

1.2

Theories of Change Linking National M&E Systems to Development Outcomes: A Scoping Review

Douglas Glandon
Guilherme Cunha Lima

Abstract

This scoping review explores a critical but underexamined topic in public sector reform: how national monitoring and evaluation (M&E) (eco)systems are theorized to contribute to improved development outcomes. Drawing on six country cases—Australia, India, Mexico, Namibia, Nigeria, and South Africa—the review describes how governments conceptualize, structure, and attempt to operationalize M&E systems as tools for governance improvement and policy effectiveness. This synthesis follows a scoping-review protocol grounded in JBI guidance, with reporting per PRISMA-ScR.

Rather than evaluating the performance or internal consistency of these systems, the review identifies the key components, assumptions, and intended causal linkages embedded in their official theories of change (ToCs). These include mechanisms such as improved decision-making through feedback loops, greater accountability through performance monitoring, and stronger institutional capacity for evidence use—each emphasized to varying degrees across country models. While some ToCs are grounded in existing institutional practices or informed by prior diagnostics, others are more aspirational, articulating long-term goals not yet fully realized. In some cases, different components within a single ToC vary in how closely they are tied to implementation experience versus future ambitions. Distinguishing between these elements is important for those seeking to adapt or apply such models in other contexts.

Across the six cases, both common features and key differences can be observed. Commonalities include multi-stage or layered causal pathways that reflect systemic thinking; explicit articulation of assumptions underpinning system performance; efforts to address both supply- and demand-side dynamics of evidence use; and initiatives to institutionalize M&E within the broader architecture of governance. At the same time, the models differ in their scope and boundaries. These differences reflect varying definitions of what constitutes a national M&E system and shape both institutional architecture and reform trajectories.

Based on features and divergences observed across the six models, the review identifies illustrative strategies that may inform the development or refinement of national M&E system ToCs. These include: (1) institutionalizing incentives and accountability mechanisms; (2) embedding feedback loops in planning and budgeting cycles; (3) strengthening knowledge brokering functions; (4) addressing workforce retention and public sector absorption challenges; (5) integrating political economy strategies to manage resistance and promote uptake; and (6) sequencing reforms according to contextual readiness. These are considerations—not prescriptive recommendations—for governments, development partners, and practitioners supporting government M&E systems strengthening.

Introduction

This scoping review explores a central but underexamined question in the field of public sector reform and evaluation system strengthening: How are national monitoring and evaluation (M&E) (eco) systems theorized to contribute to improved development outcomes? Although M&E systems have proliferated globally over the past two decades, many continue to have limited influence on policy and service delivery outcomes, as documented in evaluations and global diagnostics. Among national systems that aspire to development impact, relatively few publicly articulate how evaluation activities, institutional reforms, or capacity-building efforts are expected to contribute to improved governance, service quality, or accountability.

This review contributes to addressing this gap by describing and comparing how six national M&E system theories of change are structured, the roles they envision for evaluation, and the causal pathways they propose. The review focuses on theories of change articulated in official or quasi-official strategy documents from Australia, India, Mexico, Namibia, Nigeria, and South Africa. These models vary in their institutional location—some, like Mexico's CONEVAL, operated with technical autonomy, while

others are embedded within central government bodies. The six cases were selected for their diversity of context, institutional maturity, and ambition. Some reflect long-standing M&E traditions, while others represent newer policy efforts driven by donor influence or domestic reform momentum. Together, they illustrate a range of ways in which governments conceptualize the potential contribution of M&E systems to broader development outcomes.

The review is organized in three parts. First, the section titled “Description of National M&E System Theories of Change” presents detailed accounts of each country’s theory of change, identifying the stated components, assumptions, and causal linkages embedded in each model. Second, the section titled “Observed Common Features and Differences” highlights both recurring elements and key points of divergence, including how the models describe feedback loops, define system boundaries, address incentive structures, and engage demand-side dynamics. Third, the section “Strategic Elements for Strengthening Theories of Change Linking National M&E (Eco)systems to Improved Development Outcomes” presents a set of illustrative, context-dependent considerations that may help governments and partners refine their ToCs. The review concludes with a summary of key observations and implications for M&E system development.

Throughout the document, the review emphasizes description and comparison—not evaluation or formal analysis. It does not assess the empirical validity or performance of the six ToCs, nor does it propose an overarching synthesis model. Instead, it seeks to support clearer understanding of how national M&E systems are currently theorized to contribute to governance and development goals, and to identify potentially useful patterns or design considerations that can inform future efforts. Where possible, the review also reflects on whether the components of each ToC appear to be grounded in active implementation or aspirational intent.

Conceptual Framework

This scoping review is one of a set of twenty scoping reviews commissioned to inform the Global Evaluation Initiative (GEI) 2.0 Strategy (FY2026–30). The review set is organized around a conceptual framework that mirrors the nested structure of GEI’s theory of change which addresses four analytical levels: country, decision-maker/unit, evaluation/decision, and individual. At the *country* level, the focus is on the linkage between stronger M&E ecosystems and development and governance outcomes. At the *decision-maker/unit* level, the emphasis is on strengthening the culture and practice of evidence-informed decision-making within specific ministries or units. At the *evaluation/decision* level, the focus is on how evaluations or other evidence-generating activities can inform specific policy or program decisions. Finally, at the *individual* level, the framework addresses how training, mentorship, peer exchange, and related supports strengthen the skills, performance, and professional development of evaluators and other evaluation actors. Within each level, the review set addresses five recurring sub-topics: how the focal construct is *defined and conceptualized*; the primary *theories of change* linking the level’s activities to its intended outcomes; the *metrics* used to characterize status and change at that level; the *empirical evidence* linking activities to outcomes; and the *drivers* that help explain success or failure. The resulting 4×5 matrix — four analytical levels by five recurring sub-topics — structures the review set into twenty distinct but complementary scoping reviews.

The present review addresses the *theory of change* sub-topic at the *country* level.

