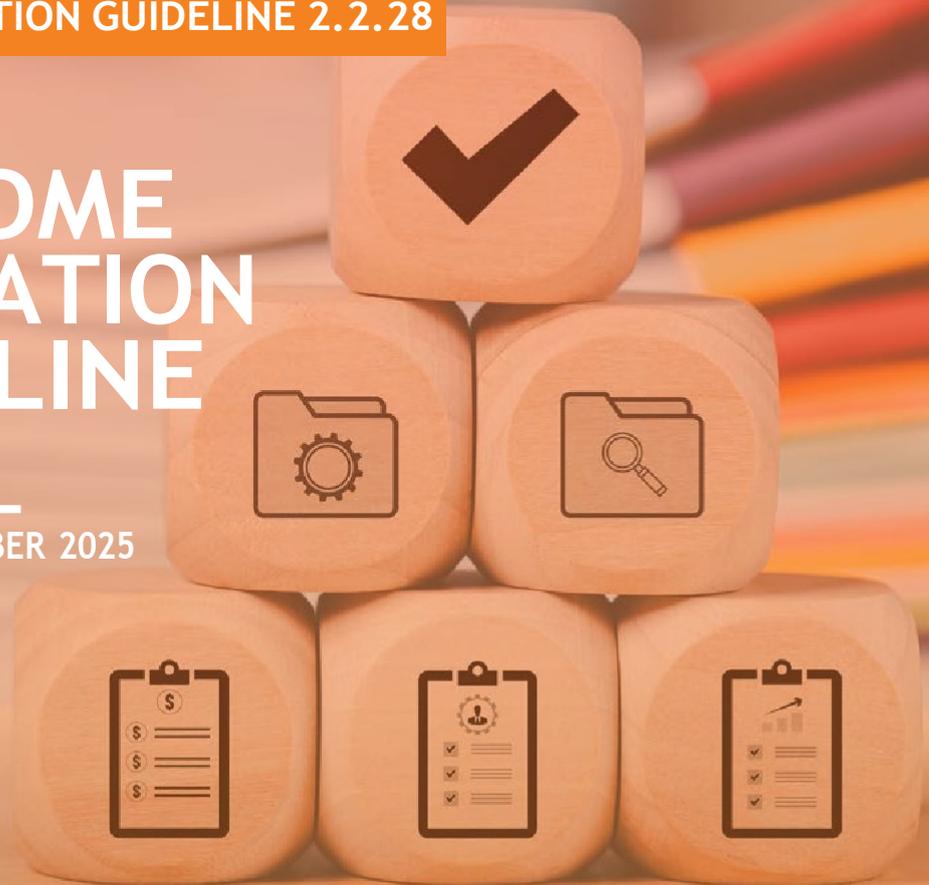the **dpme**
Department:
Planning, Monitoring and Evaluation
REPUBLIC OF SOUTH AFRICA

NDP 2030

DPME EVALUATION GUIDELINE 2.2.28

OUTCOME EVALUATION GUIDELINE

DEVELOPED: OCTOBER 2025

Addressed to	**National departments, provincial departments, municipalities and public entities**
Purpose	**To provide a practical guide on how to undertake Outcome evaluations**
Reference documents	**National Evaluation Policy Framework 2019**
Contact person	**Evaluations Unit, DPME E-mail: Evaluations@dpme.gov.za Tel: 012 312 0162**

TABLE OF CONTENTS

List of figures

List of tables

1. Introduction

Internationally, National Evaluation Policies (NEPs) and National Evaluation Systems (NESs) are closely interconnected, with systems established to operationalise the objectives set out in evaluation policies (Segone, Bamberger, & Reddy, 2015). The effectiveness of evaluation systems depends on well-defined policies that clarify the purpose, roles, and structures of the public sector evaluation function (Filgueiras & Queiroz, 2021). In the South African context, this relationship is formalised through the National Evaluation Policy Framework (NEPF), which serves as the policy framework, and the NES, which gives effect to the NEPF as part of the government-wide monitoring and evaluation system (DPME, 2019; Chirau, Blaser-Mapitsa, & Amisi, 2021). The NEPF includes various evaluation typologies, among them outcome evaluations. The Department of Planning Monitoring and Evaluation (DPME) has outlined seven distinct evaluation typology guidelines: diagnostic, outcome, design, economic, implementation, impact, and evaluation synthesis. However, an outcome evaluation guideline has not yet been formally established. This guideline has therefore been developed to offer practical guidance on planning and conducting outcome evaluations to government institutions at different spheres of government including state owned companies and state-owned enterprise.

2. Purpose and rationale for the guideline

The Outcome Evaluation Guideline has been developed to provide structured, practical, and contextually relevant guidance to government departments, public entities, and evaluation practitioners in the design and implementation of outcome evaluations within the NES. Outcome evaluations occupy a distinct position in the evaluation spectrum, focusing on whether interventions are achieving their intended results and generating meaningful benefits for the targeted beneficiaries. Unlike impact evaluations, which seek to determine causality, or process/implementation evaluations, which assess how interventions are delivered, outcome evaluations emphasise the measurement of medium-term changes in behaviour, practices, systems, or conditions that can reasonably be attributed to the intervention.

The rationale for developing this guideline is grounded in the need for greater standardisation, quality assurance, and methodological rigour in outcome evaluation practice across the public sector. While outcome evaluation has been applied in various government programmes in partnership with development partners like the United Nations Development Programme (UNDP) in Limpopo, the absence of a dedicated, comprehensive, and user-friendly guide has led to inconsistencies in approach, variable quality in evidence, and limited comparability across evaluations (Matlala, 2025). This guideline addresses these gaps by consolidating best practices, aligning with widely recognised evaluation criteria, and integrating lessons from both national and international evaluation practice.

Furthermore, the guideline responds to policy imperatives set out in the NEPF, the Medium Term Development Plan (MTDP), and sectoral planning frameworks, which collectively emphasise evidence-informed decision-making, accountability for results, and the continuous improvement of government programmes. By equipping departments and entities with a clear methodological framework, the guideline aims to enhance the credibility, usefulness, and uptake of outcome evaluations for strategic decision-making, learning, and resource allocation.

The intended users of this guideline include programme and project managers, M&E practitioners in departments and entities, evaluation specialists, and external service providers commissioned to undertake outcome evaluations. While it is primarily designed for use in the South African public sector, the principles and approaches outlined herein are adaptable to other contexts where outcome-focused evidence is required.

Finally, the guideline is intended as a companion to other DPME-issued evaluation guidance documents, such as the Impact Evaluation Guideline, Implementation/Process Evaluation Guideline, Guideline for Developing Terms of Reference for Evaluations, the Ethical Conduct Evaluation Guideline and the Evaluation Quality Assessment Framework (EQAF). It complements these resources by offering a dedicated focus on the "outcome" level of the results chain, ensuring that evaluations provide actionable insights on the extent to which government interventions are delivering the intended changes for citizens.

3. Definition of evaluation and outcome evaluation

The NEPF defined evaluation as 'the systematic collection and objective analysis of evidence on public policies, programmes, projects, functions and organisations to assess issues such as relevance, performance (effectiveness and efficiency), value for money, impact and sustainability, and recommend ways forward' (DPME, 2019). Whereas an outcome describes the effect that outputs will have on the beneficiaries in terms of changed behavior or improved performance (see fig 1).

Outcome evaluation measures the degree to which the programme is having an effect on the target population's wellbeing and/or behaviours. Outcome evaluations help determine whether or not the intended benefits of a programme are actually achieved (i.e., whether or not the programme is able to meet its intended purpose) (DPME, 2019). The timing of an outcome evaluation is when the programme has made contact with at least one person or group in the target population, however depending with the duration or life cycle of an intervention. Herein, intervention means a project, programme, policy or plan. Outcome evaluations can be conducted as:

- **Formative outcome evaluations** (often referred to as mid-term evaluations), undertaken during implementation to assess progress towards intended outcomes, identify emerging results, and provide insights for course correction (OECD/DAC, 2019; Patton, 2008).

- **Summative outcome evaluations,** conducted at or near the end of an intervention to determine the extent to which intended outcomes were achieved and to inform decisions about continuation, scaling, or replication (OECD/DAC, 2019; Patton, 2008).

Similar to other evaluation types, outcome evaluations may utilise a theory of change or logic model to help evaluators and stakeholders pinpoint short, medium and long-term changes the programme aims to achieve. After identifying these intended changes within the model, it is essential for evaluators to establish valid and reliable measures to accurately document the outcomes. Ideally, desired outcomes should be defined before the programme begins (theory of change and logical model) enabling continuous tracking throughout its duration.

In essence, this guide establishes that outcomes are an answer to the 'So what?' questions meaning 'so what difference does the intervention make'. The Kellogg Foundation defines outcomes as 'the specific changes in intervention participants behaviour, knowledge, skills, status and level of functioning' (W.K. Kellogg Foundation 2004, p.2).

