to four weeks depending on the volume and depth of information to be gathered. Thereafter, the evaluation team organizes a debriefing meeting with the main stakeholders during which feedback is provided on the preliminary findings. The debriefing meeting is also an opportunity to correct any factual error in the interpretation of facts and to get further guidance on relevant information to respond to the evaluation questions.

The data collection during the field work often includes participatory methods, which help to ensure that the views of the implementing agencies and the end users of the interventions are taken into account, increase the credibility and reliability of information and foster the ownership of the results of the evaluation by the stakeholders. Local consultants usually participate in the evaluation and they are used for the preparation of the mission, the collection of information including administration of surveys, participation in focus group discussions and direct observation to record what is seen and heard from the project's site, data and documents analysis, as well as the preparation of the reports. After the field work phase the following activities are executed: analysis of data and documents, preparation of background papers and consultations with operations departments through reference groups, internal and external reviews of the draft reports, and preparation of the summary report, which is sent to management to prepare their response to the evaluation. The summary report and the management response are sent to CODE for its consideration after which the evaluation report is disseminated in hard copies and electronically.

3.8. What Is the Normative Approach Used by IDEV for Its Evaluations?

In accordance with Multilateral Development Banks' harmonization of evaluation methodologies, performance indicators, and evaluation criteria under the Evaluation Cooperation Group (ECG), IDEV conducts *objectives-based evaluations*. The purpose of an objectives-based evaluation is to determine whether the program's objectives have been achieved while taking into account the pertinence of those objectives and the implementation efficiency and sustainability of the re-

sults. However, the evaluations go beyond anticipated results to take into account negative or positive unanticipated effects of the interventions, if any.

In practice, objectives-based evaluations respond to the evaluation questions on the following agreed evaluation criteria by OECD/DAC Network on Development Evaluation and ECG (see Appendix for more details).

- *Relevance*: Is the intervention doing the right things? The extent to which the intervention objectives and design respond to beneficiaries', global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change.
- *Coherence*: How well does the intervention fit? The compatibility of the intervention with other interventions in a country, sector or institution.
- *Effectiveness*: Is the intervention achieving its objectives? The extent to which the intervention achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups.
- *Efficiency*: How well are resources being used? The extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way.
- *Impact*: What difference does the intervention make? The extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.
- *Sustainability*: Will the benefits last? The extent to which the net benefits of the intervention continue or are likely to continue.

The downside of this evaluation approach is that, more often than not, program developers are not very sure of what would be the magnitude of the results beyond outputs. Also, the attainment of the stated objectives does not necessarily determine the worth of the intervention, which requires an assessment of the context, the objectives themselves and the mechanisms leading to success or failure. More recently there has been an emphasis on the use of the theory of change *inter alia* to try to overcome those limitations of objectives-based evaluations, while explaining why a program's activities will lead to desired outcomes.

Jess Dart (2004) identifies several normative orientations—such as experimental, pragmatic, participatory, constructivist, goal-free, and theory-based approaches—that reflect different assumptions about how merit and worth should be assessed. These alternative perspectives enrich evaluative reasoning by highlighting varied ways of defining value, interpreting evidence, and understanding program effects in complex development settings.

3.9. How and Who Evaluates the Work of IDEV? Is It Possible to Assess the Effectiveness of This Work on a Country or a Regional Basis?

In line with the Bank's Independent Evaluation Policy, the Bank's Board of Directors oversees and holds accountable the independent evaluation function through CODE. It is responsible for providing oversight and guidance to IDEV. Among

others, CODE reviews IDEV's evaluation reports and advises the Board of Directors on the adequacy of its evaluation activities, including through consideration of IDEV's reporting on its own deliverables and achievements. It also discusses and endorses the IDEV work program and budget, for approval by the Board of Directors. The Evaluator General presents to CODE an annual account of activities and achievements during the year.

In addition, the Evaluation Cooperation Group established in 1996 and composed of the evaluation departments of the Multilateral Development Banks and of which IDEV is a member, has instituted a system of peer review, through which evaluation functions can be asked to be assessed by their peers against the ECG's Good Practice Guidelines. Finally, an independent external evaluation could be commissioned as it has been the case for most of MDBs ([United States Department of the Treasury, 2014](#)).

IDEV influences the work of the Bank at project, program, country and regional levels through its findings and recommendations presented to CODE. Once those recommendations are adopted they are implemented by operations departments. While there may be instances where it is possible to trace the effect of an IDEV evaluation recommendation in similar projects in a country, or in even rarer cases in a specific country policy as confirmed by a relevant official, in most cases it would be an impossible task to disentangle the effects of IDEV evaluations at country or regional levels. In fact, IDEV recommendations are only one source of information for the Bank's management to decide. Moreover, many operations are co-financed by other development partners and they are being implemented in contexts where there are multiple actors. The effect that IDEV evaluations have on the Bank funded projects in a country or a region may eventually have a country-wide or a region-wide impact, but this could be very difficult to ascertain with any precision. Outcomes at country and regional levels are determined by multiple factors and they could rarely be attributed to an evaluation.

3.10. What Is the Role of IDEV in Dealing with the Challenges Identified in the Regional Member Countries?

IDEV attempts to address the challenges faced within the limits of its mandate and resources. A lack of reliable monitoring and results data and weak monitoring and evaluation capacity in regional member countries (RMCs) are challenges faced not only by IDEV and other parts of the Bank Group but also by other development agencies and the government authorities. Various parts of the Bank Group as well as other donors are involved in strengthening data quality and national capacities, including Quality Assurance and Results Department (ORQR) and the Chief Economist's complex. Concrete examples include the Bank's support for statistical capacity building, and capacity-building trainings provided by the African Development Institute (EADI). In this, the Bank collaborates with other partners wherever possible. Specifically, in the area of evaluation capacity development, IDEV is pursuing three initiatives, set out in the Annual Report and

the IDEV 2016-2018 Work Program: The Strengthening National Evaluation Systems initiative (aimed at governments), the African Parliamentarians' Network on Development Evaluation (aimed at parliamentarians), and the Evaluation Platform for Regional African Development Institutions (aimed at sub-regional Development Financial Institutions). In these endeavors, IDEV also works with partners, including UNICEF, UNDP, UN Women, and the Centers for Learning on Evaluation and Results (CLEAR).

IDEV has addressed the challenges associated with evaluating countries in transition by adjusting its ways of working to fit the local situation, where possible. For example, when in the course of the evaluation of the Bank's strategy and program in Burundi, field visits by IDEV staff were not possible due to the security situation, IDEV engaged local evaluators and experts and communicated with them by telephone and videoconference in order to obtain the necessary data. Only as a last resort does IDEV abandon an evaluation, as in the case of Sierra Leone, where the Ebola outbreak made conducting an evaluation there impossible.

3.11. How Does IDEV Get Informed about the Realities of a Country or a Region in Order to Better Understand Their Challenges and Specific Economic Bottlenecks?

Usually, at the design stage of an evaluation, a scoping mission is undertaken which consults all the main stakeholders both at Bank Headquarters and at country or regional level, including the Executive Directors, sector specialists, the country economist and country program officer, the regional economist and Director General, government authorities, Regional Economic Communities, private sector, Non-Governmental Organizations and associations. For instance, for its evaluation of the Bank's regional integration strategy and operations in Central Africa, the evaluation team has carried out a scoping mission in Bangui, Libreville and Yaoundé and met all main stakeholders in order to refine the theory of change of the strategy, the evaluation questions and the methodology.

Also, at the start of each evaluation, an internal Bank Reference Group is established, which provides comments and guidance throughout the evaluation process. The Reference Group is composed of the most relevant Bank departments for the evaluation. In the case of country strategy and program evaluations, this includes the Resident Representative and members of the Country Team. For the evaluation of regional integration strategies, the Reference Group includes the regional department, the Regional Integration Coordination Unit, which has replaced the Department on Regional Integration, Trade & NEPAD (ONRI) since April 2016, Resident Representatives of the regions and sectoral experts. In addition, IDEV is in close contact with both the Field Office and the government authorities throughout the evaluation. Using a broad range of data collection sources including document reviews, interviews with key informants, focus group discussions, and visits to selected project sites ensures that IDEV is well informed of the country situation.