Methods

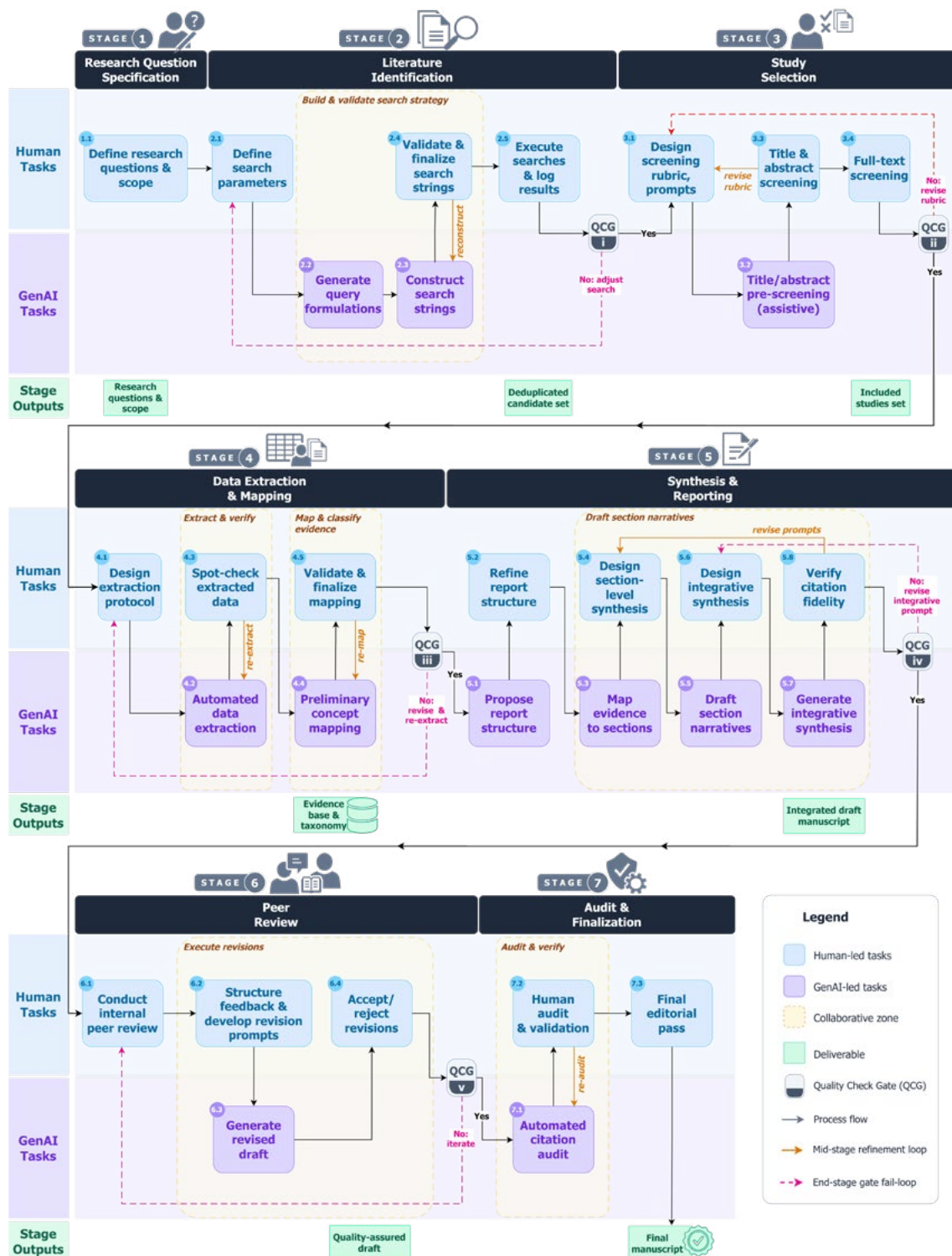
Review design

This review was conducted as a scoping review following JBI guidance (Peters et al., 2020), with reporting per PRISMA-ScR (Tricco et al., 2018). A scoping format was appropriate because the research question is causal-pathway-oriented and required breadth across heterogeneous, multidisciplinary material — peer-reviewed scholarship, national strategy and planning documents, and institutional

theory-of-change frameworks. Eligibility followed the JBI Population–Concept–Context (PCC) framing, with the review question and inclusion criteria anchored primarily to the concept of theories of change linking country-level M&E (eco)systems to improved development outcomes. Per §11.1.1 of the JBI Manual, no formal critical appraisal or risk-of-bias assessment was applied.

The review followed the seven-stage human-in-the-loop workflow shown in **Figure 1**, beginning with human-led specification of the research question and scope (Stage 1, Task 1.1). Across all stages, LLM tools supported defined tasks under human-designed prompts; two human reviewers retained authority over every inclusion, charting, mapping, and synthesis decision. Figure 1's five "collaborative zones" mark groupings of paired human and GenAI tasks that operated on the same evolving artifact through iterative cycles, with each actor's work shaping the other's next task.

Figure 1. GEI 2.0 Rapid Evidence Review (RER) Workflow



Eligibility criteria

Sources were eligible if they articulated or analyzed a theory of change — whether framed as an explicit causal pathway, logic model, results framework, or conceptual framework with articulated causal assumptions — linking the functioning of country-level M&E systems to improved development outcomes. Because both scholarly and practitioner-institutional papers were within scope, eligible source types included peer-reviewed journal articles; institutional strategy and planning documents from national governments, regional evaluation bodies, and multilateral agencies; national evaluation policies and accompanying frameworks; and ECD and evaluation-system guidance documents containing embedded theories of change. No a priori date or language restriction was applied; included sources are entirely English-language and span 2017–2024, reflecting the realized scope of the literature meeting the eligibility threshold. At full-text review, sources were excluded for failing to articulate any explicit causal logic, articulating a causal logic for a different object or pathway, or not meeting the minimum quality threshold applied to their source type.

Two source-type-specific quality thresholds were applied. Peer-reviewed sources had to have been issued through standard editorial peer-review processes. Grey-literature sources additionally had to meet three criteria: (i) articulate an explicit causal logic linking country-level M&E systems to development outcomes, rather than describing M&E activities without an articulated causal pathway; (ii) a documented evidence or reference base; and (iii) identifiable institutional quality control (verified pragmatically by checking whether the source was published through an official institutional channel and bore clear markers of institutional authorship, approval, or accountability).

Information sources and search strategy

Literature identification (Stage 2 of Figure 1) followed an iterative human–GenAI cycle, beginning with the human reviewer's definition of search parameters (Task 2.1). Because theories of change linking country-level M&E systems to development outcomes are primarily articulated in institutional strategy and planning documents rather than in peer-reviewed journals, the source base was expected to be weighted toward institutional grey literature — and was confirmed to be at inclusion. Scopus and Web of Science Core Collection provided peer-reviewed coverage, with Google Scholar searched as a complementary academic search engine. Four institutional repositories were systematically searched: the World Bank Open Knowledge Repository, the UNEG Publications Repository, the OECD iLibrary, and Global Evaluation Initiative (GEI) network publications. Targeted website searches of national evaluation office portals were conducted as a structured supplementary stream, given the centrality of national-government strategy documents to this question. Coverage was further extended through backward and forward citation chasing, expert recommendation, and key-author tracking focused on scholars whose prior work has shaped the M&E-system theory-of-change literature.

Within the "build and validate search strategy" collaborative zone, the LLM generated semantically equivalent query reformulations and drafted database- and repository-specific search strings (Tasks 2.2–2.3), which the human reviewer audited and finalized (Task 2.4); a search log archived run dates, applied limits, and result counts. Searches were executed under a pre-specified stopping rule (Task 2.5): searching continued through successive results pages until three consecutive pages yielded no further relevant records. **Quality Check Gate (QCG) i** required the deduplicated candidate set to meet coverage and retrievability criteria; where it did not, reviewers revised search parameters or strings and re-executed searches.