Figure 1: Location of outcome evaluation on results chain
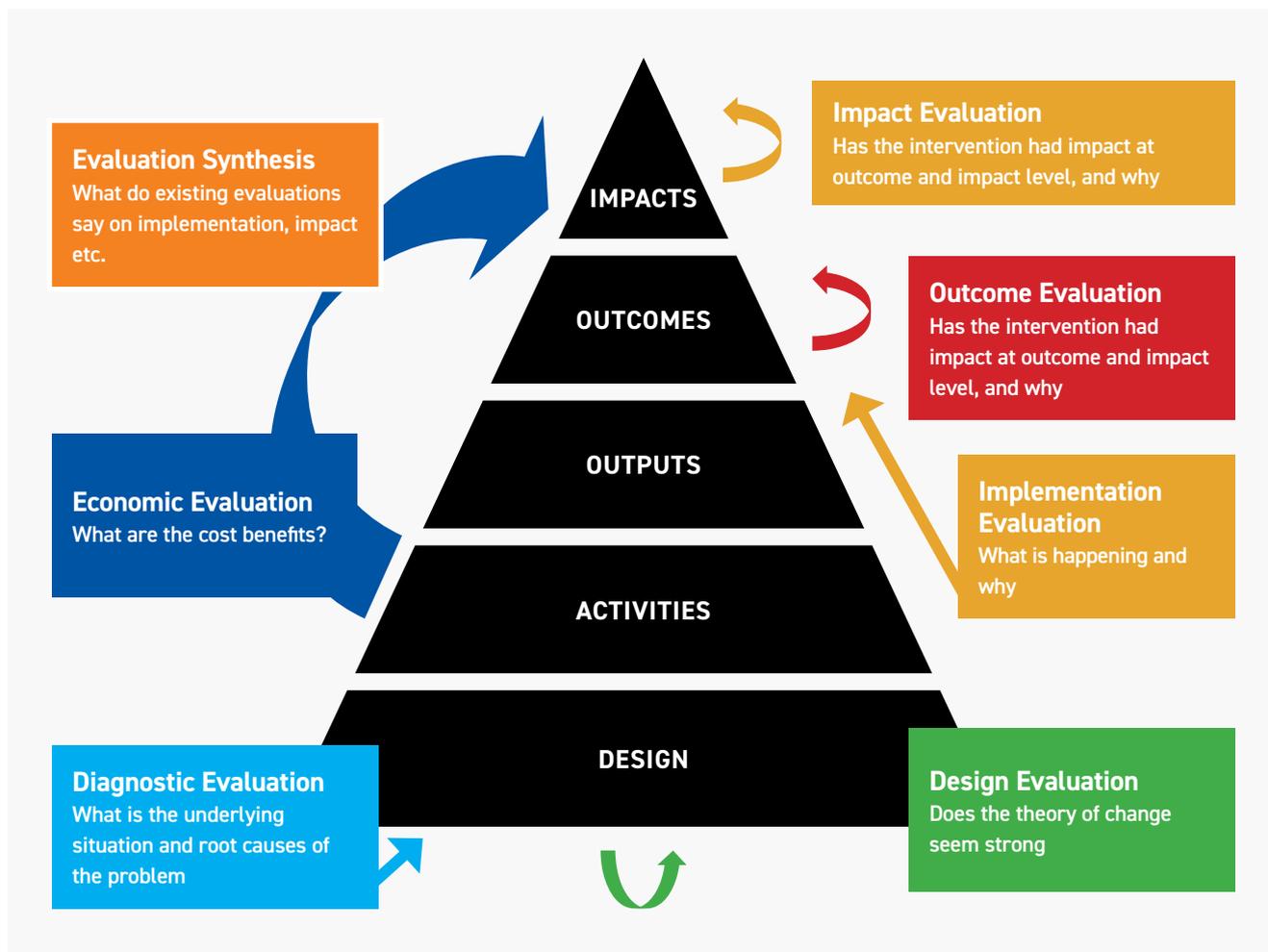
A frequent source of confusion in evaluation practice lies in the distinction between outcome evaluation and impact evaluation. While both focus on the effects of an intervention, outcome evaluations typically examine changes that are plausibly linked to the intervention within its implementation period or shortly after. They often focus on intermediate results rather than long-term societal change. Impact evaluations, by contrast, aim to establish a higher level of causal certainty about the long-term, sustained effects of an intervention, often using experimental or quasi-experimental designs to attribute observed changes directly to the intervention and rule out alternative explanations (White, 2010; Gertler et al., 2016). In practice, this distinction can blur, particularly where outcomes have long-term implications but maintaining clarity helps in selecting appropriate designs, methodologies, and timing. Ultimately, government institutions commissioning evaluations should agree upfront on a clear working definition of "outcome" and "impact" within the specific context of the intervention being assessed, ensuring alignment in expectations, terminology, and methodology.

4. Purpose of outcome evaluation and goals

Outcome evaluation systematically measures interventions results by collecting and analysing data to assess whether interventions achieve their intended goals and deliver meaningful benefits to target populations – evaluand.

The goals of outcome evaluation are value judgement of an intervention, enhancing improvement, evidence generation, strategic decision support and knowledge generation.

Figure 2: Goals of outcome evaluation

| Value Judgement | Enhancing Improvement | Evidence Generation | Strategic Decision Support | Knowledge Generation |

Goals of outcome evaluation

4.1   Value judgement of an intervention

The primary function involves measuring interventions effectiveness by systematically determining whether interventions achieve their stated goals and generate positive change for intended beneficiaries. This foundational assessment provides the groundwork for all other evaluation purposes.

4.2   Enhancing improvement

Outcome evaluation serves as a tool for programme enhancement. By revealing specific strengths and weaknesses in programme design or implementation, evaluation findings guide programme managers, implementers and stakeholders toward targeted improvements. This process informs targeted adjustments to programme design or implementation in order to achieve better results, with the specific nature and timing of changes depending on the stage of the programme lifecycle.

4.3   Evidence generation

Outcome evaluation is an accountability mechanism by generating credible, evidence-based documentation of programme performance. This evidence serves multiple audiences, including funders who need justification for continued investment, policymakers requiring proof of concept for scaling initiatives and intervention participants who deserve transparency about the services they receive.

4.4   Strategic decision support

Furthermore, outcome evaluation provides crucial evidence for strategic decision-making across the intervention lifecycle. When leaders face choices about programme design modifications, implementation strategies, or resource allocation, evaluation data helps identify which approaches deliver the greatest return on investment. This evidence-based decision- making reduces guesswork and increases the likelihood of sustained positive outcomes.

4.5   Knowledge generation

Outcome evaluation contributes to the broader knowledge ecosystem by generating insights about effective practices and unsuccessful approaches. These learnings extend beyond individual interventions to inform field-wide understanding of what works, what does not work, for whom and under what different conditions. This learning and knowledge sharing function helps by sharing both successes and failures, preventing others from repeating costly mistakes while accelerating the adoption of proven strategies that deliver outcomes.

5.   Relationship between results monitoring and outcome evaluation

Results monitoring at outcome level and ultimately outcome evaluation offer complementary forms of evidence. Results monitoring is carried out as a continuous process in which the outcomes are regularly tracked. Ongoing data is provided on what is being implemented by the programme and what short, medium and long term results are being achieved. This process is typically descriptive in nature, offering early signs of progress through tracking outcome indicators or emerging issues throughout the intervention's lifecycle (Kusek & Rist, 2004). This feeds directly to outcome evaluation which is conducted periodically and in greater depth.

A robust outcome evaluation relies heavily on the quality, consistency, and completeness of monitoring data. It involves a systematic and objective assessment of the extent to which a programme's intended outcomes as theorised through the theory of change and operationalised through logical framework namely the changes or benefits for the target population have been realised. Emphasis is placed on understanding results over a longer time frame, with attention given to questions of causality, such as why certain outcomes were or were not achieved. Establishing baseline data, that is, systematic measurements taken before or at the start of an intervention is critical. Baselines provide the reference point against which observed changes can be compared, allowing evaluators to assess progress and determine the magnitude of change attributable to the intervention (OECD/DAC, 2019).

Figure 3: Implementation and results monitoring

Input — The resources that contribute to the production and delivery of outputs

Activity — The processes or actions that use a range of inputs to produce the desired outputs and ultimately outcomes

Output — The products or goods and services produced from activities

Outcome — The short, medium and long term results for beneficiaries that are the consequences of achieving specific outputs

Impact — The development results of achieving specific outcomes

Implementation Monitoring
"Are activities being implemented as planned?"

Results Monitoring
"Are activities leading to the desired results?"

In practice, the relationship between results monitoring and outcome evaluation is iterative. Ongoing results monitoring provides a continuous flow of data on outputs and emerging outcomes, which informs the design of outcome evaluations, highlights priority focus areas and identifies any data gaps. Conversely, findings from outcome evaluations can enhance the quality and relevance of monitoring systems by refining indicators, strengthening data collection tools, and improving reporting processes. As shown in Figure 2, this integration ensures that monitoring data serves a dual purpose, meeting immediate reporting requirements and building the foundation for credible, timely, and cost-efficient evaluations that contribute to learning, accountability, and decision-making (Görgens & Kusek, 2010).

To strengthen this relationship in practice, government institutions should ensure that their monitoring systems are structured in a way that feeds directly into outcome evaluation. A short readiness checklist is provided below to guide departments in assessing whether monitoring data is robust enough to support credible outcome evaluations.

Table 1: Monitoring system readiness checklist for outcome evaluation

Monitoring element	Guiding questions	Why it matters for outcome evaluation
Baseline data	Is there a documented baseline with date, method, and data sources recorded?	Provides the benchmark against which changes are measured.
Data collection frequency	Are data collected at intervals aligned with programme milestones and reporting?	Ensures outcomes are tracked consistently and comparably.
Variable definitions	Are key indicators and variables clearly defined and standardised?	Avoids ambiguity and enhances comparability across periods.
Data quality indicators	Are there checks for accuracy, completeness, timeliness, and reliability?	Strengthens confidence in the evidence base for evaluation.
Metadata availability	Are data sources, methods, and assumptions documented and accessible?	Enables evaluators to assess credibility and replication.
Data storage & access	Is data systematically stored and accessible to evaluators and managers?	Ensures data continuity and supports independent verification.

6. Ideal conditions for conducting outcome evaluations

Outcome evaluations are most effective when conducted under specific conditions that maximise the likelihood of generating credible, useful, and actionable findings. Prematurely commissioning an outcome evaluation, before a programme has reached sufficient maturity, produced measurable outputs, or established clear causal pathways, can result in inconclusive or misleading results. As evaluability assessment literature highlights, evaluating too early, before outcomes can reasonably be observed, risks undermining both the trustworthiness and usefulness of the findings (Wholey, 2010). Establishing ideal conditions ensures that evaluation resources are used efficiently and that the evaluation contributes meaningfully to learning, accountability, and policy improvement (Centers for Disease Control and Prevention [CDC], 2024; Wholey, 2010; University of Kansas, 2024).

International good practice, including guidance from the OECD DAC (2019), UNEG (2016), and the NEPF (2025), emphasises the importance of assessing readiness for

outcome evaluation before commissioning such studies. Readiness involves ensuring that the programme has been implemented for long enough to observe meaningful changes, that robust baseline or comparable data exist, and that the theory of change is clearly articulated and tested.

Outcome evaluations should only proceed when key conditions are sufficiently met to ensure the results will be credible, useful, and actionable. A readiness assessment provides a structured way to judge whether these conditions are in place. Using a 1–5 rating scale allows evaluators to capture the degree to which each condition is met, identify areas that require strengthening, and decide whether to proceed, postpone, or redesign the evaluation approach. (See Annexure E for rating scale description.)

Table 2: Readiness assessment for outcome evaluations

Condition	Description	Key questions	Implication if condition is not met	Readiness Score (1 - 5)
Programme maturity	The intervention has been implemented for a sufficient period to allow outcomes to emerge.	Has the programme been operational long enough to influence outcomes?	Delaying the evaluation until measurable results are expected.	
Clear theory of change	A logical framework or results chain exists, linking activities, outputs, and intended outcomes.	Is there a well documented theory of change? Have assumptions been tested?	Requires ToC development or refinement before evaluation.	
Stable implementation	The programme model and delivery mechanisms are consistent and stable.	Have major design or delivery changes occurred recently?	Postpone evaluation until the programme stabilises.	
Availability of baseline/ comparat or data	Baseline data or a credible counterfactual exists for assessing change.	Do we have reliable data from before the intervention?	Conduct a baseline study or reconstruct baseline before proceeding.	
Measurable outcomes	Outcomes are specific, observable, and measurable with available tools.	Are there valid indicators and data collection tools for each outcome?	Strengthen M&E systems to measure outcomes.	
Stakeholder buy-in	Key stakeholders support the evaluation and are committed to using findings.	Is there agreement among programme managers and policymakers to act on findings?	Engage stakeholders early to secure buy-in.	
Resourcing	Adequate budget, technical expertise, and time are available.	Are resources in place for a high- quality evaluation?	Secure funding and expertise before proceeding.	