3.12. What Is the Difference between an Approach Paper and an Inception Report?

Generally, an evaluation starts with the preparation of a concept note, followed by a scoping mission if necessary and then an *approach paper*, which is a key document providing a comprehensive guidance on what the evaluation is about. Depending on evaluation, the approach paper may be replaced by an inception report or by extended terms of reference in the case of Country Strategy and Program Evaluation.

An *approach paper* typically covers the following topics: background; purpose and scope of the evaluation; a brief description of the portfolio and the strategy to be influenced by the evaluation; an initial theory of change as relevant, evaluations questions; criteria against which to evaluate; evaluation approach and design; draft evaluation matrix; management of the evaluation; plan for dissemination of the results; and initial estimate of resources and timeline required to complete the evaluation. It is advisable that the approach paper includes criteria for the quality of the evaluation report to guide the production of the reports by the evaluation team and internal and external expert reviews of the evaluation reports. When consultants are hired to support the evaluation, the approach paper informs the preparation of the terms of reference for the consultant.

The *inception report* specifies how the evaluation will be operationalized and includes results of stakeholders' consultations and possibly a literature review. It provides the final agreed scope of the exercise, any modification of the theory of change, the evaluation questions, the evaluation design, methods and tools of data collection (notably the sampling strategy, questionnaires, interview guides) and analysis, including the final evaluation matrix, and confirms the resources and time required to deliver different outputs of the evaluation, building on the approach paper. The document presents a working outline of the final evaluation report as well as the tools to be used for data collection. It is one of key deliverables, which is reviewed and shared with the IDEV management for clearance and subsequently shared with the concerned operations departments and/or country and region teams. It is also posted on internet among the other evaluation deliverables.

3.13. What Kinds of Evaluations Does IDEV Conduct; What Are Its Evaluation Products?

The Bank's Independent Evaluation Strategy 2013-2017 sets out the following evaluation types:

- Project Completion Report validation notes (public sector);
- Extended Supervision Report validation notes (private sector);
- Project Performance Evaluation Reports (public sector);
- Project Performance Evaluation Reports (private sector);
- Project cluster evaluations;
- Thematic evaluations;
- Sector evaluations;

- Country Strategy Evaluations;
- Regional Integration Strategy Evaluations;
- Corporate evaluations;
- Impact evaluations;
- Evaluation synthesis.

For each evaluation, IDEV prepares an approach paper and/or an inception report, a draft report for discussion with stakeholders, and a final summary report that is presented to CODE along with a management response. IDEV also prepares knowledge products for its evaluations, such as evaluation briefs or highlights, an annual report setting out all of its activities, and a quarterly knowledge magazine—*Evaluation Matters*—covering diverse evaluation themes.

3.14. How Does the Clustering Project Evaluation Take into Account the Specificity of Each Project?

Compared to individual project evaluations, project cluster evaluations are actually an efficient and cost-effective way to generate relevant lessons of experience from the implementation of similar projects to address common problems in different settings. The lessons can be incorporated into future project design (Sanders, 1997).

As indicated in the previous section, cluster evaluation is only one of many evaluation approaches. IDEV mainly conducts project cluster evaluations in the context of broader sector evaluations (transport, energy, water, etc.) in order to supplement the other data gathered and to gain deeper insights into specific areas that are particularly relevant for the Bank's involvement in that sector.

Cluster evaluations are based on in-depth project-level evaluations as building blocks. They provide a synthesis of these building blocks through an assessment of development effectiveness, what works, what does not, and why at the sub-sector level. They outline common findings and lessons, but also note any differences observed across the projects. As such, cluster evaluations can inform future project design. Specific projects are also used as illustrative examples. Project-level evaluations are available and can be consulted if further information is required at the project level.

3.15. What Criteria Does IDEV Use to Select a Country for Evaluation?

Usually the number of Country Strategy and Program Evaluations (CSPEs) conducted per year is around 3 - 4 yearly. However, due to the design of the Comprehensive Evaluation of the Bank's Development Results (CEDR), the number of CSPEs had to be scaled up significantly in order to achieve sufficient coverage in terms of portfolio size (60% of the Bank's lending volume) and representativeness of types of countries (Low Income Countries, Middle Income Countries, Fragile States).

The standard criteria used to select countries, applied for example to define

IDEV's 2016-2018 Work Program (IDEV, 2015), are:

- *Coverage*: each Bank country strategy and program should be covered at least once over a period of 10 years.
- *Timeliness*: evaluations are timed so that the report can inform the development of the next Country Strategy Paper (CSP). CODE's guidance is that the evaluation report should be discussed at the latest at the same time as the initial presentation of the new CSP pillars when the country strategy and program has been evaluated.
- *Relevance*: the extent of the relevance/significance of the areas of intervention of the Bank in terms of portfolio size and/or innovation, including new/pilot initiatives, move to blend status, etc.
- *Primary stakeholder interest*: the extent of the interest and/or concern expressed by the primary stakeholders of IDEV (Board/CODE, Management) during the consultations process undertaken as part of the work program definition.

3.16. Does IDEV Share the Lessons Learned from Country Evaluations with Those Countries?

Yes, of course. Prior to finalizing an evaluation, IDEV often holds a preliminary findings workshop in the country with the participation of a broad spectrum of stakeholders, including the government authorities, representatives of civil society, and local development agencies. These gatherings serve to share and discuss the findings of the evaluation team with the country stakeholders. In addition, after the evaluation has been presented to CODE, IDEV implements a tailored communication and dissemination plan, which includes country actors among the primary audience targeted. Dissemination often takes place through the Bank's Field Office, which is best placed to advise on the pertinent local audiences. A range of knowledge products publicly available on IDEV's website complements the targeted dissemination. Finally, the Evaluation Results and Lessons Learned Database, available on the Bank's website (<http://evrd.afdb.org>), provides a searchable repository of lessons from all of IDEV's evaluations.

3.17. How Does IDEV Ensure High Quality of Its Evaluation Products?

IDEV aims to produce quality evaluations which ask relevant questions, carry out good research while applying appropriate methodologies to provide lessons for future policies and programs. With the quest for quality enshrined in the Independent Evaluation Strategy 2013-2017, which was extended by one year and amended in 2019, IDEV has made extensive efforts to increase the quality of the evaluation function in the Bank. Strict adherence to international good practice standards has become the norm and new tools have also been designed and rolled out. For instance, the Project Results Assessment (PRA) template is used to systematically assess the project's relevance, coherence, efficiency, effectiveness and sustainability. It is accompanied by a detailed guidance note for each evaluation

criterion. Workshops are also conducted to strengthen internal capacity including by learning new approaches.

IDEV also ensures that all evaluations undergo both internal and external peer reviews. This comes on top of the consultation with management through a stakeholders' reference group created for every evaluation to validate the accuracy and usefulness of the findings.

3.18. What Does IDEV Do to Ensure Evaluations Are Used at the Bank?

Evaluation is not an end itself. It is meant to help improve the relevance, coherence, effectiveness, efficiency impact and sustainability of the Bank's interventions. Towards this end, feedback to both policy makers and operations staff is essential. In order to ensure that its evaluations will be used, IDEV collaborates with the operations departments from the start of the evaluation to its completion. Hence, the findings and recommendations should not come as a surprise. They need to be actionable to facilitate their implementation. Also, IDEV publishes its reports and disseminates them both physically and electronically. It prepares briefs, highlights and other knowledge products, and makes active use of its revamped website and the quarterly magazine *Evaluation Matters*. It organizes meetings and knowledge sharing events at the headquarters and in shareholder countries. This is consistent with the policy and practice in other MDBs.

In the same vein, in 2015, IDEV launched a computer-based management action record system (MARS) to track the implementation of the concrete actions the Bank Management has committed to in response to recommendations from IDEV evaluations. IDEV prepares a MARS annual report and presents it to CODE. Management also reports bi-annually to CODE on the status of the implementation of the recommendations.