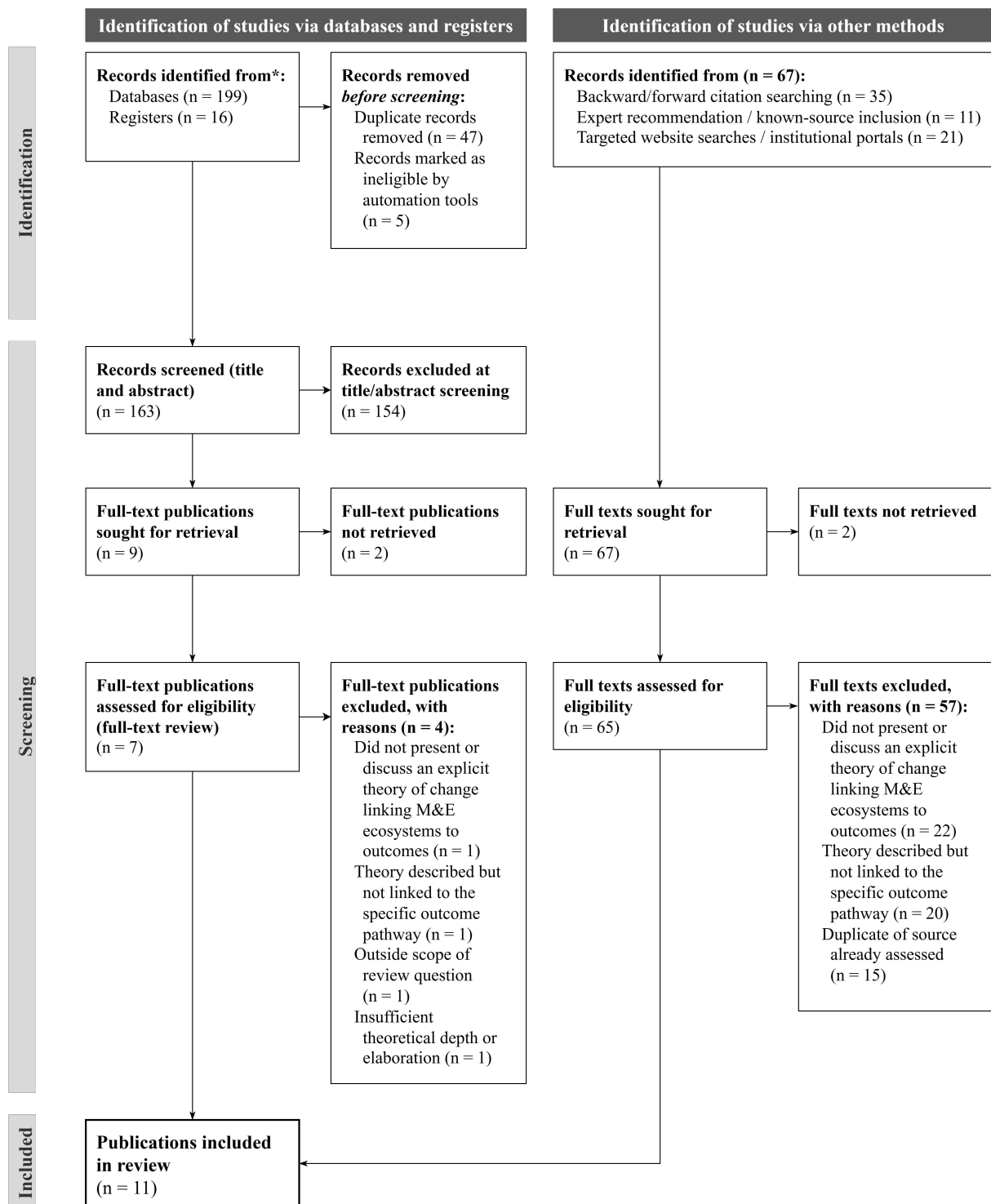
Source selection

Study selection (Stage 3) operated under a human-designed screening protocol (Task 3.1). A total of 215 records were identified through databases and registers (199 from databases; 16 from registers), with a further 67 identified through backward and forward citation searching, expert recommendation and known-source inclusion, and targeted website searches and institutional portals (**Figure 2**). After deduplication and metadata-based screening to remove records of document types outside the

eligible scope, 163 records underwent title and abstract screening: the LLM produced assistive first-pass inclusion-likelihood labels (Task 3.2), and two independent human reviewers performed the title and abstract screening (Task 3.3), informed but not bound by the LLM labels. Nine records advanced to full-text screening (Task 3.4); two could not be retrieved, leaving seven assessed, of which four were excluded with coded reasons. In parallel, 65 of the 67 supplementary-stream records were retrieved and assessed (two were not retrievable), and 57 were excluded with coded reasons. A total of 11 sources were included in the review, corresponding to six country cases (Australia, India, Mexico, Namibia, Nigeria, and South Africa) and a small set of cross-cutting institutional documents.

QCG ii required the included studies set to satisfy protocol-specified eligibility coverage; where it did not, the reviewers revised the screening protocol and re-screened the affected records.

Figure 2. PRISMA Flow Diagram



Data extraction and mapping

Data extraction and mapping (Stage 4) proceeded under a human-designed extraction protocol (Task 4.1). A structured codebook captured bibliographic metadata; the articulated causal pathway; the institutional actors positioned as drivers of change; the assumptions stated; and the contextual conditions identified, with page-anchored quotations retained for every theory-bearing element. Within the "extract and verify" collaborative zone, LLM-assisted extraction (Task 4.2) populated the codebook with page-anchored quotations and paraphrases, with iterative spot-checks and prompt refinement (Task 4.3) used to re-extract fields where verification flagged errors — including misattributions, unsupported statements, hallucinated content, and conflation of proximate constructs. Reviewers verified all causal-pathway statements and theory-bearing extractions and a random 20% sample of remaining fields. Within the "map and classify evidence" zone, a preliminary concept mapping produced under structured prompts (Task 4.4) was then validated and finalized by the reviewers (Task 4.5), comparing the eleven theories along two axes: the principal causal mechanism invoked (structural, relational, capacity-based, or accountability-based) and the primary institutional actor positioned as driver of change.

QCG iii required the resulting evidence base and taxonomy to meet extraction-fidelity criteria — including page-anchored support for every charted statement — before synthesis. Where these criteria were not met, the process looped back to Task 4.1, where the reviewers revised the extraction protocol before the affected fields were re-extracted.

Synthesis and reporting

Synthesis and reporting (Stage 5) proceeded through collaborative drafting cycles. A draft report structure was proposed under structured prompts (Task 5.1) and refined by the reviewers (Task 5.2). Within the "draft section narratives" collaborative zone, evidence was mapped to sections (Task 5.3), section-level synthesis logic was designed under human direction (Task 5.4), and section narratives were generated under human-designed prompts (Task 5.5) — producing, in this review, a synthesis architecture organized primarily by mechanism family, with secondary ordering by institutional actor, and concluding with strategic considerations for strengthening theories of change linking country M&E (eco)systems to development outcomes. An integrative prompt designed by the reviewers (Task 5.6) was then used to produce the integrated draft (Task 5.7).

A human verification of citation fidelity (Task 5.8) sampled a random 20% of cited claims and checked each against its source. Where errors, misattributions, or hallucinations appeared in the sample, the reviewer examined whether they were scattered across sections and sources or concentrated in one section or source; based on that assessment, the process looped back to Task 5.4, where the section-level synthesis prompts were revised and the affected section narratives regenerated. Where the sample was clean, the process advanced to QCG iv.

QCG iv assessed the structural and rhetorical quality of the integrated draft — whether the sections were well tied together, built on one another, and formed a coherent whole with clear flow and consistent framing. Where it did not, the reviewers revised the integrative synthesis prompt (Task 5.6) and the LLM regenerated the integrated draft.

Quality assurance

Before finalization, the integrated draft was processed through a GenAI-assisted internal peer review (Stage 6). The reviewer entered text-anchored marginal comments in the integrated draft (Task 6.1); a VBA macro converted those comments into inline, LLM-ingestible comments embedded in the body text (Task 6.2). The LLM then returned an iteratively revised draft with all changes shown in tracked changes (Task 6.3), and the reviewer accepted or rejected each proposed revision (Task 6.4). **QCG v** marked the point at which the model returned a revised draft requiring no further peer-review comments; where further comments remained, the process looped back to Task 6.1 for another cycle.

Final audit and finalization (Stage 7) used a purpose-built claim-and-citation audit pipeline (Task 7.1) within the "audit and verify" zone: the pipeline parsed every in-text citation, retrieved the corresponding extracted source passage, and used an LLM under a structured rubric to classify each cited claim as Supported, Partially Supported, or Not Supported. In the human audit and validation pass (Task 7.2), all Not Supported and Partially Supported claims were revised or removed, and a 10% random sample of the remaining Supported claims was manually re-verified against source PDFs. The audit was then re-run, with the cycle repeated until every cited claim returned a Supported verdict. A final editorial pass (Task 7.3) covered coherence and flow, formatting against the publication template, terminological consistency, and APA-compliance of the reference list.