Before initiating an outcome evaluation, commissioners should undertake a readiness assessment using the above table as a checklist. If any conditions are not met, remedial actions, such as improving data systems, refining the theory of change, or ensuring programme stability, should be prioritised before proceeding. Outcome evaluations should also align with broader evaluation planning cycles to ensure that findings are available at decision- making moments (e.g., before budget reviews, policy shifts, or scaling decisions).

7. Theory of change and identifying outcomes

The Theory of Change (ToC) is a foundational framework in outcome evaluations that explains how and why a particular programme or intervention is expected to bring about desired changes within a specific context. It works by clearly identifying the long-term goals and then mapping backward to outline the necessary short and medium outcomes or preconditions that must be achieved for those goals to be realised. This mapping includes detailing the causal pathways showing how activities and interventions lead to these outcomes, making explicit the assumptions and conditions that influence success. By doing so, the ToC provides a logical and comprehensive narrative that links what the intervention does to the changes it aims to produce, improving planning, implementation and evaluation clarity. Together, the Theory of Change and the logical model enhance understanding of the programme's theory of action by structuring the theory into measurable parts, facilitating monitoring and evaluation. This ensures that evaluations measure not only whether outcomes are achieved but also whether the programme's activities are effectively driving those outcomes, offering a basis for decision- making, learning and adaptation.

When government institutions commission outcome evaluations, the focus begins and ends with the outcomes, the benefits and changes resulting from the investment in interventions. In outcome planning, rather than starting with inputs and determining their deployment, the process should begin with clearly defined outcomes.

A practical and resource-conscious way of developing a ToC is through backward mapping. This approach begins by first clarifying the ultimate long-term goal of the programme. From this end point, planners work backwards to identify the medium- and short-term outcomes that must occur for the goal to be realised. Once these preconditions are identified, the corresponding activities and outputs are mapped forward again, ensuring logical coherence between what the programme does and the changes it seeks to achieve. This method is especially useful for government departments with limited time and resources, as it prioritises clarity on outcomes before elaborating activities, avoiding unnecessary detail at the outset (Görgens & Kusek, 2010).

A simplified five-step backward mapping protocol can be applied in ToC workshops or planning processes:
1. Define the long-term change or societal goal.
2. Identify intermediate and short-term outcomes that must occur to achieve the goal.
3. Map causal pathways showing how interventions contribute to these outcomes.
4. Test assumptions and identify external factors through rapid stakeholder consultations.
5. Refine and document the ToC in a visual format aligned with monitoring indicators.

Figure 4: Outcome planning models

Outcome-oriented planning (work back from results)

| Outcomes | Activities | Outputs | Inputs of Resources |

And what organisations normally do:

| Inputs of resources | Activities | Outputs | (and possibly outcomes) |

Source: *Hoggarth and Comfort, (2010)*

The first question is 'what is the outcome that you want to achieve?' If the desired change is to reduce teenage pregnancy in South African communities, the next question should be 'what intervention(s) would best achieve this and is there any relevant evidence that might help to answer that question?'.

Government institutions at all levels are encouraged to develop their outcome statements early, during the design and planning phases of their interventions, to facilitate assessment of achievements later on. These outcomes should be documented in the theory of change and logical frameworks. A well-defined outcome is usually a statement that encapsulate the following:

- what is being addressed? (e.g., the aspect of the client's progress or the particular conditions in a local area)
- who will benefit (e.g.an individual or a group in particular need or a whole community?)
- the direction of change (will the measure increase or reduce?)
- by how much it is predicated to change and over what time period? (Dartington 2006, p2)

One practical approach that can strengthen the identification of outcomes during ToC development is Outcome Mapping (Earl, Carden, & Smutylo, 2001). Outcome Mapping focuses on behavioural change among key stakeholders, referred to as boundary partners, rather than only on service

delivery or outputs. It defines progress markers, which are observable changes in actions, relationships, policies, or practices that indicate movement toward the desired outcome. These markers can be set at graduated levels (expect to see, like to see, love to see), enabling tracking of incremental progress throughout implementation. By using Outcome Mapping principles within the ToC process, government institutions can ensure that outcome statements are specific, measurable, realistic, and linked to clearly identified actors. This not only makes subsequent monitoring more meaningful but also provides a robust basis for outcome evaluations later in the programme lifecycle.

This guideline recognises the difficulty of clearly defining desired outcomes at the outset. However, it is important to avoid planning for failure or setting unrealistic goals. The example below is provided to assist programme managers and implementers in crafting their outcomes, which evaluators will refer to during the evaluation phase. Table 3 below depicts an example of a project 'outcome' that can help project managers and implementers to craft outcome statements.

Table 3: Example of project outcome

A hypothetical project aims to improve children's health and well-being in South African schools. It does this through a whole-school approach to health eating, exercise, the dangers of smoking and so on. Below is an example of designing a precise outcome around the issue of diet in order to answer the key question of definition of outcome. A project like this will of course have several other outcomes to work on but, ideally each one should be clearly defined.	
What is being addressed?	The need for children to eat, the government recommended intake of five portions of fruit or vegetables per person per day to reduce the risk of some cancers, heart disease and other chronic conditions.
Who will benefit?	Boys and girls 8-11 at six schools in one Limpopo
Will the measure increase or reduce (the direction of change)	The proportion of children with sufficient intake of fruit and vegetables will increase.
By how much is it predicted to change	The proportion will increase by 30 per cent from the baseline established with a survey of the children in April 2025.
Over what time period	By March 2030.
Outcome: By March 2030 an increase of 30 per cent will be achieved against the baseline position at April 2025 in proportion of children aged 8-11 at the six schools in the scheme eating the recommended five portions of fruit and vegetables per day.	

Source: Hoggarth and Comfort, (2010) revised

While the ToC provides the overarching narrative of expected change, the Logical Framework (logframe) translates this into an operational planning and monitoring tool. The logframe sets out clear objectives (impact, outcomes, outputs, activities), each with associated SMART indicators, means of verification, and assumptions. This structured matrix ensures that outcomes are not only well defined but also measurable, allowing continuous monitoring and later outcome evaluation.

Using the dietary behaviour example, the outcome ("By 2030 an increase of 30% in children's daily fruit/ vegetable intake...") is operationalised in the logframe through indicators such as "% of children consuming five portions daily" or "dietary diversity score ⩾5 of 10." These indicators, with baselines, targets, and data sources, form the foundation for robust monitoring systems and credible outcome evaluation. In this way, the logframe acts as the bridge between the ToC narrative and the evaluation evidence base (Görgens & Kusek, 2010; W.K. Kellogg Foundation, 2004).

8. Who undertakes outcome evaluation

An outcome evaluation should be conducted for all government programmes that are included in evaluation plans at departmental, provincial, municipal, or state-owned company and enterprise level. The responsibility for undertaking the evaluation may rest with internal staff, external service providers, or a combination of both, depending on the context, resources, and the need for independence.

Outcome evaluations involve the collection and analysis of data using diverse methodological approaches. It is advisable, where possible, to engage a specialised service provider with demonstrated expertise in research and evaluation, particularly in outcome-level analysis. However, where resources are constrained, internal evaluation can be undertaken by trained staff from the commissioning government institution.

- **Internal evaluators** are staff members of the government institution commissioning the evaluation. They evaluate their own intervention and remain accountable to their organisation. This approach benefits from their familiarity with the intervention's history, context, and stakeholders, and increases the likelihood that findings will be used, as staff remain invested in the results. However, it requires measures to mitigate potential bias and ensure objectivity, such as external peer review.
- **External evaluators** are independent professionals or firms, either local or international, who are not employed by the commissioning institution. They are often engaged for transparency, impartiality, and credibility, especially in outcome and impact evaluations. External evaluators typically bring specialist skills, comparative insights from other contexts, and knowledge of best practices.

- **Hybrid evaluations** combine both internal and external evaluators, aiming to leverage the strengths of each. Internal staff contribute contextual understanding and institutional memory, while external evaluators bring methodological expertise and impartiality.

Regardless of whether internal, external, or hybrid models are used, clear governance and conflict-of-interest safeguards are essential. Outcome evaluations should be overseen by an Evaluation Steering Committee (as per the NEPF) comprising senior programme managers, M&E staff, and where appropriate, representatives of civil society or beneficiaries to ensure inclusivity and accountability. The Steering Committee should have clear role descriptions covering oversight, methodological approval, ethical compliance, and the use of findings. For external providers, procurement processes should explicitly address independence and conflict-of-interest risks, requiring declarations of interest and transparent selection criteria.

Minimum qualifications and competencies for lead evaluators should include:
- Strong methodological expertise in outcome-level evaluation design (quantitative, qualitative, or mixed methods).
- Sectoral knowledge relevant to the intervention being assessed.
- Demonstrated adherence to ethical standards (DPME Evaluation Ethics Guideline; UNEG Ethical Guidelines).
- Experience in managing stakeholder engagement and facilitating participatory processes.

Importantly, both external and hybrid evaluations should explicitly include a capacity development component, where external evaluators mentor and train internal staff to strengthen institutional evaluation capability, aligning with NEPF's goal of building a professional, skilled evaluation system within government (DPME, 2019).

9. Evaluation methodology, questions and indicators

When conducting an outcome evaluation, internal and external evaluators must return to the original rationale for the evaluation and the key questions to be answered. This guideline argues that all government institutions require credible evidence of the benefits of their work to inform decision-making, improve performance, and strengthen accountability. All the outcomes identified in the theory of change and logical framework should be tracked for evidence of achievement. The evaluation process must therefore seek evidence on whether or not these outcomes have been realised.

Crucially, outcome evaluation should not be treated as an afterthought. Embedding outcome evaluation questions at the inception stage, within an intervention's M&E plan or framework, ensures that the necessary baselines, monitoring systems, and resources are in place to generate robust, actionable findings. This approach aligns with international best practice, which emphasises planning evaluations early in the programme cycle to enable credible causal analysis and maximise utilisation of findings (OECD/DAC, 2019; UNEG, 2016).

Once the evaluation questions and sub-questions are framed, evaluators must then determine what information could be gathered to provide credible evidence to answer them. This guideline distinguishes between evidence and data: evidence refers to the facts or circumstances supporting or refuting a proposition (Tulloch, 1996), while data are the specific facts collected as a basis for inference or reckoning. Data on outcomes may be quantitative or qualitative and should be triangulated to strengthen validity. Quantitative data captures measurable, countable results, such as percentages, proportions, and statistical trends, while qualitative data explores meanings, perceptions, and contextual explanations, often through narrative accounts. Both forms are essential for a comprehensive understanding of outcomes. In addition, strong outcome evaluations rely heavily on monitoring data collected

throughout implementation. Baseline data, whether drawn from primary data collection, programme records, or secondary sources, provide the point of comparison for assessing change over time. Without a reliable baseline, the credibility of outcome findings is diminished, as there is no benchmark against which to measure progress (Görgens & Kusek, 2010).