3.19. What Is an Evaluation Approach; How to Select One?

Like in any other discipline, evaluation contains different schools of thoughts or approaches which are a set of distinct conceptual options used to design and carry out part or all of the tasks involved in evaluation. **Stufflebeam (2001)** notes that evaluation approaches can be grouped into distinct models that reflect different assumptions about what evaluation is for and how it should be conducted, ranging from objectives-oriented and management-oriented models to expertise-oriented, consumer-oriented, and participant-oriented approaches. Those approaches may be complementary or conflicting about what evaluation is and how it should be carried out. These different views have different sources including philosophical and ideological beliefs; methodological preferences or practical choices. The choice of an evaluation approach needs to be a function of what is useful in each evaluation approach depending on the specific evaluation questions or challenges.

There are many evaluation approaches, and some of them are listed below (a comprehensive list is available here: <http://betterevaluation.org/approaches>):

- *Objectives-oriented*: focus on goals/objectives and the degree to which they are achieved.
- *Decision-management-oriented*: identifying and meeting informational needs of decision makers.
- *Experimental*: produces an estimate of the mean net impact of an intervention by comparing results between a randomly assigned control group and treated group or groups; *quasi-experimental approach* is used when random assignment is not feasible.
- *Participatory evaluation*: stakeholders (clients and users) centrally involved in process and/or making decisions about the evaluation.
- *Theory-based*: construction of program theory while specifying causal models that explain how a program is intended to achieve intended outcomes.
- *Contribution Analysis*: developed to address questions about the feasibility of concluding that an intervention has contributed to an outcome in circumstances where a direct causal relationship is difficult to demonstrate.
- *Case study*: research design focusing on in-depth understanding of an evaluation and in its context using a combination of qualitative and quantitative data.

3.20. What Methods/Methodologies Are Employed in Evaluation?

There are three types of methods employed in evaluation: Quantitative, qualitative, and mixed methods. Use of any methods depends upon the evaluations, and any evaluation usually needs a combination of different methods for different aspects of the evaluation. This is because, no single evaluation methodology could fully capture all of the complexities of how programs operate in the real world.

Quantitative methods put a high emphasis on impartiality, accuracy, objectivity and the validity of the information generated. Included under scientific-experimental models would be: experimental and quasi-experimental designs; cost-benefit and cost-effectiveness analyses.

Qualitative methods emphasize the importance of observation, the need to understand the evaluation context, and the value of subjective human interpretation in the evaluation process. They are essential when quantitative methods are not possible or feasible, and they can provide valuable insights and findings that quantitative methods may not provide. Qualitative methods include: participants observation, direct observation, unstructured interviews and structured interviews; focus group interviews; and case studies.

Mixed method approaches link and combine qualitative and quantitative data to create a synergistic evaluation, whereby one method enables the other to be more effective and together both methods to provide a fuller understanding of the evaluation problem.

There are sometimes hot debates among quantitative and qualitative methods proponents with each claiming the superiority of their position. However, in reality, most good evaluators are familiar with all the above categories and borrow from each as the need arises. Each of them brings something valuable to the eval-

uation. In recent years, attention has been put on how one method might integrate results from evaluations that use different strategies.

3.21. What Is Triangulation?

Triangulation is an attempt to produce evaluation results using different theories, analyses, methods and sources of information. This is especially important when use of a rigorous approach has not been possible. Therefore, triangulation is often essential. Evaluation findings are strengthened when several pieces of evidence from different sources broadly arrive at the same conclusion. Information may come from interviews, document analysis, surveys, and other evaluation methods: quantitative, qualitative and mix methods. Even when rigorous methods are used, qualitative information can reinforce findings and add depth to findings. Triangulation helps avoid different sorts of biases from sampling, measurement, and other evaluation procedures. Use of triangulation can help innovate in the ways of thinking and doing business. IDEV evaluations are based on triangulation of different sources of information to respond to evaluation questions. However, triangulation may lead to different or inconsistent results. In this case, the evaluation needs to try to determine the reason(s) for difference(s) in order to arrive at a more consistent set of findings. If this does not resolve the issue, there is the option to report the inconsistency (ies) and to recommend that additional evaluative research be undertaken and/or that the evaluation be redesigned.

Triangulation is particularly important in contribution-oriented evaluations, where conclusions are based on the convergence of evidence from multiple sources rather than on direct attribution established through experimental or quasi-experimental methods.

3.22. What Is Evaluability Assessment and How Is It Carried Out?

An *evaluability assessment* is a process to identify whether an intervention can be evaluated, and whether an evaluation is justified, worth the effort and resources to be spent, feasible and likely to provide useful information. It assesses whether the evaluation can achieve its objectives within given constraints: time frame to evaluate outcomes or impacts, the proposed budget and available time. It indicates whether the objectives are clear, relevant and measurable. The evaluability indicates any particular challenges which could stand in the way of the evaluation: political, social, cultural, technical, geographic accessibility, weather, etc. In addition to helping decide whether an evaluation is worthwhile in terms of its likely benefits, consequences and costs, an evaluability assessment can provide information to help decide whether a program needs to be modified, whether it should go ahead or be stopped.

The assessment answers the questions: a) does the quality of the design of the program allow for the evaluation; b) are the results of the program verifiable based on the planned collection systems; c) would the evaluation be feasible, credible and useful? Each question has detailed sub-questions to help assess the level of

quality while using a rating system.

The evaluability assessment is based on different sources of information including: a) review of program documentation; b) analysis of the information system defined in the program (or related to the program) and determination of the information needs; c) evaluations of similar operations; d) interview of the main stakeholders (task managers, policy makers and representatives of the assumed beneficiaries and program managers; e) analysis of the program while reconstituting its theory including assumed mechanisms for successful implementation, and an assessment of their likely success in the prevailing conditions of the program (see the internet site of Better Evaluation).

3.23. What Is Theory of Change and Why Is It Important?

According to Funnell and Rogers (2011), along with a *Theory of Action*, *Theory of Change* is part of a *Program Theory*. A program theory explicates how and why an intervention is understood to contribute to its intended or observed outcomes. The theory of change “explains how activities are understood to produce a series of results that contribute to achieving the final intended outcomes and impacts”. The theory of action shows “the ways in which interventions are constructed to activate the theory of change”. Hence, for any intervention, it becomes possible to distinguish between a *theory failure*, things may be done right but they do not succeed because the theory is wrong, and *implementation failure*, the theory may be right but not well implemented. Carol H. Weiss (1995, 1997) popularized the term theory of change as a way to describe the set of assumptions that explain both the ministeps that lead to the long-term goal of interest and the connections between program activities and outcomes that occur at each step of the way.

The theory of change can be used for the designers especially of complex interventions to be specific about the theories guiding their work in laying out the sequence of outcomes that are expected to occur as the result of an intervention, and planning an evaluation strategy around tracking whether these expected outcomes are actually produced. Any good results framework requires clarity with respect to the theory of change. A useful theory of change must articulate clear causal pathways, make underlying assumptions explicit, and link activities to outcomes in ways that are both plausible and verifiable, thereby strengthening the credibility and evaluability of the results framework (Mayne, 2015).

Formulating a theory of change requires using different sources of information including (Rogers, 2014): a) needs assessment or determinant analysis that identifies what must be in place for success; b) documented objectives that are clear, relevant and measurable; c) previous evaluations and research on similar programs or policies, particularly those that include analysis of how the programs/policies work; d) expert opinion on these types of programs/policies; e) perspectives of staff, managers, partners and beneficiaries about how (not whether or not) the intervention works, or fails to work; f) feedback from relevant stakeholders on draft versions of the theory of change; and g) research-based theories

about how change occurs.

The interest of using theory of change in evaluation is that it helps not only to know if the program produced the intended outcomes but also why those outcomes appeared—or failed to appear. Nevertheless, theory of change has its own limitations. For one program, there may be several theories. Also, one of the greatest challenges is that evaluation methods are not always up to the questions to be answered. Furthermore, evaluators themselves are not always trained in relevant social science theories and/or in rigorous methods to test the multiple relationships between stated variables.