Description of National M&E System Theories of Change

Australian Centre for Evaluation (ACE) Strategy

The Australian Centre for Evaluation (ACE) Strategy theory of change outlines a cohesive and interconnected approach to strengthening evaluation practice across the Australian Public Service (APS) (Australian Government, 2024). Rather than delineating outcomes by time horizon (e.g., short-, medium-, or long-term), the strategy articulates a vision of improved public policy and program performance through the use of high-quality, ethical, and culturally appropriate evaluation. This vision is supported by four interdependent streams of activity: Evaluation Leadership and Promotion, Impact Evaluation Delivery and Advice, Evaluation Planning and Use, and Evaluation Capability Building. Each stream is intended to contribute to specific outcomes while reinforcing the others, collectively working toward a public sector that embeds evaluation into decision-making processes.

The Evaluation Leadership and Promotion stream focuses on fostering an enabling environment for evaluation by engaging key networks, promoting the Productivity Commission's Indigenous Evaluation Strategy as the whole-of-government framework for evaluations of policies and programs affecting Aboriginal and Torres Strait Islander peoples, and facilitating leadership through entities such as the Commonwealth Evaluation Reference Group and the Evaluation Community of Practice. This stream is designed to help raise system-wide awareness of the value and requirements of high-quality evaluation, strengthen senior executive support, and encourage appropriate resourcing for evaluation activities. These outcomes depend not only on internal coordination but also on leadership commitment to positioning evaluation as a core governance function.

The Impact Evaluation Delivery and Advice stream centers on delivering and promoting rigorous impact evaluations, particularly through experimental and quasi-experimental methods. ACE supports this by forming partnerships with government departments and providing technical guidance to ensure methodological rigor and ethical integrity. While the strategy promotes experimental and quasi-experimental methods, it explicitly recognizes that not all programs are suitable for such designs—listing "feasibility of conducting a high-quality impact evaluation" among six selection criteria for identifying candidate policies and programs, and naming the identification of suitable candidates as a key assumption in the theory of change. The strategy is less specific, however, about how candidate identification will work in complex or politically sensitive policy areas. These efforts are intended to demonstrate the potential benefits of such evaluations in practice and foster increased uptake across government. This stream may provide concrete practice examples that help shift perceptions of what impact evaluation can achieve, feeding into leadership buy-in and systemic demand.

Through the Evaluation Planning and Use stream, ACE works to embed evaluation considerations early in the policy cycle, including during Cabinet and Budget processes. By offering guidance on integrating evaluation into new policy proposals, post-implementation reviews, and overall planning frameworks, this stream seeks to ensure that policy and program proposals are grounded in evaluative foresight, that a diverse range of government initiatives are subject to quality evaluations, and

that evaluative thinking becomes a routine input into policy formulation and budget decisions. This stream supports the view that evaluation should be a deliberate component of public administration rather than an afterthought or discrete, ex-post exercise.

The Evaluation Capability Building stream underpins all others by developing the skills, tools, and professional infrastructure needed to sustain evaluation practice. It supports the design and delivery of tailored training, the development of foundational and specialist courses, and the establishment of an APS Evaluation Profession in collaboration with the Australian Public Service Commission. These initiatives are expected to elevate capability across agencies to commission, manage, and use evaluation effectively. This technical foundation is positioned as critical for realizing the broader cultural and institutional shifts envisioned in the other streams.

These four streams are intended to function as a mutually reinforcing system: leadership and advocacy cultivate an enabling environment; rigorous delivery demonstrates value; embedded planning creates structural demand; and capability building ensures that public servants are equipped to meet that demand. However, while the Commonwealth Evaluation Reference Group provides cross-entity coordination, the strategy does not specify mechanisms—such as shared cross-stream indicators, joint workplans, aligned performance targets, or others—to align its four streams with one another in practice. Together, the four streams are intended to advance the ACE vision: a public sector where evaluation informs decisions, improves program performance, and ultimately enhances the wellbeing of Australians.

The theory of change rests on several key assumptions: that political leaders, central agencies, and senior executives will sustain their commitment to evaluation; that suitable programs will be identified for impact evaluation; that capability development will lead to actual behavioral change; that demand will increase in response to demonstrated value and support; and that leaders will treat evaluation as central to the stewardship of public resources and prioritize investment in it accordingly. These assumptions highlight the need for continuous engagement and strategic alignment across stakeholders in the broader evaluation ecosystem.

Although the ACE theory of change diagram does not depict feedback loops between streams, the strategy explicitly describes several adaptive learning mechanisms in its monitoring approach—including routine planning sessions, after-action debriefs, stakeholder feedback, and the State of Evaluation survey—and characterizes both the strategy and its theory of change as living documents to be updated as lessons emerge (Australian Government, 2024). This orientation suggests a potential for iterative learning over time.

While several components of the ACE Strategy—such as the establishment of the Evaluation Community of Practice, the rollout of pilot evaluations in select departments, and the development of capability-building resources—are already being implemented and build on existing APS infrastructure (Australian Government, 2024), the broader goal of embedding evaluation as a routine input into all stages of Commonwealth decision-making remains aspirational. Realizing this goal will depend on the wider APS reform context — including the central workforce, capability, and stewardship challenges identified in the 2023 Capability Review of the Australian Public Service Commission (Australian Public Service Commission, 2023) — within which the ACE Strategy is being operationalized.

India's Development Monitoring and Evaluation Office (DMEO)

India's Development Monitoring and Evaluation Office (DMEO) theory of change outlines a highly structured roadmap for strengthening the national monitoring and evaluation (M&E) ecosystem in pursuit of more effective, efficient, equitable, and sustainable public programs (Development Monitoring and Evaluation Office [DMEO], 2021). While articulated through six distinct stages—foundational activities, short-term objectives, medium-term objectives, immediate results, intermediate outcomes, and the overarching goal—the theory is not linear. Instead, it presents a systemically oriented frame-

work in which progress is framed as contingent on multiple reinforcing elements and critical assumptions. Together, these components are intended to embed a culture of evidence use across institutions and contribute incrementally to the achievement of India's development priorities and the Sustainable Development Goals (SDGs).

The theory begins with a set of foundational activities that create the institutional and operational conditions thought necessary to support M&E reform. These activities include securing financial resources, establishing fit-for-purpose organizational structures and technical systems, recruiting and training a skilled M&E workforce, and fostering a culture of evidence use through partnerships, outreach efforts, cross-learning, and joint studies. These investments form the bedrock of the broader strategy, with their success contingent on the assumption that political stability and sustained support for M&E exist and will continue to underpin institutional efforts. Without this foundational commitment, downstream reforms would lack the legitimacy, resources, and traction needed to take hold across the public sector.

Building on this foundation, the theory articulates a set of short-term objectives grouped into three interconnected thematic pillars. The first, enabling evidence-based policymaking, involves integrating M&E into government scheme expenditure proposals and institutionalizing performance tracking through the adoption of a national evaluation policy. The second, strengthening data systems and data architecture, centers on building infrastructure for key sectors and enhancing the technical capacities of data and strategy units. The third, expanding M&E capacity, is especially broad and foundational—it includes supporting nascent M&E systems at multiple levels of government, introducing academic M&E courses in leading institutions, developing a national curriculum and competency framework, and providing staff training through collaborations with experts. These short-term objectives aim to normalize M&E practices across institutions and establish the structural and human capacity needed for sustained evidence generation and use.