9.1 Sampling and sample-size heuristics

A critical element of outcome evaluation is ensuring that sampling strategies are appropriate for the intervention and resources available. Under-sampling undermines credibility, while over-sampling wastes resources. A five-step protocol can guide evaluators:

1. Define the population of interest clearly (e.g., programme beneficiaries, schools, clinics, municipalities).
2. Select the sampling frame (administrative lists, registries, or constructed frames).
3. Choose the sampling strategies that are feasible and appropriate to the evaluation design (random, stratified, cluster, purposive, mixed).
4. Estimate the required sample size based on desired confidence level, margin of error, and design effect.
5. Adjust for non-response and attrition by inflating sample size by 10–30% as appropriate.

Table 4: Sampling heuristics for outcome evaluations

Design type	Typical sample size guidance	Margin of error/ retention	When to use
Cross- sectional survey	400–600 respondents total; ⩾30 per subgroup	±5% at 95% confidence (~385 respondents for large populations)	Suitable for large, diverse populations; enables subgroup analysis if stratified.
Cohort/ longitudinal study	200–400 respondents, with ⩾70% of baseline sample retained	±7–10% depending on retention	Tracks changes in the same individuals over time; useful for assessing behavioural and practice change.
Cluster sampling (e.g., schools, clinics, wards)	20–30 clusters × 15–25 respondents each; inflate sample by 1.5–2× for design effect	±7–10% (varies with intra-cluster correlation)	Cost-efficient for dispersed populations; requires adjustment for clustering in analysis.
Key informant interviews	5–30 participants (purposive)	Not statistical	Best for insights from policymakers, managers, or experts.
Focus group discussions	6–10 participants per group, 6–10 groups	Not statistical	Effective for exploring perceptions, behaviours, and unintended outcomes.

9.2 Indicators

Selecting the indicators is a crucial part of evaluating outcomes of an intervention. Indicators are proxy measures for something you cannot measure directly, and you find an indicator that tells you about it and gives you some reasonable measurement. It is important for the evaluators to revisit the results chains/logical model to understand the following:

- What are the intended outcomes for the intervention to be evaluated?
- What indicators could tell us about those outcomes?
- What information are we therefore choosing to collect?
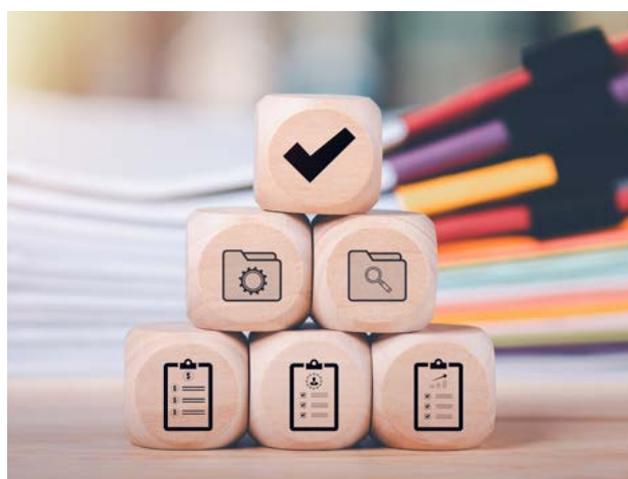
Table 5: Sample indicator template – School Health and Well-being Programme example

Indicator	Definition	Numerator	Denominator	Disaggregation	Source	Frequency	Baseline	Target
% of children (8– 11 consuming at least 5 portions of fruit/ vegetables daily	Proportion of surveyed children aged 8–11 who report eating ⩾5 portions of fruit/veg in the past 24 hrs.	# of children eating ⩾5 portions daily	Total # of children surveyed (8–11 yrs)	Sex; province; school	24-hr recall survey	Annual	25% (Apr 2025)	55% (Mar 2030)

It is the evaluators responsibility to picture the situation (result–outcome) to be achieved but ask the following:

- What would it look like if we achieved it?
- How will we know we have arrived?

Table 6 below provides generic outcome evaluation questions that can be adapted to different contexts. These questions should ideally be formulated at the start of the intervention and embedded in the evaluation plan to ensure data availability and readiness for future evaluation.

Table 6: Outcome evaluation questions

Purpose	Key Outcome Question	Sub-Outcome Evaluation Questions	Methodology	Common Evaluation Methods and Approaches
Understanding differences between the situation before and after the programme intervention	What changes occurred as a result of the programme?	• What measurable improvements are observed among participants compared to the baseline? • Are there unintended outcomes (positive or negative) emerging during or after implementation?	• Mixed-methods (combining quantitative and qualitative data collection and analysis) • Comparative analysis using baseline and follow-up data • Theory of Change review	• Pre- and post-intervention surveys or assessments (using baseline data) • Focus group discussions • In-depth interviews • Review of programme monitoring records
Intended and Unintended Outcomes	Did the programme produce or contribute to the intended outcomes in the short, medium, and long term?	• Which outcome targets were met or exceeded? • Which outcomes fell short of expectations? • What unexpected effects (positive or negative) were identified, and what are their possible causes?	• Integration of quantitative data from monitoring systems and primary surveys with qualitative insights from stakeholders • Document review of programme records and reports	• Statistical analysis of outcome indicators • Thematic analysis of qualitative evidence • Review of programme records/data

Purpose	Key Outcome Question	Sub-Outcome Evaluation Questions	Methodology	Common Evaluation Methods and Approaches
Related elements contributing to the achievement of outcomes	What factors influenced the achievement of outcomes?	• What internal and external factors contributed to or hindered progress towards outcomes?	• Qualitative inquiry (focus on causal and contextual factors) • Contribution analysis • Document and secondary data review	• Key informant interviews • Thematic analysis • Case studies • Review of monitoring and administrative data
Sustainability	Is there a sustained impact of the programme outcomes over time?	• Are the observed outcomes maintained after programme completion? • What follow-up evidence exists on the durability of changes?	• Longitudinal tracking (repeated measures over time) • Follow-up monitoring using primary and secondary data sources	• Follow-up surveys • Repeat interviews over intervals • Trend analysis using programme records and monitoring data

9.3   Outcome harvesting

In case where ToC has not been explicit on the desired outcomes. This guideline encourages the government institutions to use outcome harvesting. Outcome harvesting is an evaluation approach that focuses on identifying and documenting the actual changes that result from development intervention. Rather than starting with predetermined objectives, this methodology works in reverse, first capturing evidence of what changes have occurred, then tracing back to determine how the intervention contributed to those outcomes. The approach draws from Outcome Mapping principles, and the two methods are frequently used together to provide a comprehensive understanding of interventions outcome[2]. A fundamental distinction between these approaches lies in their timing and application. Outcome Mapping serves a dual purpose it can be implemented both during the initial planning phase and throughout (or following) intervention execution. Outcome harvesting, however, is specifically designed for use during or after implementation, rather than as a planning tool.

The outcome harvesting approach employs specific terminology that is essential to understanding the methodology. These key terms are defined below and are adapted from Wilson-Grau and Britt (2013):

• A change agent is an individual or organisation that influences an outcome. In outcome harvesting the change agent is often an organisation running a project or programme.

• A social actor is an individual, group, community, organisation or institution that changes because of a change agent's intervention.

• The harvest user is the stakeholder who needs the findings of an outcome harvest to make decisions or take action. This may include one or more people within the change agent organisation, or third parties such as a donor.

• The harvester is the person or people responsible for managing the outcome harvest. The harvester is often an internal or external evaluator. The harvester leads the outcome harvesting process and facilitates and supports participation within the process.

2 https://www.intrac.org/app/uploads/2024/12/Outcome-harvesting.pdf

9.3.1 When to use outcome harvesting

According to see Wilson-Grau (2015), there are three preconditions that should be present for outcome harvesting to take place.

- **Focus on outcomes over activities:** Outcome harvesting is most suitable when the primary interest lies in understanding results and impact rather than tracking implementation activities or immediate outputs. This methodology is specifically designed to explore what changes occurred and the underlying reasons, helping to illuminate change processes. It is not intended for assessing whether planned activities were executed as originally designed.

- **Complex and dynamic contexts:** This approach excels in complicated environments where cause-and-effect relationships are unclear or where multiple stakeholders influence outcomes. In such settings, both the desired changes and the strategies to achieve them are often unpredictable, requiring continuous adaptation of plans. Outcome harvesting proves particularly valuable in areas such as policy advocacy, community mobilisation, capacity building, empowerment initiatives, and network development.

- **Learning oriented evaluation:** Outcome harvesting is most appropriate when the evaluation's primary purpose is generating insights to enhance future performance rather than simply measuring success.

Figure 5: Steps in outcome harvesting
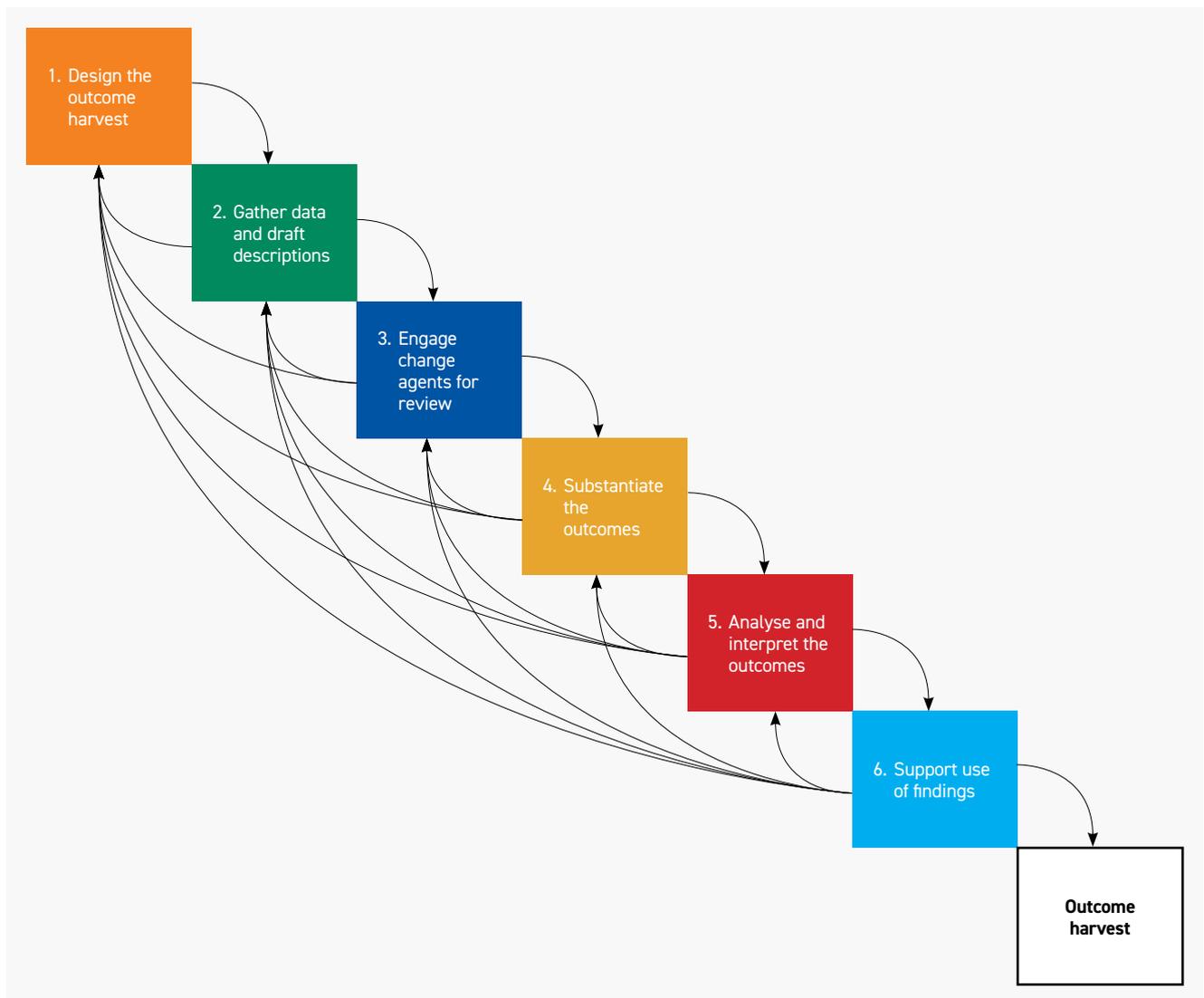
Step 1: Design the outcome harvest