3.24. What Are Different Sampling Methods Used?

In many cases, an evaluation cannot cover all projects and is obliged to focus on a sample of projects to evaluate. However, more often, the unit of analysis is not a project but a participant or a beneficiary or some other stakeholder. There are two main types of sampling methods (probability and non-probability) and several techniques used to select a sample for the evaluation. The method and technique depend on the evaluation questions, resources, time, and desired level of accuracy.

Probability sampling. It involves the random selection of evaluation participants in a manner that gives each member of the population a non-zero chance of being selected for the sample. The benefit to using probability sampling is that the sample will be fully representative of the population, and the results will be generalizable to the population. The disadvantage to using probability sampling techniques is that they can be very time-consuming and costly. They also require more technical skills not always available in most evaluators.

Nonprobability sampling. When using nonprobability sampling techniques the potential for bias is high and applicability of results is low because the sample is made up of units that were selected specifically by the evaluator, referred by others, or were simply convenient or available at the time. However, these techniques are less resource intensive and are generally more accessible and convenient to work with.

3.24.1. Types of Probability Sampling

Simple Random Sampling. This technique gives all units in the population an equal opportunity for being selected by using a method that will select units completely at random. One way of performing a simple random sample commonly used in evaluation for selecting units includes using a basic lottery system or drawing numbers/names from a hat. This method is very effective if the size of the population is small or moderate. If the population is large it is recommended to use sampling software to select a sample or a random numbers table usually found in statistics books.

Systematic random sampling. This sampling technique selects the sample size based on the 'sampling interval'. For example, the selection of n individuals out of N will begin by the random selection of a first individual in the interval $[1, N/n]$. Given this random start point, every N^{th}/n individual will be selected systematic.

cally through the end of the list, yielding the sample of n individuals.

Stratified random sampling: This technique divides the population into meaningful homogenous or similar groups based on a certain characteristic (e.g., gender, language, regions, socioeconomic status, ethnicity) and then selects a simple random sample from each group. Stratified sampling ensures greater representativeness on a characteristic of interest within the population. This method is a more commonly used sampling technique than simple random sampling because it allows you to study a wider range of the population without a larger sample size. A limitation of stratified sampling is that as the number of stratifications increases, the size of the sample must also increase if statistically significant results are to be obtained.

Cluster Random sampling: It is a form of random sampling in which the population is divided into comparable groups (schools, farms, regions, etc.) and some of the groups are randomly chosen to be part of the sample. The main difference between strata and clusters is that strata are composed of similar elements from which to sample whereas in clusters, elements are heterogeneous and sampling is only on the selected clusters to study.

3.24.2. Types of Non-Probability Sampling

Convenience sampling: This technique uses whatever units from the population that are available to participate at a given time. Convenience sampling is often called grab sampling. This technique has very little structure. The only criterion for selection is that the unit selected is a member of the population and is available to participate in the evaluation at the time required. For example, using a convenience sampling technique, an evaluator would select the first 100 people who received a copy of a program's college planning handbook to take a survey on its user-friendliness.

Purposive sampling: This technique is conducted by allowing the evaluator to select a sample that he/she feels is representative of the population. Purposive sampling relies on the knowledge of an evaluator and his/her understanding of the design of the program to choose the most appropriate and representative sample. The evaluator subjectively chooses units which he/she believes are representative of the entire population and try to ensure that the full spectrum of variation and diversity of the population is represented in the sample. It is advisable to use a panel of expert advisors to assess the representativeness and the objectivity of the sample.

Quota sampling: This technique predefines specific groups within the population to investigate and then samples from the population to fill a quota for each group. Quota sampling is similar to stratified sampling but does not use random sampling to choose units from each group. The goal, however, is to still choose a sample that closely matches the population ratio/breakdown on a characteristic or set of characteristics of interest (e.g., age, gender, race, etc.). The evaluator should continue to select units to participate in the sample until all groups are filled.

Snowball sampling: This technique uses the process of referrals to build the sample size. The snowball sampling begins with a few key individuals to include in the sample. The interviewer asks them to participate in the evaluation, as well as recommend other people they might know with similar characteristics to the population he/she is studying that could also participate in the evaluation. Then he/she contacts the individuals he/she was referred to and ask them to recommend others for the evaluation as well, thus growing the sample size through referrals.

Self-selection sampling: This technique allows participants to volunteer and decide that they would like to be part of the evaluation sample. Self-selection sampling is usually done by advertising the evaluation and asking for volunteers. This allows potential participants to contact the interviewer and to volunteer to participate in the evaluation.

Given the complexity of the Bank's projects which are composed of multiple components, the selection of projects to be evaluated for IDEV evaluations are mainly based on non-probability sampling with a blend of purposive and quota selection techniques. This is the case for most quality at entry evaluations as well as for the Comprehensive Evaluation of the Development Results (IDEV, 2016b). However, some evaluations use probability sampling at the level of interviewees when enough information is available to design and implement them. This was the case of two cluster evaluations undertaken in 2015 in power interconnection and rural electrification subsectors.

3.25. What Do You Do When the Sample Size or Survey Response Is Not a Good Size or Statistically Relevant?

In such a situation there are several possibilities depending on prevailing constraints: the time required to complete the evaluation, resources available, the pressure to get results. If those constraints are removed, therefore the ideal is to collect more data until you get a relevant statistically significant sample. This would require better organization to follow-up the survey and push people to respond.

Otherwise, if there is not enough responses to be able to say that there is a statistically representative sample, then one must report that the findings represent only the opinions expressed by those in the sample (those who responded). Indeed, in these circumstances, one cannot conclude that the opinions of the sample represent those of the people who were supposed to respond and did not. There are tables indicating for each population size and the precision required, the minimum number of people that must be surveyed to get a representative sample at a certain level of confidence.

3.26. How Do You Evaluate with Limited or Poor Quality Data?

An evaluation with limited and poor data is illustrated by the evaluation of microfinance carried out by IDEV in 2013. The Bank did not have any consolidated database from four implementing departments of microfinance and at each de-

partment level the existing information was scant and poor. Many of the projects were components of bigger sectoral projects and no information were collected on the components to help assess the performance of those projects. Of the 91 microfinance projects approved over 10 years, 35 were ongoing. To collect relevant information to analyze the relevance, performance and sustainability, it was decided to limit the work to a sample of projects composed of all those 35 ongoing projects. A questionnaire with a set of questions on relevant information was prepared and sent to the task managers. A junior consultant and a research assistant were tasked to closely follow-up the exercise and push the task managers to respond. Almost all of them responded comprehensively and this helped create a database, which facilitated responses to the evaluation questions. From the sample, 7 projects were selected for case studies in order to compare the information provided by the task managers and those collected by IDEV team from the ground (Gakusi et al., 2018).

In such a situation, it is paramount to work with sectoral or subject experts having hands-on experience with the kind of projects being evaluated as the evaluation could rely also on their experience in other settings.

3.27. What Is an Economic Evaluation?

Economic evaluation refers to a range of different evaluation techniques, including cost-benefit analysis and cost-effectiveness analysis to measure efficiency. Cost-benefit and cost-effectiveness analyses are used for assessing whether or not the costs of an activity can be justified by the outcomes and impacts. The two techniques are used to inform the decisions regarding alternative use of resources. An intervention can only be cost-effective relative to some other intervention or doing nothing. Thus, both techniques assume that the decision makers face a set of alternatives from which choices must be made and that the criteria for making decision includes benefits or effects and the resources which will be expended to get them (the costs).

Cost-benefit analysis measures both inputs and outcomes of alternatives in monetary terms. Cost-benefit analysis permits the calculation of Net Present Value (NPV) or Internal Rate of Return (Financial and Economic), which may be compared against an expected opportunity cost. In the cases where both costs and benefits can be measured in monetary units, decision makers would choose only among alternatives whose net benefits exceed costs. The most attractive alternatives would be those with the highest net benefit-cost ratios (or NPV or IRR).

Cost-effectiveness analysis estimates input costs in monetary terms and outcomes in non-monetary quantitative terms, such as improvements in student reading scores or in health status. For almost all social interventions, outcomes cannot be easily converted into market outcomes. They are evaluated in terms of effectiveness with relation to the objectives of competing interventions while using agreed upon evaluation criteria.