The medium-term objectives deepen the reforms initiated in the previous stage. Within the evidence-based policymaking pillar, the emphasis evolves toward strengthening outcome-based monitoring and facilitating multidimensional knowledge-sharing platforms. For data systems and architecture, the focus shifts to supporting the development of state-level data systems (DMEQ, 2021) and assessing administrative data governance and quality at the central ministry/department level through initiatives such as the Data Governance Quality Index (DGQI) (DMEQ, 2021; Aggrawal, 2024), alongside the application of advanced analytics and evidence-informed tools. In the capacity pillar, attention turns to targeted technical assistance for key stakeholders and the formation of enduring partnerships to bolster the broader M&E ecosystem. The successful realization of these objectives depends on two critical and interrelated assumptions: that stakeholders possess sufficient capacity and resources to implement M&E reforms, and that they are willing to invest in and strengthen these systems in practice. These assumptions reflect the model's sensitivity to both technical readiness and institutional will.

If these enabling conditions hold, the short- and medium-term objectives are expected to yield three clusters of immediate results, aligned with the same thematic structure. Within the policy pillar, these include the formal adoption and implementation of a national M&E policy, the routine application of evaluation and outcome-based performance monitoring across government schemes, and demonstrable government action grounded in M&E findings. On the data front, strengthened data strategy units begin consistently collecting high-quality monitoring data, supported by the integration of data governance and quality assessment frameworks like DGQI and infrastructure like the Government Integrated Repository for Governance (GIRG), thus enhancing evidence-informed decision-making. In the capacity domain, Ministries/Departments (M/Ds) and states develop the capability to commission and conduct high-quality evaluations, as well as to track SDG performance, while a critical mass of trained evaluators emerges. DMEQ itself is envisioned as a national hub of M&E professionals and experts.

These immediate results are expected to contribute to four intermediate outcomes that reflect systemic transformation, assuming that stakeholders will sustain their commitment and institutional effort to ensure quality and consistency in M&E implementation. The intermediate outcomes include measurable improvements in developmental results at central and state levels; increased action on course corrections, rationalization, and program strengthening at central and state levels, based on evidence-driven policy recommendations; a rise in demand for evidence among stakeholders; and the emergence of a standardized and strengthened national M&E capacity, practice, and ecosystem.

The theory of change concludes with a clearly articulated goal: to enhance the effectiveness, efficiency, equity, sustainability, and overall impact of government programs across India, ultimately contributing to the achievement of national development goals and the SDGs. This is framed as both a technical and developmental objective, with broad implications for public sector performance and the achievement of national and global development goals. Although the model does not explicitly reference feedback loops or adaptive learning mechanisms, its phased structure and assumption-based sequencing suggest a flexible, iterative approach—one that could evolve based on political realities, institutional learning, and stakeholder responsiveness.

Many components of DMEO's ToC are currently active. The DGQI framework has evolved from version 1.0 to 2.0, with the latter expanding to encompass 74 Ministries/Departments and 650 schemes (Agrawal, 2024). However, the broader goal of establishing a truly integrated, nationwide M&E culture across India's federal and state systems remains aspirational, and its realization will depend on overcoming persistent institutional and capacity challenges.

Mexico's National Council for the Evaluation of Social Development Policy (CONEVAL)

CONEVAL's theory of change presents a structured and institutionally embedded approach to advancing evidence-informed policymaking in the domain of social development (CONEVAL, 2018). Prior to its dissolution by the Senate in June 2025, CONEVAL was an institution of the Mexican State with technical autonomy, mandated to evaluate social development programs and policies and generate reliable information on poverty levels. Anchored in its mandate to generate, synthesize, and disseminate data on poverty and social programs, the model envisioned improved social development policy performance and stronger public accountability as its ultimate impact. This vision was pursued through an iterative cycle of four interacting processes — planning, implementation, evaluation, and feedback (CONEVAL, 2018, p. 20) — operationalized through interconnected activities, outputs, and feedback loops that collectively aimed to embed rigorous evidence into decision-making processes across multiple branches of government and society.

The theory of change was organized around four interacting pillars (CONEVAL, 2018, pp. 44–45): multidimensional poverty measurement; planning; monitoring and evaluation of policies and social programs; and follow-up on recommendations and use of information. The technical core comprised three primary activities — poverty and social-lag measurement, evaluations of policies and programs, and ongoing program monitoring. These activities were carried out by CONEVAL's specialized teams and coordinated across the institution's administrative, evaluative, analytical, and executive functions. Coordination with the Ministry of Finance and Public Credit (SHCP) and, historically, the Secretariat of the Public Function (SFP) was constitutive of the system: these institutions jointly issued the General Evaluation Guidelines in 2007 and co-developed the Annual Evaluation Program (CONEVAL, 2018, pp. 39–41). Together, they formed the foundation for generating credible and relevant evidence. The immediate outputs of these activities included detailed poverty statistics, evaluation reports, diagnostic analyses, and tailored policy recommendations—each intended to highlight programmatic gaps and opportunities for reform.

These outputs fed directly into a critical intermediary process: the formulation of social development analysis and evidence-based recommendations. This phase played a pivotal mediating role by translating technical findings into accessible and actionable knowledge for diverse audiences. The model posited that this analytical layer would enhance the relevance and usability of evidence, thereby increasing its likelihood of uptake. A distinctive operational element of the model is the Aspectos Susceptibles de Mejora (ASM) mechanism (CONEVAL, 2018, pp. 94–96), through which evaluation findings are classified by the institutional scope of action required — specific, institutional, interinstitutional, or intergovernmental — and converted into formal, tracked improvement commitments, operationalizing the feedback loop between evaluation outputs and program reform.

From this intermediary stage, the theory of change outlined multiple pathways for evidence use. On the institutional side, the National Commission for Social Development and the Inter-Secretarial Commission for Social Development served as coordination bodies — across federal, state, and municipal levels and within the executive branch, respectively — and as the institutional venues for resolving cross-cutting improvement recommendations requiring action by multiple actors (CONEVAL, 2018, p. 96 and fn. 53). On the demand side, a broad array of stakeholders—including the legislative branch, the Ministry of Finance and Public Credit, and citizens—were expected to engage with and act upon the evidence produced. These users were envisioned to incorporate evaluative insights into decisions regarding budget allocation, program design and operation, and the strategic direction of social programs.

The intended outcome of this process was an increase in the actual use of evidence in policymaking, budgeting, and oversight. At the impact level, CONEVAL's theory of change aspired to enhance both the effectiveness of social development policy and the quality of public accountability. These dual goals reflected the model's broader vision of a public sector that is not only informed by evidence but held accountable for performance.

While many components of CONEVAL's ToC were actively operational at the time of the *Memorias* — including multidimensional poverty measurement, the Annual Evaluation Program, the ASM follow-up mechanism, and the use of evaluation evidence in some budget decisions — other elements remained partially realized, including the systematic incorporation of findings into the Federal Expenditure Budget, the consolidation of subnational M&E systems, and the resolution of cross-cutting recommendations requiring coordinated action by multiple actors (CONEVAL, 2018, pp. 170–178).

Namibia's National Monitoring and Evaluation Policy

Namibia's National Monitoring and Evaluation Policy (National Planning Commission [NPC], 2023) presents an internally coherent theory of change designed to embed monitoring and evaluation into the core of public sector governance. The model is grounded in a systemic vision of evidence-informed, results-oriented decision-making, and aims to build an institutional culture in which M&E informs planning, budgeting, and implementation at all levels of government. Using a standard results chain structure—inputs, activities, outputs, outcomes, and impact—the theory of change offers a logically sequenced pathway for strengthening both the demand and supply of M&E, supported by clearly articulated assumptions and a strong emphasis on institutional coordination.