Harvesters and harvest users collaborate to develop guiding questions, prioritising the specific needs of harvest users. Based on these questions, they determine what information to collect and from whom, plus define what details to include in outcome descriptions. Essential information must cover:

The Outcome

- Who did the change agent influence?
- What specific changes occured? When and where?
- What observable differences exist in the social actor's behaviour?

Contribution

- How did the change agent contribute to this change?
- What specific actions influenced the outcome?
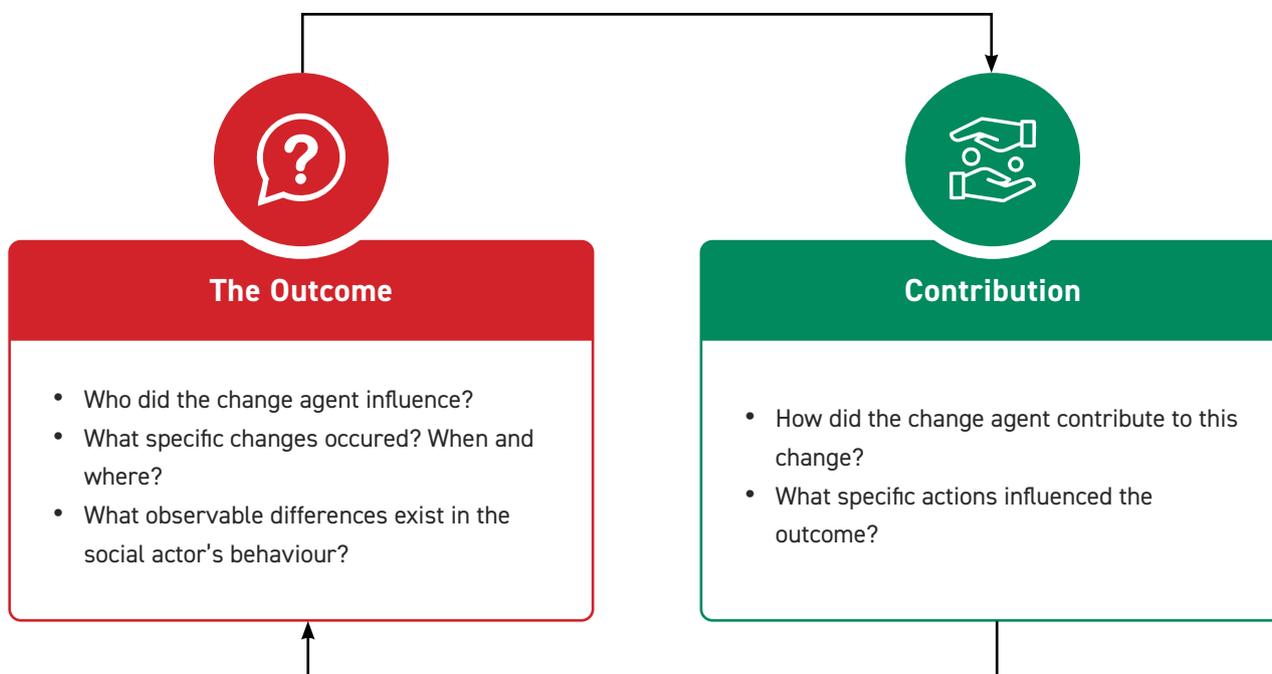
At this stage, they may also establish how information will be categorised for later analysis and interpretation.

Step Two: Gather data and draft descriptions

This step has two components. First is review existing documentation (reports, evaluations, press releases) for evidence of potential outcomes. Sometimes primary data is collected from various sources, including social actors themselves. Part Two is to draft outcome descriptions incorporating information agreed upon in Step One. Descriptions vary in length from single sentences to multiple pages and may include context, others' contributions, different perspectives, or other relevant details beyond the core change and contribution elements.

Step Three: Engage change agents for review

Change agents review draft outcome descriptions through surveys, questionnaires, interviews, or workshop sessions. During this engagement, additional outcomes may be identified and documented while harvesters request supplementary information when needed. Change agents may consult colleagues or external contacts for further details, which sometimes includes interviewing direct beneficiaries or policy targets.

The harvester's role is to rigorously examine each outcome for specificity and coherence, including verifying plausible connections between change agent actions and outcomes. The result is a revised set of outcome descriptions.

Step Four: Substantiate the outcomes

The harvester and harvest users review outcome descriptions and select a sample for verification to enhance accuracy and credibility. Verification involves interviewing individuals who are independent of the change agent but knowledgeable about the outcomes and interventions. These interviews often provide deeper insights into the outcomes and the change agent's contribution.

Step Five: Analyse and interpret the outcomes

The harvester categorises all outcomes using classifications from step one, often as a participatory exercise with change agents. For limited outcomes, classification can be done manually, while larger harvests may require a database for multidimensional criteria.

Once categorised, the harvester interprets the information to answer the harvesting questions. Outcome harvesting doesn't prescribe specific analytical methods standard qualitative analysis techniques can be applied.

Step Six: Support use of findings

The harvester discusses analysis and interpretation with harvest users, potentially suggesting specific actions based on findings. However, outcome harvesting analysis is just one factor in decision-making political, legal, financial, or ethical considerations often influence final decisions.

9.3.2 Triangulation requirements for harvested outcomes

Triangulation is a critical step in strengthening the credibility and robustness of harvested outcomes. Since outcome harvesting often documents changes retrospectively and traces them back to contributions, the risk of relying on single or biased data sources is high. Without triangulation, findings may lack credibility, reduce confidence in their validity, and weaken their utility for decision-making. By drawing on multiple, independent sources of evidence, evaluators can cross-check claims, reduce subjectivity, and build a more reliable picture of whether and how change occurred.

In evaluation practice, triangulation typically involves combining data from different sources (e.g., monitoring systems, stakeholder testimonies, policy documents), applying multiple methods (e.g., interviews, surveys, document reviews), or validating findings through independent substantiators. The UNEG Ethical Guidelines for Evaluation (2020) and Wilson-Grau & Britt (2013) emphasise that the credibility of outcome harvesting depends not only on systematic documentation but also on rigorous verification. This ensures that harvested outcomes are not just anecdotal, but supported by a reasonable standard of evidence, particularly where evaluations inform high-stakes decisions.

To operationalise this, the guideline proposes minimum triangulation standards differentiated by the importance of outcomes. These standards balance rigor with feasibility, ensuring that commissioners and evaluators can prioritise verification efforts while still producing credible results.

Table 7: Minimum triangulation standards for harvested outcomes

Outcome Importance	Minimum Evidence Requirements	When to use
High-importance outcomes	At least two independent evidence sources (e.g., monitoring data, reports, stakeholder testimonies) and one independent substantiator who can verify the change	Legislative reform attributed to advocacy efforts
Medium-importance outcomes	At least two sources of evidence (may include programme records and primary accounts from affected stakeholders)	Improved service Uptake among target groups
Lower-importance outcomes	At least one credible source of evidence, but still reviewed by evaluators for plausibility	Localised awareness or training outcomes

To further support evaluators, harvested outcomes should be consistently documented using a structured description template. This template ensures key information, such as who changed, what changed, when it occurred, and how it was substantiated, is recorded in a transparent and standardised manner.

Table 8: Template for documenting harvested outcomes

Field	Guidance
Who changed	Identify the social actor(s) – individual, group, organisation, or institution.
What changed	Specify the observable change in behaviour, practice, policy, capacity, or relationships.
When	State the timing or period when the change occurred.
Evidence	Provide the sources used to substantiate the outcome (documents, interviews, monitoring data).
Contribution analysis	Explain how the intervention plausibly contributed to the change, noting other influencing factors.
Substantiation	Record details of independent verification (e.g., name/role of substantiator, method used).

9.4 Selecting evaluation approaches for outcome evaluation

Selecting the right evaluation approach is a critical design decision that influences the relevance, credibility, and utility of findings. Outcome evaluations in government should be designed to both respond to policy and planning cycles and meet the diverse needs of stakeholders.

The NEPF highlights three approaches as especially relevant in the South African context:

- **Rapid Evaluation –** to generate timely evidence for urgent decision-making;
- **Collaborative Learning through Sector Reviews –** to foster shared understanding and joint problem-solving;
- **Transversal Evaluations –** to assess results across multiple institutions or sectors.

However, aligning with international best practice (OECD/DAC, 2019; UNDP, 2019), outcome evaluations can also draw on a broader set of approaches to ensure inclusivity, adaptability, and context-responsiveness. The choice of approach should be determined early in the evaluation planning phase (as emphasised in the NEPF's requirement for evaluation plans) and embedded within the Terms of Reference to ensure methodological clarity from the outset. Table 8 below presents an expanded set of evaluation approaches relevant to outcome evaluation in government. In addition to describing their key features and the typical point in the evaluation cycle when they are most applicable, the table also provides guidance on their relative complexity, expected time to deliver, and the likely strength of evidence they can generate. This enables commissioners and practitioners to make more informed choices about which approach is most suitable given the purpose, resources, and decision-making timelines of their evaluations.