Economic evaluation needs, however, to adopt a broader view to take into ac-

count other benefits and costs to society—secondary effects consisting of positive or negative externalities—accounting for all resources used by the project, whether human, technological or natural, and gauges the value the project generates to all stakeholders, to determine whether society at large gains from the investment. Nevertheless, in spite of their obvious value when public and private resources are limited and several courses of action are feasible, cost-benefit and cost-effectiveness analyses are not used extensively in program evaluation. Thus far, the Bank has not yet undertaken any evaluation of use of economic evaluation in its projects, which is mostly prevalent for infrastructure operations. In 2010, the Independent Evaluation Group (IEG) of the World Bank evaluated the use of cost-benefit analysis in development projects and concluded that: “The percentage of Bank projects that are justified by cost-benefit analysis has been declining for several decades, owing to a decline in adherence to standards and to difficulty in applying cost-benefit analysis (...). The Bank’s use of cost-benefit analysis for decisions is limited because the analysis is usually prepared after the decision to proceed with the project has been made.” (IEG, 2010).

3.28. How to Evaluate Policy-Based Operations (PBOs)?

The policy lending instruments were introduced in the 1980s essentially to provide support for the balance of payments, and later extended to serve broader objectives of policy and institutional reforms through general and sectoral budget support. Thus, *Policy Based Operations* were set up to support a set of policies and public spending actions of a partner country government in order to foster economic growth and poverty reduction. For evaluation, the chief question is to assess to what extent and under which circumstances those operations have successfully enhanced the policies, strategies and spending actions of the partner government to efficiently achieve sustainable national and/or sector level development outcomes and a positive impact on economic growth and poverty reduction.

The daunting challenge for the evaluation is that of *attribution* of the results to the operations financed by one donor given the fact that PBOs often operate at the economy-wide level or at the sectoral level, where different resources are put in the general budget or a sectoral common fund. While following the principle of objectives-based evaluation, the evaluation of PBOs focuses on policy and institutional actions supported and the resulting changes in macroeconomic, social, environmental, and human development outcomes. The PBO’s intended outcomes are taken from the program’s statement of objectives and results chain.

Within this framework, it is possible to evaluate the results of PBOs as a whole program while using the development evaluation criteria regardless of the sources of financing:

- *Relevance*: alignment with strategy papers and performance assessment framework;
- *Coherence*: compatibility of the intervention with other interventions in a country, sector or institution;

- *Effectiveness*: collective outputs and outcomes;
- *Efficiency*: reduction of transaction costs;
- *Impacts*: collective long-term impacts;
- *Sustainability*: technical, financial and economic, institutional sustainability, political and governance, ownership, partnerships, environmental and social sustainability, resilience to exogenous factors and risk management.

For these operations, disbursement delay is of a limited concern as they are disbursed in a limited number of instalments. Different aspects to look at and methods to use are developed in the Big Book on Evaluation Good Practice Standards (ECG, 2012).

Overall, the daunting challenge for the evaluation is that of attribution, given that policy-based operations are implemented alongside other government actions and development partner interventions. In such contexts, evaluations often focus on assessing the contribution of the operation to observed policy and development outcomes rather than establishing direct attribution.

4. Conclusion

This paper provides a practical and comprehensive overview of development evaluation, drawing on relevant literature and the author's experience within the Independent Development Evaluation function of the African Development Bank. Structured around frequently asked questions, it explains key evaluation concepts, approaches, methods, criteria, and institutional arrangements, while offering guidance on the conduct, quality assurance, and use of evaluations. The paper serves as a useful reference for both practitioners and stakeholders seeking to better understand development evaluation and its role in promoting accountability and learning. However, the paper reflects the state of evaluation practice up to 2020 and was prepared before the widespread adoption of artificial intelligence (AI) tools in evaluation. Consequently, it does not discuss the opportunities and risks associated with AI-assisted evaluation, including data analysis, evidence synthesis, and quality assurance (Naeem et al., 2025; UK Evaluation Society, 2025; UNICEF, 2024). In addition, it does not address the methodological, operational, and ethical challenges that emerged during the COVID-19 pandemic, such as restrictions on fieldwork and the increased reliance on remote data collection. Nor does it consider the more recent uncertainties affecting development cooperation, including shifts in development aid priorities and funding associated with changes in United States administrations, which have significant implications for evaluation planning, resource allocation, and the assessment of development effectiveness.

This paper has other limitations. First, although it draws on internationally recognized evaluation literature, including OECD/DAC criteria and Multilateral Development Bank good practice standards, it does not provide a comprehensive or systematic review of the evaluation field. Second, the questions that structure the paper originated from IDEV staff and therefore reflect the information needs and

practical concerns of evaluators working within a multilateral development bank context. While the paper benefited from external review by evaluation professionals from other international organizations and independent experts, some topics of interest to other evaluation communities may receive less attention. Finally, many of the examples and institutional arrangements discussed are informed by the experience of the African Development Bank's Independent Development Evaluation function, although the concepts and principles presented are intended to be relevant to development evaluation more broadly.

Acknowledgements

This paper updates a document prepared by the author while serving as Chief Evaluation Officer at the Independent Development Evaluation (IDEV) of the African Development Bank until December 2020. It acknowledges the guidance of Rakesh Nangia, Evaluator General of IDEV until August 2018, who suggested its preparation. IDEV managers and staff contributed by identifying frequently asked questions on evaluation and reviewing drafts. The paper benefited from careful review by Indran Naidoo, Director of the Independent Evaluation Office of UNDP; Fabrizio Felloni, Deputy Director of the Independent Office of Evaluation of IFAD; the late John Eriksson, an esteemed evaluation mentor; and Maximilien Tereraho, independent consultant. It also incorporates the recent revision of development evaluation criteria by the OECD/DAC Network on Development Evaluation and benefited from comments by Andrew Anguko, Chief of Quality and Methods at IDEV. The author is also grateful to the anonymous reviewers of this publication for their constructive comments and suggestions, which helped improve the clarity, rigor, and overall quality of the paper. The author alone remains responsible for the content of the paper.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Bamberger, M., & Segone, M. (2011). *How to Design and Manage Equity-Focused Evaluations*. New York: UNICEF Evaluation Office.
http://mymande.org/sites/default/files/EWP5_Equity_focused_evaluations.pdf
- Burt, P. (2006). *Moving from Outputs to Outcomes: Practical Advice from Governments Around the World*. Washington, DC: IBM Center for The Business of Government & The World Bank. January.
<https://web.archive.org/web/20220121012248/https://www.businessofgovernment.org/sites/default/files/PerrinReport.pdf>
- Cook, T. (2025). *Getting Back to "What Works"? Débats de l'IEEP n° 9, 4-7*.
<https://sciencespo.hal.science/hal-04926448v1>
- Dart, J. (2004). *Six Normative Approaches to Evaluation*.
<https://www.semanticscholar.org/paper/Six-Normative-Approaches-to-Evaluation-Dart/9d4b9bae005200bb6df0f3519c8bedeca533fea9>