The theory of change begins with a single foundational input considered necessary for catalyzing reform: sustained commitment to the allocation of resources for building national evaluation capacity and system (NPC, 2023, Fig. 2). These inputs are conceptualized not only as prerequisites for initiating activities, but also as critical for sustaining momentum and institutional credibility. In particular, the policy highlights the importance of high-level endorsement and funding to establish and maintain the M&E infrastructure within Offices, Ministries, and Agencies (OMAs), Regional Councils (RCs), and Local Authorities (LAs). Without these conditions in place, the entire causal chain is assumed to be at risk of stalling.

Building on these inputs, the policy outlines a diverse set of activities that target multiple dimensions of the M&E ecosystem. These include public awareness campaigns to increase appreciation for the value of M&E and the development of national tools and frameworks (such as guidelines, evaluation plans, and performance tracking systems). A feature of Namibia's model is the integration of performance-based budgeting (PBB) into the M&E reform agenda, positioning M&E as a potential driver of strategic resource allocation rather than merely a compliance mechanism. Capacity-building efforts—through training programs and curriculum development—are also central to this component, aiming to cultivate a professionalized and technically competent M&E workforce.

The immediate outputs envisioned include the establishment and operationalization of M&E units across OMAs, RCs, and LAs; the development and dissemination of technical tools to support evaluation practice; and the expansion of academic offerings in M&E, such as postgraduate diplomas. These outputs are intended to institutionalize both the infrastructure and knowledge base necessary to anchor M&E within Namibia's public administration. Importantly, these outputs serve as the visible manifestations of policy commitment, intended to signal progress to both internal stakeholders and external partners.

The anticipated outcomes encompass both capacity enhancement and institutional transformation. Trained officials are expected to apply their skills to improve the quality and consistency of M&E practices across sectors. Simultaneously, an expanded academic pipeline is projected to increase the supply of qualified M&E professionals available to support both governmental and non-governmental actors. Institutional coordination among OMAs, RCs, and LAs—while not itself an outcome in the model—is treated as a key assumption underpinning the outcome of an improved national M&E system (NPC, 2023, Fig. 2). The model suggests that these outcomes will collectively lead to a more integrated, efficient, and responsive national M&E system.

The long-term impact envisioned by the theory of change is the full embedding of M&E in governance processes. This includes routine integration into policy formulation, budget planning, and decision-making at all levels of government. At this stage, M&E is expected to evolve from a technical function into a strategic instrument for improving public sector performance. The ultimate goal is a governance culture that values learning, adapts based on evidence, and is accountable to its citizens.

The theory of change clearly articulates key assumptions that underpin the transition from individual outputs to their corresponding immediate outcomes, from specific immediate outcomes to their corresponding intermediate outcomes, and from individual intermediate outcomes to the intended impact.

For example, the progression from the adoption of tools to support performance-based budgeting (output) to the development of an enabling environment for planning, using, conducting, and managing M&E (outcome) depends on the assumption that adequate resources will be allocated to system strengthening (NPC, 2023). Similarly, the transition from M&E unit officials acquiring enhanced knowledge and skills (outcome) to a national capacity to generate M&E evidence that informs policymaking, budgeting, programming, and planning (impact) relies on the assumption that students pursuing M&E degrees will be motivated to enroll in postgraduate programs. Likewise, the advancement from improvements in Namibia's M&E system—such as the establishment of evaluation units, plans, guidelines, and policies (outcome)—to the broader goal of strengthened national capacity for evidence-informed decision-making (impact) hinges on the sustained establishment and nurturing of partnerships to support M&E development. The model further identifies two cross-cutting assumptions that condition the entire pathway: that OMAs, RCs, and LAs will be willing to change their behavior and attitudes toward M&E, and that buy-in and ownership from government and non-government stakeholders will continue to strengthen (NPC, 2023, Fig. 2). These assumptions underscore the political and organizational dependencies that must be in place for the theory of change to hold.

Despite these frameworks, systemic uptake remains partial. The NPC (2021) Situation Analysis found that only 17% of OMAs have established central M&E units, with just 22% having clearly defined M&E roles and responsibilities (as cited in Nshimiyimana & Rabie, 2024b). Documented weaknesses include a lack of internal capacity (NPC, 2023) and the absence of high-placed champions within sector ministries to drive the system, despite relatively stronger central coordination by the NPC (Nshimiyimana & Rabie, 2024b). Furthermore, while the Child Welfare Directorate has a programme-wide M&E plan with an associated theory of change and logic model, no dedicated theory of change, logic model, or M&E framework exists for the Child Support Grant programme specifically (Nshimiyimana & Rabie, 2024a).

Nigeria's National Monitoring and Evaluation Policy

Nigeria's National Monitoring and Evaluation Policy (Federal Ministry of Finance, Budget and National Planning [FMFBNP], 2022) outlines a structured, layered theory of change—organized across five tiers (inputs, outputs, intermediate outcomes, outcomes, and impact) with multiple parallel pathways at each level—aimed at strengthening evidence-informed governance through the institutionalization of results-based management (RBM). The model's overarching vision is to improve public sector accountability, transparency, and service delivery by embedding M&E systems that generate, disseminate, and apply timely and high-quality evidence. It positions M&E as both a technical and governance reform instrument—potentially essential for shifting decision-making processes toward greater effectiveness, responsiveness, and learning. The theory of change serves as a roadmap for transforming Nigeria's institutional culture by enabling state and non-state actors to use performance evidence to shape development outcomes, reinforce democratic accountability, and elevate citizens' quality of life.

The theory of change begins with a set of foundational inputs intended to create the enabling conditions for system reform. These inputs include adequate funding from government and development partners, a skilled human-resource base, and material resources (guidance, tools, software, equipment, systems, and office space) that together create the operational platform on which the M&E system is to be built. Training, professional cadre formation, and individual-capacity gains for RBM are not treated as inputs but as outputs and intermediate outcomes further up the chain (FMFBNP, 2022, p. 5). These resources are not just technical prerequisites; they reflect a broader commitment to institutional readiness and policy coherence across ministries, departments, and agencies (MDAs).

Building on these inputs, the model envisions a series of outputs that collectively shape the architecture of a national M&E system. These outputs operate on both the supply and demand sides of evidence production and use. On the supply side, the development of a national M&E operational framework—including standardized guidelines, procedures, and tools—is expected to improve the consistency and quality of evaluation practice. Legislative efforts to embed RBM principles into the regulatory framework are intended to enhance institutional accountability, while the creation of formal M&E structures within MDAs establishes clear lines of responsibility and internal capacity. On the demand side, the model seeks to stimulate political and civic appetite for evidence use, particularly by promoting policy uptake among decision-makers, encouraging inter-institutional coordination, and building the capacity of civil society organizations (CSOs) to monitor public investments. Outputs also include the establishment of evaluation repositories and knowledge management systems to facilitate access, transparency, and reuse of evaluative findings.

These outputs are designed to generate a series of intermediate outcomes that mark the functional deepening of M&E practices across the public sector. As evidence systems mature, the policy anticipates greater availability, accessibility, and use of high-quality, timely data at national and subnational levels. Performance monitoring and reporting mechanisms are expected to become more institutionalized, embedded in routine planning and budgeting cycles. Human and organizational capacities for RBM are projected to expand, supporting more consistent implementation and oversight of development initiatives. Importantly, the model also anticipates greater demand for evaluation—both from

within the public administration and from non-governmental actors—along with strengthened legal frameworks and more robust citizen engagement in governance processes. These intermediate outcomes act as both markers of progress and potential drivers of further system consolidation.

The theory of change culminates in two high-level outcomes: strengthened results-based planning and management across all tiers of government, and a deepened culture of learning, transparency, and accountability within public institutions (FMFBNP, 2022).