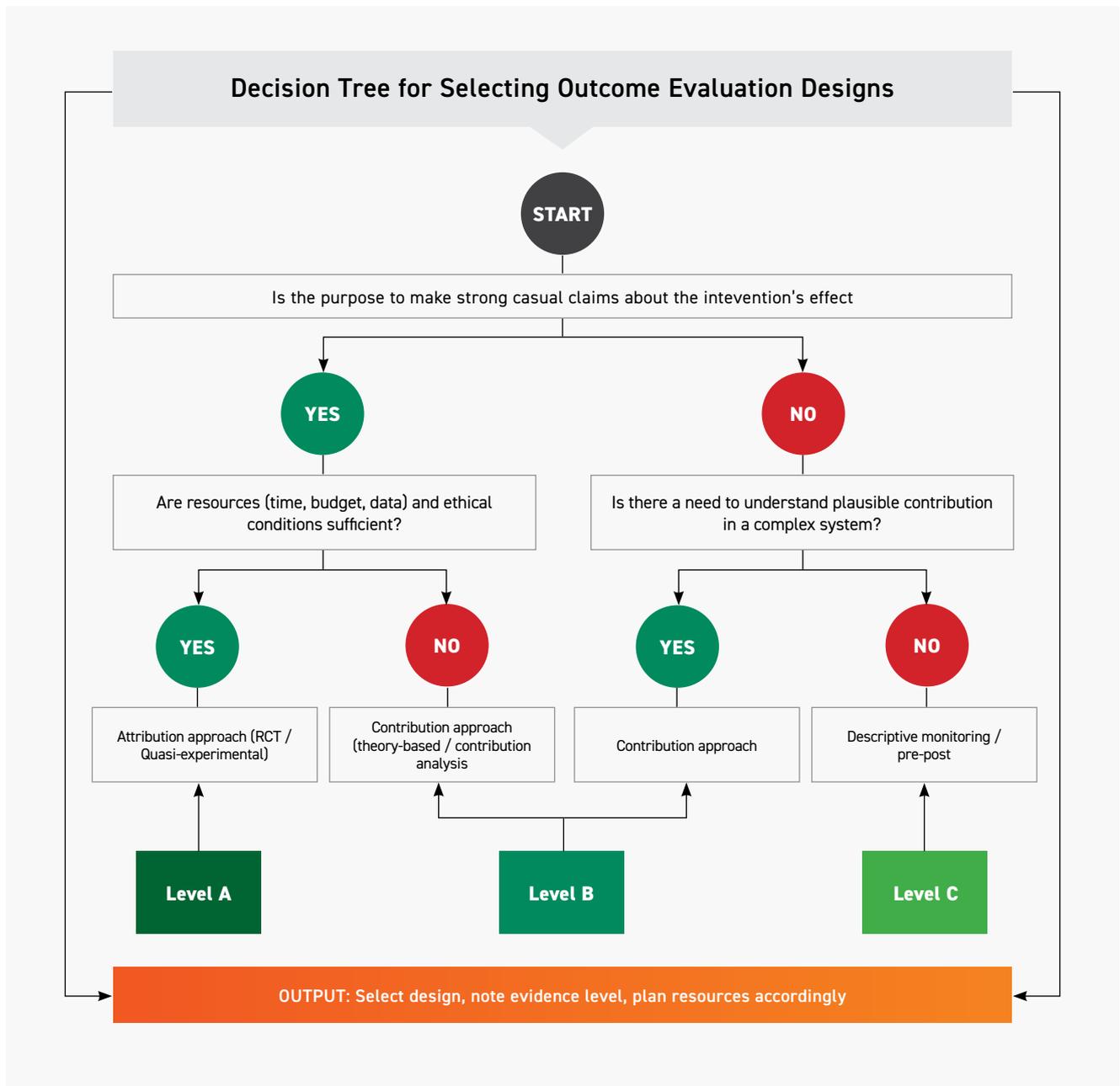## Table 9: Evaluation approaches and selection considerations

Category	Approach	Key Issues	Timing	Typical Complexity	Time to Deliver	Evidence Strength
Use Orientation	Utilisation-Focused Evaluation	Centres on the needs of intended users, ensuring they are identified and engaged from the outset. Maximises the likelihood that findings will be applied to policy and programme improvement.	From start to end of an intervention	Medium	6–12 months	Medium
	Participatory Evaluation	Involves intended beneficiaries directly in the design, data collection, and interpretation of results, fostering ownership and relevance.	From start to end of an evaluation	High	9–18 months	Medium
	Empowerment Evaluation	Builds capacity of stakeholders to evaluate their own performance, using the evaluation process as a tool for empowerment and self-determination.	From early stages and throughout implementation	Medium	6–12 months	Medium
Equity & Rights Orientation	Gender-Responsive Evaluation	Assesses how gender and power relations change due to the intervention. Integrates feminist principles, inclusivity, and the "Leave No One Behind" standard.	From start to end of an evaluation	High	9–18 months	Medium–High
	Disability-Responsive Evaluation	Integrates the rights and needs of people with disabilities at all stages of evaluation. Improves accessibility and promotes inclusive programme design.	From start to end of an evaluation	Medium–High	9–18 months	Medium
	Social-Equity Responsive Evaluation	Examines how interventions address or perpetuate inequality. Identifies who benefits or is excluded, focusing on equity of both process and results.	From start to end of an evaluation	High	9–18 months	Medium–High
	Transformative Evaluation	Applies a social justice and transformation lens, prioritising marginalised voices and structural change alongside technical rigour	From start to end of an evaluation	High	12–24 months	High
Causal/ Explanatory Orientation	Theory-Based Evaluation	Uses a ToC or logic model to test causal pathways and understand why outcomes occurred. Enables both accountability and learning.	Any time, depending on ToC availability	Medium	6–12 months	Medium–High
Adaptive & Innovation Orientation	Developmental Evaluation	Supports innovation in complex, adaptive environments. Provides real-time feedback to refine strategies and inform emergent decision-making.	Continuous, during implementation	High	Continuous (ongoing)	Medium
Timeliness Orientation	Rapid Evaluation	Produces timely evidence to inform urgent policy or programme adjustments. Prioritises speed while maintaining methodological rigour	At any time; quick turnaround	Low	6–12 weeks	Low–Medium
Environmental Orientation	Environmentally-Responsive Evaluation	Assesses environmental outcomes and climate impacts, including mitigation, adaptation, and ecosystem health.	From start to end of intervention and evaluation	High	12–24 months	Medium–High

9.5 Attribution vs Contribution in Outcome Evaluation

Outcome evaluations differ in whether they aim to attribute change directly to interventions or to assess contribution within complex systems. Choosing the right approach depends on the evaluation purpose, available resources, and the complexity of the intervention environment (OECD/DAC, 2019; White, 2010; Gertler et al., 2016). The diagram below provides a decision tree to guide commissioners and evaluators in selecting the most appropriate design. It distinguishes between high-rigour attribution approaches (Level A evidence), theory-based contribution approaches (Level B evidence), and descriptive designs that provide useful but limited insights (Level C evidence).

Figure 6: Selecting attribution, contribution, or descriptive designs

Decision Tree for Selecting Outcome Evaluation Designs

START

Is the purpose to make strong casual claims about the intevention's effect

YES — Are resources (time, budget, data) and ethical conditions sufficient?
- YES: Attribution approach (RCT / Quasi-experimental) — Level A
- NO: Contribution approach (theory-based / contribution analysis — Level B

NO — Is there a need to understand plausible contribution in a complex system?
- YES: Contribution approach — Level B
- NO: Descriptive monitoring / pre-post — Level C

OUTPUT: Select design, note evidence level, plan resources accordingly

For ease of reference, a summary table of these approaches, their evidence levels, and when to use them is provided in Annexure G.

9.6 Evidence grades in outcome evaluation

The range of approaches outlined in Table 10 highlights the flexibility available to evaluators in designing outcome evaluations. However, different approaches carry different strengths of evidence, costs, and feasibility. To assist commissioners and practitioners in making informed choices, this guideline introduces an evidence-grading rubric that maps common designs to levels of evidentiary confidence. This ensures that methodological decisions are not made in isolation but are guided by both purpose and practical considerations (OECD/DAC, 2019; Gertler et al., 2016; White, 2010).

Table 10: Evidence grades in outcome evaluation

Evidence Grade	Description	Typical Designs	Cost/Time Implications
Level A (High)	Strong causal confidence, outcomes can be attributed with reasonable certainty.	RCTs, robust quasi- experimental designs with pre/post and matched controls, strong baseline & comparator data.	High cost, 12–24 months.
Level B (Medium)	Plausible contribution established; multiple lines of triangulated evidence and independent substantiation.	Contribution analysis, realist evaluation, theory- based designs with mixed methods.	Medium cost, 6–12 months.
Level C (Low)	Descriptive evidence of change; useful for programme learning and monitoring but limited for causal claims.	Pre-post surveys, routine monitoring data, case studies, qualitative inquiry.	Lower cost, shorter duration (weeks–6 months).

10. Planning, prioritising, implementing and managing outcome evaluation

Effective outcome evaluations require careful and deliberate planning from the earliest stages of intervention design. In line with the NEPF and international best practice (OECD/DAC, 2019; UNEG, 2020), government institutions should ensure that key considerations such as ethics, budgeting, timelines, and use of findings are integrated into evaluation plans before implementation begins. This section outlines critical dimensions of managing outcome evaluations: ethical safeguards, budgeting and resource allocation, realistic timelines, and decisions based on evaluation results.

Key Dimensions of Planning Outcome Evaluations

Evaluability Assessment

Check whether the intervention is ready for evaluation:

- Clarify the ToC and outcome definitions.
- Confirm baseline and monitoring data availability and quality.
- Assess timing, scope, budget, and stakeholder readiness.
- Ensure outcomes are measurable and aligned to evaluation purpose.

Selecting the Evaluation Approach

Match the design to purpose, complexity, and resources:

- Clarify evaluation purpose (accountability, learning, decision-making).
- Align approach with intervention type and context (e.g., rapid, participatory, theory-based).
- Balance stakeholder needs with available time, budget, and data.
- Ensure the chosen method supports

Ethics and safeguarding evaluands

Protect the dignity, rights, and safety of all participants:

- Secure informed consent and voluntary participation.
- Guarantee anonymity and confidentiality in reporting.
- Put measures in place to safeguard children and vulnerable groups.
- Ensure evaluators act with honesty, integrity, and independence.
- Follow DPME ethical guidelines and international standards (e.g., UNEG).

Budgeting for outcome evaluation

Allocate resources realistically to ensure quality and credibility:

- Budget 0.1%–5% of programme costs for evaluation.
- Account for data collection (surveys, records, interviews, QA).
- Include costs for specialist expertise (evaluators, analysts, subject experts).
- Provide for capacity building of internal staff to ensure sustainability.
- Don't forget logistics, travel, and project management.
- Avoid underfunding weak budgets risk poor evidence and unused reports.

Result of the outcome evaluation

Evaluation findings should guide clear decisions and next steps:

- Decide whether to continue, scale up, modify, or discontinue the programme.
- Address implementation gaps and strengthen weak areas.
- Determine if additional resources are required for better outcomes.
- Integrate lessons learned into policy, strategy, and operations.
- Share results with key stakeholders for accountability and support.

Peer review and quality assurance

Ensuring credibility and quality requires systematic checks at all stages:

- Build peer review into design, mid-term, and final reporting.
- Use internal reviews for alignment with institutional priorities.
- Commission external reviews to provide independence and rigour.
- Follow the DPME Guideline on Peer Review of Evaluations for roles and timing.
- Apply the EQAF to assess planning, design, implementation, reporting, and use.
- Strengthen learning, transparency, and uptake of findings through consistent QA.