- Davidson, E. D. (2005). *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*. Sage.
- ECG (2010). *Good Practice Standards on Independence of International Finance Institutions' Central Evaluation Departments*.
<https://www.ecgnet.org/document/gps-independence-international-financial-institutions-central-evaluation-departments>
- ECG (2012). *Big Book on Evaluation Good Practice Standards*. December.
<https://www.ecgnet.org/sites/default/files/FINAL%20Big%20Book%20on%20GPS.pdf>
- Funnell, S. C., & Rogers, P. J. (2011). *Purposeful Program Theory: Effective Use of Theory of Change and Logic Models*. Jossey-Bass.
- Gakusi, A.-E., Negre, A., Diomanade, M. et al. (2018). Statistical and Qualitative Assessment of the Design Quality, Performance, Effectiveness and Sustainability of the Assistance of the African Development Bank to Microfinance, 2000-2012. *The African Statistical Journal*, 20, 14-55.
- Herson, M., & Mitchell, J. (2005). Real-Time Evaluation: Where Does Its Value Lie? *Humanitarian Exchange*, No. 32, 43-45.
<http://odihpn.org/wp-content/uploads/2005/12/humanitarianexchange032.pdf>
- IDEV (2013). *African Development Bank Independent Evaluation Strategy 2013-2019*. African Development Bank.
<https://idev.afdb.org/en/document/african-development-bank-independent-evaluation-strategy-2013-2019>
- IDEV (2015). *Work Program 2016-2018*. African Development Bank.
https://idev.afdb.org/sites/default/files/documents/files/IDEV%20Work%20Program%202016_2018%20EN%20Web%20final.pdf
- IDEV (2016a). *Revised Independent Evaluation Policy (Revised Version 4)*. African Development Bank.
<https://idev.afdb.org/sites/default/files/Evaluations/2020-06/Revised%20AfDB%20Evaluation%20Policy%20EN.pdf>
- IDEV (2016b). *Comprehensive Evaluation of the Development Results of the African Development Bank Group 2004-2013, Synthesis Report*. African Development Bank.
https://idev.afdb.org/sites/default/files/Evaluations/2020-03/IDEV%20CEDR%20Report%20EN_web.pdf
- IDEV (2019). *Independent Evaluation Policy, Amended in July*. African Development Bank.
<https://www.ecgnet.org/sites/default/files/Revised%20AfDB%20Evaluation%20Policy%20EN.pdf>
- IDEV (2020). *Steps and Instructions for Completing an Evaluation*. African Development Bank.
- IEG (2010). *Cost-Benefit Analysis in World Bank Projects*. Washington DC: The World Bank.
<https://openknowledge.worldbank.org/bitstream/handle/10986/2561/624700PUB0Cost00Box0361484B0PUBLIC0.pdf?sequence=1>
- Mayne, J. (2015). Useful Theory of Change Models. *Canadian Journal of Program Evaluation*, 30, 119-142. <https://doi.org/10.3138/cjpe.230>
- Morra-Imas, L. G., & Rist, R. C. (2009). *The Road to Results: Designing and Conducting Effective Development Evaluations*. The World Bank.
<https://doi.org/10.1596/978-0-8213-7891-5>
- Naeem, M., Smith, T., & Thomas, L. (2025). Thematic Analysis and Artificial Intelligence: A Step-By-Step Process for Using ChatGPT in Thematic Analysis. *International Journal*

- of *Qualitative Methods*, 24, 1-18. <https://doi.org/10.1177/16094069251333886>
- OECD (2002). Policy Coherence for Development. In *Development Co-Operation Report 2001: Efforts and Policies of the Members of the Development Assistance Committee*, Chapter 2 (pp. 23-53). OECD Publishing.
- OECD (2005). *Policy Coherence for Development: Promoting Institutional Good Practice*. OECD Publishing.
- OECD (2015). *Better Policies for Development 2015: Policy Coherence and Green Growth*. OECD Publishing.
- OECD (2018). *Policy Coherence for Sustainable Development 2018: Towards Sustainable and Resilient Societies*. OECD Publishing.
- OECD (2021). *Applying Evaluation Criteria Thoughtfully*. OECD Publishing. <https://doi.org/10.1787/543e84ed-en>
- OECD (2023a). *Driving Policy Coherence for Sustainable Development: Accelerating Progress on the SDGs*. OECD Publishing.
- OECD (2023b). *Glossary of Key Terms in Evaluation and Results Based Management for Sustainable Development* (2nd ed.). OECD Publishing.
- OECD/DAC (1991). *Principles for Evaluation of Development Assistance*, Paris. <http://www.oecd.org/development/evaluation/50584880.pdf>
- OECD/DAC (2010a). *Glossary of Key Terms in Evaluation and Results Based Management*, Paris. <http://www.oecd.org/development/peer-reviews/2754804.pdf>
- OECD/DAC (2010b). *Quality Standards for Development Evaluation*, Paris. <http://www.oecd.org/development/evaluation/qualitystandards.pdf>
- OECD/DAC (2019). *Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use*. <https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf>
- Patton, M. (2008a). *State of the Art in Measuring Development Assistance*. https://www.researchgate.net/publication/268337535_State_of_the_Art_in_Measuring_Development_Assistance
- Patton, M. (2008b). *Utilization Focused Evaluation* (4th ed.). Sage Publications.
- Patton, M. (2010). *Developmental Evaluation Applying Complexity Concepts to Enhance Innovation and Use*. Guilford Press.
- Picciotto, R. (2004). *Policy Coherence and Development Evaluation: Concepts, Issues and Possible Approaches*. The OECD Workshop on Policy Coherence and Development.
- Rogers, J. P. (2014). *Theory of Change*. UNCEF Methodological Briefs. https://www.unicef-irc.org/publications/pdf/brief_2_theoryofchange_eng.pdf
- Sanders, J. R. (1997). Cluster Evaluation. In E. Chelimsky, & W. R. Shadish (Eds.), *Evaluation for the 21st Century*. Sage.
- Schwandt, T. A. (2015). *Evaluation Foundations Revisited: Cultivating a Life of the Mind for Practice*. Stanford University Press.
- Scriven, M. (1991). *Evaluation Thesaurus* (4th ed.). Sage Publications, Inc.
- Scriven, M. (2004). *Michael Scriven on the Difference between Evaluation and Social Science Research. The Evaluation Exchange*, Winter 2003/2004.
- Shadish, W. (1998). Some Evaluation Questions. *Practical Assessment, Research & Evaluation*, 6, No. 3. <https://doi.org/10.7275/1434-s056>
- Stufflebeam, D. L. (2001). *Evaluation Models: New Direction of Evaluation*. No. 89, 7-98.

[https://bpb-us-e1.wpmucdn.com/sites.psu.edu/dist/7/144284/files/2021/12/3-
Stufflebeam_Evaluation_Models.pdf?utm_source=copilot.com](https://bpb-us-e1.wpmucdn.com/sites.psu.edu/dist/7/144284/files/2021/12/3-Stufflebeam_Evaluation_Models.pdf?utm_source=copilot.com)

UK Evaluation Society (2025). *AI in Evaluation: Good Practice Guidelines for Practitioners*.

[https://evaluation.org.uk/wp-content/uploads/2025/11/UK-Evaluation-Society-AI-in-
Evaluation-Good-Practice-Guidelines-for-Practitioners-V-2.0-1125.pdf](https://evaluation.org.uk/wp-content/uploads/2025/11/UK-Evaluation-Society-AI-in-Evaluation-Good-Practice-Guidelines-for-Practitioners-V-2.0-1125.pdf)

UNICEF (2024). *An Operational Framework for Machine Learning in Evaluation*.

[https://www.unicef.org/evaluation/media/5606/file/ML%20Operational%20Frame-
work%20FINAL.pdf.pdf](https://www.unicef.org/evaluation/media/5606/file/ML%20Operational%20Framework%20FINAL.pdf.pdf)

United States Department of the Treasury (2014). *Evaluation Standards and Practices at the Multilateral Development Banks, a Report to Congress*.

[https://www.treasury.gov/resource-center/international/development-banks/Docu-
ments/2015-03-01%20\(Evaluation%20Report\).pdf](https://www.treasury.gov/resource-center/international/development-banks/Documents/2015-03-01%20(Evaluation%20Report).pdf)

Weiss, C. H. (1995). "Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families". In J. I. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New Approaches to Evaluating Community Initiatives, Concepts, Methods and Context*. Aspen Institute.

<http://files.eric.ed.gov/fulltext/ED383817.pdf>

Weiss, C. H. (1997). Theory-Based Evaluation: Past, Present, and Future. *New Directions for Evaluation*, 1997, 41-55. <https://doi.org/10.1002/ev.1086>

Wood, B. et al. (2011). *The Evaluation of Paris Declaration. Final Report*.

[https://www.gov.uk/government/uploads/system/uploads/attach-
ment_data/file/197482/evaluation-paris-declaration.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/197482/evaluation-paris-declaration.pdf)

Annex 1. What Are the Criteria Used for Development Evaluation?

OECD/DAC Evaluation Criteria

A. Relevance: Is the intervention doing the right things?

The extent to which the intervention objectives and design respond to beneficiaries', global, country, and partner/institution needs, policies, and priorities, and continue to do so if circumstances change. Beneficiaries are defined as "the individuals, groups, or organisations, whether targeted or not, that benefit directly or indirectly, from the development intervention." Other terms, such as rights holders or affected people, may also be used.