At the highest level, the theory of change articulates two complementary impact statements: an overarching vision of “Better quality of life and prosperity for all Nigerians,” and the establishment of an effective governance system anchored on result-based management, learning, transparency, and accountability that guarantees sustainable socio-economic development (FMFBNP, 2022, p. 6). The policy text further specifies two corollary impacts: improved citizen confidence in government through systematic MDA accountability for program results, and improved program effectiveness and public accountability. Together, these impacts tie M&E reform directly to Nigeria’s development agenda, framing evidence as a public good that supports state responsiveness and citizen well-being.

The theory of change is explicit in acknowledging the contextual risks that could undermine its success. These include macroeconomic volatility tied to unpredictable oil revenue, delays in the approval of the Evaluation Policy itself and corresponding gaps in government funding for M&E activities, insecurity in the Northeast that affects external investment, weak public financial management, and a broader deficit of good governance (FMFBNP, 2022, p. 5).

While the policy is grounded in recent government commitments, implementation remains subject to these systemic challenges (FMFBNP, 2022).

South Africa’s National Evaluation System (NES)

South Africa’s NES, as articulated in its revised theory of change (Department of Planning, Monitoring and Evaluation [DPME], 2018), is presented as building the foundational blocks of a longer institutionalization trajectory. DPME (2018) frames the NES as a “20-year project” in which the first five years (2011–2016) established core building blocks — evaluation plans, capacity-building structures, quality assurance mechanisms, and communication channels — while identifying institutionalization, including more systematic budget allocations, dedicated evaluation posts, and possible legislative anchoring, as the next phase rather than an already-achieved state. This framework is grounded in a multi-tiered causal logic, specifying how strategic investments in inputs and activities are assumed to generate outputs, catalyze outcomes, and contribute to transformative impacts. The revised ToC differs from the initial version in three respects (DPME, 2018, §7.6): inputs are made explicit rather than assumed; quality assurance is added as an outcome; and improvement plans are added — described in the report as “important in closing the evaluation cycle.”

The ToC begins with a set of foundational inputs that form the institutional and operational bedrock of the NES. These include the formal establishment of an evaluation unit within the Department of Planning, Monitoring and Evaluation (DPME); the allocation of funding for evaluations at departmental, national, and provincial levels; and the DPME’s provision of co-financing for evaluations under the National Evaluation Plan (NEP). Human resources—comprising both government officials and external service providers—are also identified as critical inputs. Additionally, funding is earmarked specifically for capacity building, alongside the development of a comprehensive capacity-building plan that serves dual purposes: strengthening demand for evaluation and enhancing technical understanding and utilization of evaluation findings across stakeholder groups.

These inputs enable a coordinated set of core activities necessary to institutionalize evaluation practice. These activities include the collaborative management and resourcing of the evaluation unit, the implementation of capacity-building initiatives, and the provision of technical advice to departments conducting evaluations. Other operational activities encompass the facilitation of peer review processes, the post hoc quality assessment of evaluations, and the development of guidelines, policies, and

standards to ensure coherence and quality. Crucially, the model specifies the actual implementation of evaluations and the development of Departmental Evaluation Plans (DEPs) and Provincial Evaluation Plans (PEPs), reflecting a commitment to mainstream evaluation into planning cycles across levels of government.

The outputs of these activities comprise a more structured and systematized evaluation architecture. These include the development of the NEP and associated systems, which together establish the procedural and institutional framework for evaluating government programs. Outputs also include the operationalization of effective systems to manage these evaluations, supported by a functioning quality assurance mechanism. Equally important are outputs related to the promotion of evaluation culture—namely, initiatives focused on learning, sharing, and advocacy designed to raise stakeholder awareness, strengthen evaluation capacity, and build commitment across the public sector and civil society.

This phase marks the institutional consolidation of the NES, where systems move from design to implementation and from isolated evaluations to coordinated processes with established quality control mechanisms. These outputs are essential for enabling structured engagement with evaluation findings and preparing the ground for their systematic application in policy and programming.

The outcomes articulated in the ToC are oriented around fostering the effective use of evaluation and research evidence in decision-making processes. One critical outcome is the enhanced capacity of stakeholders to manage and apply such evidence—a result directly tied to the earlier emphasis on training and capacity-building inputs. This growing capacity is expected to facilitate the sharing of evaluation and research evidence, promote its comparison with existing departmental and service delivery plans, and enable the development of improvement plans that respond directly to identified gaps or weaknesses.

The revised ToC includes improvement plans as outcomes — developed, implemented, funded, and monitored for impact — explicitly framed in DPME (2018) as “closing the evaluation cycle.” At the time of the evaluation, however, improvement-plan tracking was described as a gap rather than an operational mechanism, with DPME (2018) recommending the development of a centralized tracking system as a precondition for routinely closing this loop.

The theory also suggests that evaluation and research evidence should inform substantive changes to government interventions, marking the critical moment at which evaluation shifts from being a technical exercise to becoming an instrument for public sector reform. This link between evaluation use and policy adaptation is central to the NES’s vision of a responsive and evidence-informed administration.

At the highest level of the causal chain, the NES aims to contribute to a set of transformative impacts. These include the improved effectiveness of public services and the eradication of poverty, achieved through better-targeted and more responsive interventions. Government and public accountability are also expected to improve as a result of increased transparency and the institutionalization of evidence-informed decision-making processes. A further impact is the replication of successful interventions, facilitated by systematic learning from evaluation findings. Another major goal is to support more rational and equitable allocation of public resources, based on the credibility and relevance of evaluation evidence. Notably, the ToC emphasizes that these processes must involve affected stakeholders extensively and consistently—reflecting a normative commitment to inclusion, legitimacy, and democratic accountability.

The recursive nature of the ToC becomes evident at this stage: impact-level improvements—such as better service delivery or poverty reduction—are expected not only as end goals but as catalysts that reinforce political commitment to evaluation, stimulate further demand for evidence, and strengthen institutional engagement across government.

Compared to other models reviewed in this study, South Africa's NES is among the most mature, with many of its components having already undergone implementation, refinement, and institutionalization over more than a decade. Since 2011, 67 national evaluations had been completed or were underway within the NES, covering over US \$10 billion of government expenditure (Goldman et al., 2019). Although the evaluation found it difficult to assess use reliably across the system, preliminary evidence from nine departmental and provincial case studies suggested that instrumental use (in seven of nine cases) and process use (in five of nine cases) of evaluations were widespread; system-wide tracking remained limited by erratic progress reports against improvement plans (Goldman et al., 2019). Goldman et al. (2019) characterize capacity building, communication, and quality assurance as areas of considerable progress alongside identified needs for further strengthening, while flagging the absence of legislative anchoring (a draft Planning Bill was presented but delayed) and the limited tracking of improvement plans as more substantive constraints—together indicating that the NES's broader vision of systematic, evidence-informed governance across all sectors remains a work in progress.

Observed Common Features and Differences

Despite their differing institutional contexts, the six national theories of change reviewed share several recurring design features. Most notably, all six adopt some form of layered or phased causal architecture. These frameworks reflect efforts to move beyond linear models and capture the complexity of public sector reform. For example, Australia's strategy is structured around four interrelated streams, each intended to function in parallel while contributing to a shared vision of improved public sector performance (Australian Government, 2024). India's DMEO model also reflects this systemic orientation, combining a sequenced presentation of results with a theory of change that emphasizes interdependencies and critical assumptions at each stage—rather than a simple linear progression (DMEO, 2021).

Another commonly observed feature is the explicit identification of assumptions that underpin each model's causal logic. India's theory details assumptions about political stability, institutional capacity, and stakeholder engagement at multiple stages of the change pathway. Namibia's model similarly articulates dependencies related to higher education systems, intergovernmental coordination, and financial commitment (NPC, 2023). These assumptions serve as internal anchors for each model, helping to clarify the conditions under which the theory is expected to hold.