How long should the outcome evaluation take

Outcome evaluations may take 6–12 months or a few weeks (participatory/rapid), depending on scope and methods.

10.1 Evaluability assessment

Planning for an outcome evaluation should begin with an evaluability assessment to determine whether the intervention is ready to be evaluated. This step clarifies the ToC, ensures outcomes are clearly defined and measurable, and reviews the availability and quality of baseline and monitoring data. It also assesses whether the timing and scope of the evaluation are appropriate, whether relevant stakeholders are engaged, and whether the evaluation can be conducted within the available budget and timeframe. Conducting an evaluability assessment before commissioning an evaluation reduces the risk of wasted resources and ensures the process produces meaningful and actionable evidence.

10.2 Selecting the evaluation approach

Selecting the most appropriate evaluation approach is central to ensuring relevance and utility. The choice should be guided by the purpose of the evaluation, the complexity of the intervention, stakeholder needs, and available resources. The NEPF highlights rapid evaluations, collaborative learning through sector reviews, and transversal evaluations as particularly relevant for government. However, additional approaches, including theory-based, participatory, gender-responsive, disability-responsive, social equity, environmental, developmental, and utilisation-focused evaluations, may be appropriate. The decision should also consider the intended use of findings, the degree of participation desired, and the level of methodological rigour required.

10.3 Ethics and safeguarding evaluands

DPME guidelines on ethical conduct and other ethics procedures for evaluation should be followed at all times. Special ethical considerations should be considered for evaluations conducted by internal and external team members. When evaluating interventions that has people or whether it is done with them there is the possibility that their lives could be affected in some way through participation in the evaluation. There is therefore a duty on evaluators to work to minimise the possibility of having adverse effects on the evaluands (Denscombe 2009). When conducting the evaluation, it is important for all evaluators to put safe measures in place so that participants can response with confident and without fear of pressure or misrepresentation. Therefore, the evaluators need to work in a manner that respects the rights and dignity of

evaluands and to act with honesty and integrity. Protection of children and vulnerable adults who are involved in any way in evaluation is important. Hoggarth and Comfort (2010) argues that there are four main ethical issues for all evaluation to be observed are:

- **Informed consent:** People need to know what they are agreeing to
- **Voluntary participation:** People need a free choice about whether to participate and to be able to stop or refuse to answer questions
- **Anonymity:** Readers of any evaluation report should not be able to recognise the person who gave their views
- **Confidentiality:** Other should not know what the respond said or did.

Evaluators conducting the outcome evaluation should also be objective. Matters of honesty, integrity and respect are enshrined in the principles of all codes of ethical practice. These principles as described by Hoggarth and Comfort (2010) are the need for evaluators to:

- Report findings accurately
- Observe legal requirements
- Consider their own safety and the safety of their participants
- Avoid intimate personal relationships with those involved in a study
- Make clear any possible conflict of interest

10.4 Budgeting for outcome evaluation

A general 'rule of thumb' is that an evaluation should be between 0.1% to 5% of an intervention's budget. The cost of conducting an outcome evaluation depends on several critical factors that influence both the scope and complexity of the process. It is the duty of the evaluation manager to ensure that the budget is realistic, as over ambitious and under budgeted scope of work is likely to yield a weak base of evidence and an unused report.

One of the primary considerations is data collection. This includes the costs associated with obtaining relevant data, whether through existing administrative records or by designing and administering new instruments such as surveys, interviews or focus groups. The scale of data collection how many participants, sites or geographic locations are involved also significantly impacts costs, as does the need for quality assurance activities like data cleaning and verification to ensure accuracy. Another crucial component is the expertise required to execute the evaluation successfully. Hiring external

evaluation specialists such as lead evaluators, data analysts and subject matter experts represents a substantial portion of the budget. These professionals bring technical rigor and objectivity, especially important for outcome evaluations which are complex – testing the theory of change and logical frameworks.

Additionally, it is valuable to allocate part of the budget toward capacity building where evaluators provide training and mentoring to staff within the commissioning government institution. This does not only support effective current evaluation participation but also builds internal skills for future internal evaluations contributing to sustainability. The level of evidence demanded by the evaluation also affects costs. Further, evaluations that measure multiple outcomes will necessitate broader data collection efforts, thereby increasing both the time and financial resources needed.

Beyond these core factors additionally, travel and logistical expenses, particularly for evaluations conducted across multiple sites, and project management costs for coordination and administration should be factored in. Thoughtful budgeting across these dimensions ensures the evaluation delivers valuable, trustworthy insights for intervention learning and accountability.

10.5 How long should the outcome evaluation take

If a participatory process is carried out with stakeholders, this could be several weeks including meetings and writing up the results. If a full survey is carried out this could take between 6-12 months. However, it is advisable that commissioning institutions embrace the rapid evaluation approach to produce quality evaluation findings that can inform decision and policy making, resource allocation and programme improvement within reasonable time frame.

10.6 Result of the outcome evaluation

The process of the outcome evaluation is very important. At the end of the outcome evaluation some decisions are needed. These are:
- Whether to continue, scale up, modify or discontinue the programme based on evidence of outcomes.
- How to address identified implementation challenges or gaps to strengthen achievement of identified outcomes.

- What adjustments or additional resources may be necessary to enhance outcome achievement.
- How to integrate lessons learned into policy strategy or operational plans.
- Identifying stakeholders to engage and communicate results with for accountability and further support.

10.7 Peer review and quality assurance

Peer review strengthens the credibility and quality of an outcome evaluation. It should be built into key stages, design, mid-term, and final reporting, to validate methods, data, and findings. Internal reviews ensure alignment with institutional priorities, while external reviews provide independence and methodological rigour.

To ensure systematic consistency, evaluations should adhere to the DPME Guideline on Peer Review of Evaluations, which outlines who should conduct reviews, when they should occur, and how they should be structured, particularly emphasising the roles of methodology and sector experts. Additionally, the Evaluation Quality Assessment Framework (EQAF) provides detailed criteria to assess the planning, design, implementation, reporting, and use phases of evaluations, guiding evaluators toward national quality standards.

By integrating these guidelines into the evaluation cycle, commissioners and evaluators can ensure their evaluation processes are credible, transparent, and fit for use. Quality assurance through these mechanisms also helps institutional learning, bolsters stakeholder confidence, and supports the sustainable uptake of findings in policy and programme decision-making.

....

Signed
Adv. Melanchton Makobe
Acting Director-General
Department of Planning Monitoring and Evaluation
Date: 12/01/2026

Annexure A: Glossary

Term	Definition	Source / Reference
Activity	Actions taken or work performed to produce specific outputs as part of a programme or intervention.	OECD-DAC (2019)
Attribution	The ascription of a causal link between observed changes and a specific intervention, considering other influencing factors.	OECD-DAC (2019)
Baseline	Information collected at (or just before) the start of an intervention that serves as a point of reference for measuring progress and change.	DPME (2019)
Contribution Analysis	An approach to assess the contribution an intervention is making to observed results, recognising that multiple factors influence change.	Mayne (2012)
Counterfactual	The hypothetical scenario of what would have happened to beneficiaries in the absence of the intervention.	White (2010)
Developmental Evaluation	Supports innovation in complex, adaptive environments by providing real-time feedback for strategy adaptation.	Patton (2011)
Disability-Responsive Evaluation	Integrates the rights and needs of people with disabilities at all stages of evaluation. Improves accessibility and promotes inclusive programme design.	UNDP (2018)
Effectiveness	The extent to which an intervention's outcomes were achieved, or are expected to be achieved, taking into account their relative importance.	OECD-DAC (2019)
Efficiency	A measure of how economically resources/inputs are converted to results.	OECD-DAC (2019)
Empowerment Evaluation	Builds the capacity of stakeholders to evaluate their own performance, fostering ownership and learning.	Fetterman (2001)
Environmentally-Responsive Evaluation	Assesses environmental outcomes and climate impacts, including mitigation, adaptation, and ecosystem health.	UNDP (2019)
Evaluation	The systematic and objective assessment of an ongoing or completed intervention, its design, implementation and results.	OECD-DAC (2019)
Formative Evaluation	Evaluation intended to improve performance, typically conducted during implementation.	Scriven (1991)
Gender-Responsive Evaluation	Assesses how gender and power relations change due to the intervention, integrating inclusivity and equity.	UN Women (2015)
Impact		OECD-DAC (2019)

Term	Definition	Source / Reference
Indicators	Quantitative or qualitative variables that provide a simple and reliable means to measure achievement, reflect changes connected to an intervention, or help assess performance.	UNDP (2009)
Inputs	The financial, human, and material resources used to implement an intervention.	OECD-DAC (2019)
Logical Framework (Logframe)	A management tool used to improve the design of interventions, most often summarising objectives, indicators, means of verification, and assumptions.	DPME (2011)
Outcome	The likely or achieved short- and medium-term effects of an intervention's outputs.	OECD-DAC (2019)
Outcome Evaluation	An evaluation that examines the extent to which an intervention has achieved its intended outcomes and contributed to changes in the target population or system.	DPME (2025, proposed)
Outcome Harvesting	A qualitative evaluation approach that collects evidence of what has changed and works backwards to determine whether and how an intervention contributed to these changes.	Wilson-Grau & Britt (2012)
Outputs	The products, capital goods, and services that result from an intervention; may also include changes resulting from the intervention that are relevant to the achievement of outcomes.	OECD-DAC (2019)
Participatory Evaluation	Involves intended beneficiaries directly in the design, data collection, and interpretation of results, fostering ownership and relevance.	Cousins & Whitmore (1998)
Relevance	The extent to which the intervention's objectives and design respond to beneficiaries' needs, country priorities, and partners' policies.	OECD-DAC (2019)
Results Chain	The causal sequence for a development intervention that stipulates the necessary sequence to achieve desired objectives—typically inputs → activities → outputs → outcomes → impacts.	DPME (2019)
Social-Equity Responsive Evaluation	Examines how interventions address or perpetuate inequality, focusing on who benefits or is excluded.	UNDP (2018)
Summative Evaluation	Evaluation conducted after implementation to determine the extent to which anticipated outcomes and impacts were achieved.	Scriven (1991)
Theory-Based Evaluation	Uses a ToC or logic model to test causal pathways and understand why outcomes occurred.	Weiss (1997)
Theory of Change (ToC)	An explanation of how and why a set of activities will bring about the changes the intervention seeks to achieve.	Taplin & Clark (2012)

Term	Definition	Source / Reference
Transformative Evaluation	Applies a social justice and transformation lens, prioritising marginalised voices and structural change.	Mertens (2009)
Triangulation	The use of multiple data sources, methods, and perspectives to increase the credibility and validity of evaluation findings.	Patton (2015)
Utilisation-Focused Evaluation (UFE)	An evaluation approach that prioritises utility and actual use by intended users.	Patton (2008)

Annexure B: Monitoring system readiness checklist

Monitoring Element	Guiding Questions	Status (Yes/No/Partial)	Notes / Gaps
Baseline data	Is there a documented baseline with date, method, and data sources recorded?		
Data collection frequency	Are data collected at intervals aligned with programme milestones and reporting?		
Variable definitions	Are key indicators and variables clearly defined and standardised?		
Data quality checks	Are there processes to verify accuracy, completeness, timeliness, and reliability?		
Metadata availability	Are data sources, methods, and assumptions documented and accessible?		
Data storage & access	Is data systematically stored and accessible to evaluators and managers?		
Disaggregation	Are data disaggregated (e.g., by gender, age, geography, vulnerability)?		
Responsibility/ownership	Is responsibility for data collection and management clearly assigned?		
Data use & feedback loops	Is monitoring data being used regularly for programme reviews and adjustments?		
Ethics & data protection	Are privacy, confidentiality, and ethical safeguards applied to sensitive data?		