"Respond to" means that the objectives and design of the intervention are sensitive to the economic, environmental, equity, social, political economy, and capacity conditions in which it takes place. "Partner/institution" includes government (national, regional, local), civil society organisations, private entities and international bodies involved in funding, implementing and/or overseeing the intervention. Relevance assessment involves looking at differences and trade-offs between different priorities or needs. It requires analysing any changes in the context to assess the extent to which the intervention can be (or has been) adapted to remain relevant.

The following questions are suggested to be considered:

- 1) Relevance of project objectives:
 - Are the activities and outputs of the intervention consistent with the overall goal and the attainment of its objectives in terms of intended effects and impacts?
 - To what extent are the objectives of the intervention still valid?
- 2) Relevance of project design to achieve those objectives:
 - The extent to which the intervention's objectives are clearly stated and focused on outcomes rather than on outputs.
 - The realism of intended outcomes in the prevailing circumstances.
 - The extent to which project design adopted the appropriate implementation arrangements to the identified problems including the assessment risks and of potential negative impacts and how they should have been avoided.
 - The relevance of modifications made to project design.

B. Coherence: How well does the intervention fit?

The compatibility of the intervention with other interventions in a country, sector or institution.

The extent to which other interventions (particularly policies) support or undermine the intervention, and *vice versa*. Includes internal coherence and external coherence:

Internal coherence addresses the synergies and interlinkages between the in-

tervention and other interventions carried out by the same institution/government, as well as the consistency of the intervention with the relevant international norms and standards to which that institution/government adheres. **External coherence** considers the consistency of the intervention with other actors' interventions in the same context. This includes complementarity, harmonisation and coordination with others, and the extent to which the intervention is adding value while avoiding duplication of effort.

The following questions are suggested to be considered:

- 1) Were there a clear articulation and effective coordination of the Bank's interventions?
- 2) Were there complementarities and synergies among those interventions rather than contradictions?
- 3) Were there a clear articulation and effective coordination of donors' interventions?
- 4) Were there complementarities and synergies among those interventions rather than contradictions?
- 5) Were the instruments used the most appropriate to attain the stated objectives?

C. Effectiveness: Is the intervention achieving its objectives?

The extent to which the intervention achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups.

Analysis of effectiveness involves taking account of the relative importance of the objectives or results.

The following questions are suggested to be considered:

- 1) To what extent were the objectives achieved or are likely to be achieved?
- 2) What were the major factors influencing the achievement or non-achievement of the objectives?
- 3) Were there unintended results, positive or negative?

D. Efficiency: How well are resources being used?

Efficiency: how well are resources being used?

The extent to which the intervention delivers, or is likely to deliver, results in an economic and timely way.

"Economic" is the conversion of inputs (funds, expertise, natural resources, time, etc.) into outputs, outcomes and impacts, in the most cost-effective way possible, as compared to feasible alternatives in the context.

"Timely" delivery is within the intended timeframe, or a timeframe reasonably adjusted to the demands of the evolving context. This may include assessing operational efficiency (how well the intervention was managed).

The following questions are suggested to be considered:

- 1) Were activities cost-efficient: Did the benefits of the project (achieved or expected to be achieved) exceed project costs (Cost-benefits analysis); and were the benefits of the project achieved at least cost (cost-effectiveness

- analysis)?
- 2) Were objectives achieved on time?
 - 3) Was the program or project implemented in the most efficient way compared to alternatives?
 - 4) What are the aspects of project design that contributed to or reduced efficiency?

E. Impact: What difference does the intervention make?

The extent to which the intervention has generated or is expected to generate significant positive or negative, intended or unintended, higher-level effects.

Impact addresses the ultimate significance and potentially transformative effects of the intervention. It seeks to identify social, environmental and economic effects of the intervention that are longer term or broader in scope than those already captured under the effectiveness criterion. Beyond the immediate results, this criterion seeks to capture the indirect, secondary and potential consequences of the intervention. It does so by examining the holistic and enduring changes in systems or norms, and potential effects on people's well-being, human rights, gender equality, and the environment.

The following questions are suggested to be considered:

- 1) What has happened as a result of the program or project?
- 2) What real difference has the activity made to the beneficiaries?
- 3) How many people have been affected and what categories have been most affected?

F. Sustainability: Will the benefits last?

The extent to which the net benefits of the intervention continue, or are likely to continue.

Includes an examination of the financial, economic, social, environmental, and institutional capacities of the systems needed to sustain net benefits over time. Involves analyses of resilience, risks and potential trade-offs. Depending on the timing of the evaluation, this may involve analyzing the actual flow of net benefits or estimating the likelihood of net benefits continuing over the medium and long-term.

The following questions are suggested to be considered:

- 1) *Technical Soundness*: The extent to which the project achievements rely on sound technology using inputs efficiently and providing productivity gains. It includes operation and maintenance facilitation, availability of recurrent funding, spare parts, workshop facilities etc.
- 2) *Financial and economic Viability*: The extent to which funding mechanisms and modalities (e.g. tariffs, user fees, maintenance fees, budgetary allocations, other stakeholder contributions, aid flows, etc.) have been put in place to ensure the continued flow of benefits after project completion, with particular emphasis on financial sustainability.

- 3) *Institutional sustainability and strengthening of capacities*: The extent to which the project has contributed to strengthen institutional capacities—including for example through the use of country systems—that will facilitate the continued flow of benefits associated with the project. An appreciation should be made with regards to whether or not improved governance practices or improved skills, procedures, incentives, structures, or institutional mechanisms came into effect as a result of the operation.
- 4) *Political and governance environment*: The extent political and governance developments that could impact the government's priorities with respect to the project. This includes (but is not limited to) upcoming elections or an impending change in government; and other factors that could impact the political commitment to the operation or operational engagement and the political decisions required for sustainability of project results (including laws and the provision of counterpart financing). Special attention should be paid to fraud, corruption and other unethical practices resulting from governance failures.
- 5) *Ownership and sustainability of partnerships*: Whether the project has effectively involved relevant stakeholders, promoted a sense of ownership amongst the beneficiaries (both men and women) and put in place effective partnerships with relevant stakeholders (e.g. local authorities, civil society organizations, private sector, donors) as required for the continued maintenance of the project results.
- 6) *Environmental and social sustainability*: This criterion would normally only apply to Environmental Category I and II projects. It assesses the extent to which the environmental and social mitigation/enhancement measures of the project were implemented, the capacity of country institutions and systems and the availability of funding to ensure the environmental and social sustainability of the operation. The Environmental and Social Safeguards rating in the implementation progress report should be used as a guidance.
- 7) *Resilience to exogenous factors and risk management*: The extent to which the achievements depend on exogenous factors, such as the terms of trade, the world market prices or the political situation in the country and/or in neighboring countries, atmospheric conditions.

Annex 2. Selected Concepts

Attribution	The ascription of a causal link between observed (or expected to be observed) changes and a specific intervention. Attribution refers to that intervention which is to be credited for the observed changes or results achieved. It represents the extent to which observed development changes can be attributed to a specific intervention or to the performance of one or more partners taking account of other interventions, (anticipated or unanticipated), confounding factors, or external shocks.
Cluster Evaluation	An evaluation of a set of related interventions which aims at identifying commonalities and synergies across projects as well as determining the progress made towards the achievement of the related programming objective.
Contribution Analysis	An approach for assessing causal questions and inferring causality in real-life program evaluations. It offers a step-by-step approach designed to help managers, researchers, and policymakers arrive at conclusions about the contribution their program has made (or is currently making) to particular outcomes.
Cost-benefit analysis (CBA)	A quantitative analysis performed to establish whether the present value of benefits of a given project exceeds the present value of costs in monetary terms as compared to other alternatives.
Cost-effectiveness analysis (CEA)	CEA is an alternative to cost-benefit analysis when analysts face the problem to monetize benefits. CEA is commonly used in healthcare, for example, where it is difficult to put a value on outcomes such as the number of lives saved. Like CBA, CEA measures costs in a common monetary value but the effectiveness in terms of physical units.
Desk review	A review of the existing documentation related to the subject to be evaluated.
Developmental evaluation	An evaluation approach to understand the activities of a program operating in dynamic, novel environments with complex interactions. It focuses on innovation and strategic learning rather than standard outcomes and is as much a way of thinking about programs in context and the feedback they produce.
Economic rate of return (also referred to as EIRR)	The ERR of a project is the average annual return to society on the capital invested over the entire life of the project. It is the interest rate at which the project's discounted benefits equal discounted costs, both valued from the entire society's point of view. A project is accepted if the ERR is equal to or exceeds a certain threshold (the social discount rate).
Economic net present value	ENPV is the difference between discounted benefits and costs at a given discount rate. Projects are accepted if the ENPV is positive.
Evaluable	A generic term for whatever is being evaluated: person, performance, program, proposal, product, possibility, and so on. In development evaluation the term refers to a program or system rather than a person.
Development Goal	The higher-order objective to which a development intervention is intended to contribute.
Evaluability assessment	Assessment to determine whether a program meets the preconditions for evaluation and, if so, how the evaluation should be designed to ensure maximum utility. It is the extent to which an activity, a project or a program can be evaluated in a reliable and credible fashion.
Evaluation	The systematic and objective assessment of an on-going or completed project, program or policy, its design, implementation and results. The aim is to determine the relevance and fulfillment of objectives, development efficiency, effectiveness, impact and sustainability.
Experimental evaluation design	Ex-ante evaluation design requiring random assignment of a population to at least two groups, one of which serves as a control or counterfactual.
Financial Rate of Return (also refer to financial internal rate of return)	FRR is the interest rate at which the cost and benefits of a project, discounted over its life, are equal. FRR differs from financial rate of return (or FRR) in that it takes into account the effects of factors such as price controls, subsidies, and tax breaks to compute the actual cost of the project to the economy.

Continued

Focus Group	Structured discussion facilitated by an evaluator. It brings together typically 6 - 10 beneficiaries or program staff for confronting opinions. The facilitator uses a discussion guide and the participants interact to enhance and consolidate the information on themes or domains needing be reconciled, or which concern complex questions that have to be explored in depth. Note-takers record comments and pertinent observations.
Formative evaluation	Evaluation conducted during the implementation, usually focusing on operational issues and oriented towards identifying strengths and shortcomings. Sometimes referred to as process evaluation and mid-term evaluation.
Impact	Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.
Independent evaluation	An evaluation carried out by entities and persons free of the control of those responsible for the design and implementation of the development intervention. The credibility of an evaluation depends in part on how independently it has been carried out. Independence implies freedom from political influence and organizational pressure. It is characterized by full access to information and by full autonomy in carrying out investigations and reporting findings.
Implementation failure	The program does not adequately perform the activities specified in the program design that are assumed to be necessary for bringing about intended social improvements. It includes situations in which no service, no enough service, or the wrong service is delivered, or service varies excessively across the target population.
Intervention	Policy, strategy, program, project, reform, institution or other activity being evaluated.
Internal rate of return	IRR is the interest rate at which the net present value of all the cash flows (both positive and negative) from a project or investment equals zero. It is used to evaluate the attractiveness of a project or investment. If the IRR of a new project exceeds a required rate of return, that project is desirable.
Logic model	A visual representation of a program theory in a diagram or in a table. It helps to clarify objectives of any project, program, or policy. It aids in the identification of the expected causal links—the program logic—in the results chain: inputs, outputs, outcomes, and impact. To get a theory of change, we need to add the assumptions to the causal links in the impact pathways of the logic model.
Meta-analysis	An approach to synthesizing quantitative studies on a common topic.
Meta-evaluation	An evaluation of the merit of evaluations: validity, utility, conduct, credibility, and cost. The term is also used for evaluations designed to aggregate findings from a series of evaluations.
Mixed methods	An evaluation research approach which combines different evaluation frameworks, tools and techniques to use their different comparative advantages in addressing particular concerns and needs.
Non-experimental evaluation design	An evaluation design where no attempt is made to compare intervention and nonintervention groups, and the emphasis is on description. It includes theory and case-based approaches.
Outcome	The likely or achieved short-term and medium-term effects of an intervention's output. Outcomes represent changes in development conditions which occur between the completion of outputs and the achievement of impact.
Outputs	The products and services which result from the completion of activities within a development intervention. Outputs are expected to produce outcomes which lead to impacts.
Participatory methods	Provide active involvement in decision-making for those with a stake in an intervention and generate a sense of ownership in the monitoring and evaluation, results and recommendations.
Program	A program is a portfolio comprising multiple projects that are managed and coordinated as one unit with the objective of achieving outcomes and benefits for the organization. In some organizations program is also used for a national-coverage project or just a large project.

Continued

Program objectives	Specific statements detailing desired accomplishments of a program together with one or more measurable criteria for success.
Program theory	An explicit theory of how and why an intervention contributes to the intended or observed outcomes. It is composed of: a) a theory of change which explains how an intervention is expected to bring about the desired results for an individual, organization or community; and b) a theory of action explaining how an intervention is constructed to activate the theory of change.
Project	A temporary entity established to deliver specific outputs in line with predefined time, cost and quality constraints. A project should always be defined and executed and evaluated relative to an approved business case which balances the costs, benefits and risks of the project. Often projects form a clear and distinct portion of a larger, less precisely identified program.
Quasi-experimental evaluation design	An evaluation design where an intervention and non-intervention or comparison group (counterfactual) are formed either before or after the intervention, but without randomly assigning the population into two groups.
Stakeholders	Individuals, groups or organizations having a significant interest in how well a program functions. They include individuals with decision-making authority over the program, funders and sponsors, administrators and personnel, and clients or intended beneficiaries.
Summative evaluation	Evaluation conducted at the end of an intervention to determine the extent to which anticipated outcomes and impacts were produced. Summative evaluation is intended to provide information about the worth or value of the program.
Theory-based evaluation	An analysis that establishes a plausible association between the various links in the project's results chain, using quantitative and qualitative evidence as well as evidence from other evaluations and academic literature.
Theory failure	The program is implemented as planned but its services do not produce the immediate results on the participants that are expected or the ultimate benefits that are intended, or both.
Triangulation	The use of three or more theories, sources or types of information, or types of analysis to verify and substantiate an assessment. The combination of multiple data sources, methods, analyses or theories, aims to help evaluators overcome the bias that comes from single informants, single methods, single observer or single theory studies.

These concepts and definitions presented in this Annex are aligned with the OECD/DAC Glossary of Key Terms in Evaluation and Results-Based Management (OECD, 2023b), which standardizes terminology across development agencies.

Annex 3. Acronyms

ADOA	Additionality and Development Outcomes Assessment
AfDB	African Development Bank Group
AUFI	Audit and Finance Committee
CEDR	Comprehensive Evaluation of the Bank's Development Results
CLEAR	Center for Learning on Evaluation and Results
CODE	Committee on Operations and Development Effectiveness
COPM	Delivery and Performance Management Office
CSP	Country Strategy Paper
CSPEs	Country Strategy and Program Evaluations
DAC	Development Assistance Committee
EADI	African Development Institute
ECG	Evaluation Cooperation Group
ENPV	Economic Net Present Value
FAQ	Frequently Asked Questions
FRR	Financial Rate of Return
IDEV	Independent Development Evaluation of the AfDB (administratively referred to as BDEV)
IFAD	International Fund for Agricultural Development
IRR	Internal Rate of Return
NPV	Net Present Value
MARS	Management Action Record System
MDBs	Multilateral Development Banks
OECD	Organisation for Economic Co-operation and Development
ONRI	Department on Regional Integration, Trade & NEPAD
OPEV	Operations Evaluation Department
ORQR	Quality Assurance and Results Department
PCR	Project Completion Report
PBOs	Policy Based Operations
PRA	Project Results Assessment
RMCs	Regional Member Countries
UNICEF	The United Nations Children's Fund
UNDP	United Nations Development Programme
UN Women	United Nations Entity for Gender Equality and the Empowerment of Women