Monitoring Element	Guiding Questions	Status (Yes/No/Partial)	Notes / Gaps
System integration	Are monitoring data linked to departmental MIS or national systems where relevant?		
Capacity & training	Do staff responsible for monitoring have adequate skills and resources?		

Annexure C: Prioritisation matrix for outcome evaluations

Criteria	Description	Weight (%)	Score (1–5)	Weighted Score
Strategic Relevance	Alignment with national priorities, MTDP, sector plans, and SDGs	20%		
Decision-timing / Urgency	Findings needed for upcoming policy, budget, or programme decisions	15%		
Materiality	Size of budget, population coverage, or scale of potential impact	15%		
Risk / Public Interest	Political sensitivity, reputational risk, or public demand for accountability	10%		
Evaluability / Data Readiness	Existence of ToC, baseline, outcome indicators, and data access	15%		
Feasibility	Timeframe, resources, and logistical feasibility	10%		
Equity and Inclusion Significance	Extent to which the programme addresses marginalised groups, gender equality, or "leave no one behind"	10%		
Learning Potential	Likelihood that the evaluation will generate transferable insights	5%		
Total		100%		/5

Using the Prioritisation Matrix

Score each potential outcome evaluation (1–5) against the listed criteria where 1 = very low and 5 = very high alignment/ performance. Apply the given weights and sum the weighted scores to rank priorities. Select the highest-scoring evaluations within available budget and capacity, ensuring sectoral and geographic balance. Document the process for transparency and approval.

Annexure D: Outcome evaluation readiness checklist

Readiness Element	Guiding Questions	Implication if Condition is Not Met	Readiness Score (1–5)	Notes / Gaps
Programme maturity	Has the intervention been operational long enough to influence outcomes?	Delay evaluation until measurable results are likely to be observed.		
Current and plausible Theory of Change	Is there a well- documented ToC linking activities, outputs, outcomes with clear assumptions?	Develop or refine ToC before evaluation begins.		
Stable implementation	Has the programme model and delivery approach remained consistent?	Postpone evaluation until the programme stabilises.		
Baseline or comparator data	Is there reliable baseline or credible comparator group data?	Conduct or reconstruct baseline before proceeding.		
Measurable outcome indicators	Are indicators clear, measurable, and linked to the ToC?	Strengthen M&E systems to measure outcomes effectively.		
Monitoring data availability	Are key output and outcome monitoring data available and reliable?	Improve data systems and collection methods before proceeding.		
Stakeholder commitment	Are decision-makers and implementers supportive of the evaluation and willing to use findings?	Engage stakeholders early to secure buy-in.		
Adequate time	Is there enough time for design, data collection, analysis, and validation?	Adjust timelines or scale to fit available time.		
Adequate budget	Is the budget sufficient for high-quality evaluation	Secure additional funding or adjust scope.		
Ethics pathway clear		Resolve ethics requirements before starting.		

Scoring:

- 1 = Not ready at all
- 2 = Major gaps exist
- 3 = Partially ready, gaps need addressing
- 4 = Largely ready, minor gaps remain
- 5 = Fully ready to proceed

This Outcome Evaluation Readiness Checklist is a practical tool for assessing whether a programme is prepared to undergo an outcome evaluation. It covers key elements such as programme maturity, the presence of a credible Theory of Change, stability of implementation, data availability, stakeholder commitment, resources, and ethical readiness. Each element includes guiding questions, the implications of not meeting the condition, and a readiness score (1–5) to help identify strengths and gaps. The checklist supports informed decision- making on whether to proceed with an evaluation immediately, delay it, or address specific gaps to ensure credible, useful, and timely findings.

Annexure E: Ethics and safeguarding checklist for outcome evaluations

Focus Area	Minimum Requirements	Practical Guidance
Informed Consent & Assent	Written or verbal consent for adults; assent forms for children plus parental/guardian consent	Provide forms in plain language and local languages; explain purpose, risks/benefits, voluntary participation, right to withdraw
Voluntary Participation	Participants free to decline or withdraw without consequences	Reconfirm rights at data collection; provide contact info for complaints/questions
Protection Pathways for Disclosures of Harm	Protocol for reporting harm or abuse cases	Establish referral contacts (e.g., social services, child protection, health facilities); ensure evaluators act within 24–48 hrs; document referrals
Anonymisation & Data Protection	Protect personal identifiers and ensure data security	Use encryption for electronic data, lockable storage for hard copies, define retention (e.g., 3–5 years), secure deletion protocols
Accessibility & Inclusion	Ensure participation of people with disabilities and vulnerable groups	Provide materials in accessible formats (Braille, large print, audio); ensure venues are accessible; adapt tools as needed
Quality Assurance & Independence	Peer review and QA mechanisms in place	Build peer review at design, mid-term, and final report; submit for DPME Quality Assurance Framework (EQAF) review where applicable

Annexure F: Readiness assessment rating scale

Rating	Description	Interpretation
1	Condition not met at all.	Significant gaps exist; evaluation should not proceed until major preparatory work is done.
2	Condition minimally met.	Key elements are missing; substantial improvements are required before proceeding.
3	Condition partially met.	Some elements are in place, but important gaps could affect the credibility or usefulness of results.
4	Condition mostly met.	Minor gaps exist; evaluation could proceed with mitigation measures in place.
5	Condition fully met.	All necessary elements are in place; evaluation is ready to proceed without significant risk to quality.

Annexure G: Attribution vs Contribution approaches and evidence grades for outcome evaluation

Approach	Evidence Grade	Typical Methods / Designs	Use Case	Strengths	Limitations
Attribution	Level A (High Evidence)	RCTs; quasi-experiments (e.g., difference- in-difference, matching methods).	When strong causal claims are needed, e.g., scaling up interventions, reallocating significant resources.	Provides highest rigour and causal certainty.	Resource- and time-intensive; ethical and political feasibility may be constrained.
Contribution	Level B (Moderate Evidence)	Theory-based designs (contribution analysis, realist evaluation, process tracing).	When multiple factors shape change, and attribution is unrealistic (e.g., complex policy/ programme environments).	Generates credible insights into plausible causal links; adaptable to complexity.	Moderate resource requirements; less certainty than counterfactual designs.
Descriptive	Level C (Low Evidence)	Monitoring data; administrative data; pre/post comparisons without counterfactuals.	When resources are constrained and rapid or resource- efficient insights are needed.	Feasible, timely, and useful for programme management and learning.	Limited for causal claims; findings must be interpreted with caution.

References

Centers for Disease Control and Prevention. (2024). Framework for program evaluation in public health. MMWR Recommendations and Reports, 73(6), 1–49. https://www.cdc.gov/mmwr/volumes/73/rr/rr7306a1.htm

Chirau, T., Blaser-Mapitsa, C., & Amisi, M. (2021). Policies for evidence: A comparative analysis of Africa's national evaluation policy landscape. Evidence & Policy, 17(3), 535–548. https://doi.org/10.1332/174426421X16104826256918

Dartington. (2006). Defining outcomes for children's services: A framework for local authorities. Dartington Social Research Unit.

Denscombe, M. (2009). Ground rules for social research: Guidelines for good practice. McGraw-Hill Education (UK).

Department of Planning, Monitoring and Evaluation (DPME). (2014). Evaluation Quality Assessment Framework (EQAF). Pretoria: DPME.

Department of Planning, Monitoring and Evaluation (DPME). (2014). Guideline on Peer Review. Pretoria: DPME.

Department of Planning Monitoring and Evaluation (DPME). (2019). The National Evaluation Policy Framework (NEPF). DPME.

Earl, S., Carden, F., & Smutylo, T. (2001). Outcome mapping: Building learning and reflection into development programs. International Development Research Centre (IDRC).

Filgueiras, F., & Queiroz, L. de F. N. (2021). The governance of public policy evaluation systems: Policy effectiveness and accountability. Organizações & Sociedade, 28(96), 208–232. https://doi.org/10.1590/1984-92302021v28n9609EN

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). Impact evaluation in practice (2nd ed.). World Bank.

Görgens, M., & Kusek, J. Z. (2010). Making monitoring and evaluation systems work: A capacity development toolkit. World Bank.

Hoggarth, L., & Comfort, H. (2010). A practical guide to outcome evaluation. Jessica Kingsley Publishers.

Matlala, L. S. (2025). Factors affecting the use of evidence in public sector programmes in South Africa: A systematic review of Outcome 8 programmes. Paper presented at the Conference on Digital Government Research. [DOI:10.59490/dgo.2025.1024]

Organisation for Economic Co-operation and Development – Development Assistance Committee (OECD/DAC). (2019). Better criteria for better evaluation: Revised evaluation criteria definitions and principles for use. OECD Publishing.

Patton, M. Q. (2008). Utilization-focused evaluation (4th ed.). Sage Publications.

Segone, M., Bamberger, M., & Reddy, S. (2015). National evaluation policies for sustainable and equitable development: How to integrate gender equality and social equity in national evaluation policies and systems. UNICEF.

South African Monitoring and Evaluation Association (SAMEA). (2014). Guideline on evaluation ethics. SAMEA.

United Nations Evaluation Group (UNEG). (2020). UNEG ethical guidelines for evaluation. UNEG.

University of Kansas. (2024). Developing a framework for outcome evaluation. In Community toolbox. https://ctb.ku.edu/en/table-of-contents/evaluate/evaluation/framework-for- evaluation/main

White, H. (2010). A contribution to current debates in impact evaluation. Evaluation, 16(2), 153–164. https://doi.org/10.1177/1356389010361562

Wholey, J. S. (2010). Handbook of practical program evaluation (3rd ed.). Jossey-Bass. https://www.eval.fr/wp-content/uploads/2020/04/Handbook-of-Practical-Program- Evaluation_compressed.pdf

Wilson-Grau, R. (2015). Outcome harvesting. BetterEvaluation. http://betterevaluation.org/plan/approach/outcome_harvesting

Wilson-Grau, R., & Britt, H. (2013). Outcome harvesting. Ford Foundation.

W.K. Kellogg Foundation. (2004). Logic model development guide. W.K. Kellogg Foundation.