Several theories also describe both supply- and demand-side dimensions of M&E system functioning. Nigeria's policy includes actions aimed at expanding technical capacity for evidence generation while also seeking to engage civil society and promote the use of findings among policymakers (FMFBNP, 2022). Namibia and South Africa incorporate public awareness campaigns, professionalization strategies, and institutional coordination mechanisms that seek to foster demand and build a supportive evidence culture (NPC, 2023; DPME, 2018). These approaches reflect an emerging emphasis on evaluation ecosystems rather than isolated technical functions.

A fourth common feature is the presence of institutional mechanisms intended to support the routine integration of evaluation into public administration. These include dedicated evaluation units, formal quality assurance systems, professional development pathways, and national evaluation plans. For example, South Africa's model identifies departmental and provincial evaluation plans as instruments for embedding evaluation into planning and budgeting cycles. Australia is establishing a formal APS Evaluation Profession (scheduled to launch in late 2024) in partnership with the Australian Public Service Commission, and already supports communities of practice and leadership structures—including the Commonwealth Evaluation Community of Practice and the Commonwealth Evaluation Reference Group—to anchor evaluation across the bureaucracy (Australian Government, 2024). Such mechanisms aim to create organizational stability and enable continuity of practice beyond political or administrative turnover.

At the same time, there are notable differences across the models—particularly in the boundaries they draw between monitoring, evaluation, and data systems. Some models, such as Australia’s ACE strategy, focus explicitly and narrowly on evaluation. Others, like India’s DMEO theory of change, incorporate broader elements including administrative data architecture, performance monitoring, and data governance assessment frameworks like the Data Governance Quality Index (DGQI). This divergence in scope reflects unresolved questions in the field about what constitutes a “national M&E system” and where institutional responsibility for monitoring ends and evaluation begins. It also underscores an ongoing tension within the field of evaluation capacity development (ECD): while some stakeholders view evaluation as the primary entry point for system strengthening, others argue that evaluation cannot be disentangled from the monitoring and data ecosystems that shape its uptake and utility. In turn, this variation in system boundaries influences not only institutional architecture but also the types of capacity-building strategies and reform pathways each model emphasizes.

Strategic Elements for Strengthening Theories of Change Linking National M&E (Eco)systems to Improved Development Outcomes

Based on the common features and notable differences observed across the six national theories of change, this section presents a set of strategic elements that may be useful for governments, development partners, and evaluation professionals engaged in designing or revising M&E system theories of change. These elements are not offered as prescriptions or conclusions drawn from formal analysis. Rather, they aim to illustrate recurring and contrasting design choices that may support reflection and adaptation across settings.

The first element involves the potential integration of incentives and accountability mechanisms into M&E system design. Several of the reviewed ToCs emphasize leadership commitment and normative promotion of evaluation use but do not specify binding requirements or consequences for inaction. In some contexts, governments may wish to explore mechanisms such as legal mandates for evaluation, budgetary earmarks for evidence generation, performance contracts, or requirements for departments to document the use of evaluation findings. Such mechanisms could help create structural incentives that encourage uptake, particularly among middle management and operational units.

A second element concerns the operationalization of feedback loops and learning mechanisms. Many theories reference feedback or adaptation but stop short of specifying how lessons from evaluation are reviewed or used to inform decisions. Governments might consider linking evaluation findings to structured moments of policy reflection—such as midterm budget reviews, implementation forums, or executive retreats—to enable systematic learning. South Africa’s emphasis on improvement plans and follow-up tracking illustrates one approach. Identifying clear triggers for reflection and specifying institutional responsibilities may enhance the use of evidence in real-time decision-making.

A third element is the development of knowledge brokering functions to help translate evaluation findings into actionable insights. While most ToCs focus on evidence generation, few describe how this evidence is curated, synthesized, and communicated in accessible formats. While some countries may benefit from establishing a central M&E coordination body or observatory to support these functions, such bodies are unlikely to meet the needs of all evidence users across policy sectors. In complex or decentralized settings, it may be more effective for a central unit to provide overarching guidance and frameworks for knowledge translation, while delegating implementation to sector-specific or subnational units with greater contextual familiarity. Alternatively, governments might consider developing knowledge brokering capacity at multiple levels of administration, tailored to programmatic scope, technical complexity, and audience needs. Regardless of the structure, brokering functions may help bridge gaps between technical outputs and policy audiences, while also supporting cross-government learning.

A fourth element relates to human capital development and retention. Several models call for investing in academic training pipelines to increase the supply of M&E professionals, but they offer limited discussion of how these individuals will be absorbed into government roles or retained over time. Countries might explore strategies such as creating dedicated evaluation career tracks, linking training to promotion criteria, or offering secondments in evaluation units to build capacity and institutional demand. Aligning talent development with public sector hiring, incentives, and long-term career pathways may increase the sustainability of workforce investments.

A fifth area of focus is the consideration of political economy factors in theory of change design. Many models assume that credible evidence will be used if it is timely and accessible, but do not explicitly account for the institutional and political dynamics that may constrain or enable uptake. Stakeholder mapping, identification of internal reform champions, and efforts to create political space—such as briefings for legislators or public campaigns to build demand—could help strengthen the enabling environment for evaluation use. In some contexts, protections for decision-makers acting on politically sensitive findings may also be relevant. These strategies aim to align evidence systems with the realities of policymaking environments.

Finally, countries may wish to sequence reforms in accordance with system readiness and institutional capacity. Some models envision the rapid rollout of complex reforms, such as performance-based budgeting or national evaluation mandates, without detailed attention to foundational infrastructure or subnational disparities. In such cases, piloting reforms in data-ready ministries, phasing capacity-building investments, or simplifying initial evaluation protocols may help build momentum while managing risks. Tailored sequencing can allow systems to adapt gradually, while maintaining alignment with long-term goals.

The potential relevance of these strategic elements will vary by context. Governments and partners engaged in strengthening national M&E systems may find it useful to reflect on which approaches align with their existing institutional arrangements, political dynamics, and implementation capacities.

Conclusion

Synthesis of Key Findings

The six national theories of change reviewed in this study illustrate how governments conceptualize and organize the role of monitoring and evaluation (M&E) within broader governance and development agendas. Although developed in and for different institutional contexts, each model offers a view into how public officials and policymakers envision causal pathways linking M&E system strengthening efforts to intended improvements in policy, service delivery, or national development outcomes.

Taken together, these models show considerable variation in how theories of change are structured and specified. South Africa's revised theory of change, developed through a formal evaluation of the National Evaluation System, emphasizes feedback loops, quality assurance processes, and embedded learning mechanisms. India's and Namibia's models follow sequenced results chains that emphasize capacity building, standardization, and coordination. Several models—such as those from Australia and Mexico—highlight leadership, evaluation quality, and the promotion of a culture of use as central drivers of change. Across cases, common features include multi-layered causal logics, attention to supply- and demand-side enablers of evidence use, and a focus on institutionalizing M&E functions through professionalization and planning integration. Just as importantly, these models vary in the extent to which their components are grounded in prior implementation experience versus forward-looking aspirations.

The theories of change reviewed also reflect different assumptions about how change happens. Some models emphasize long-term investments in capability and culture, while others give priority to structural or procedural reforms. Some locate leadership at the center of the model; others place emphasis on technical infrastructure or the design of individual evaluations. Across this diversity, the models collectively reflect the evolving ways in which governments seek to embed M&E in the work of governance and public decision-making.

Rather than offering a uniform blueprint, the country cases illustrate the range of approaches governments use to theorize how M&E systems contribute to national objectives. This diversity reflects underlying differences in institutional histories, policy priorities, and administrative capacities, as well as the distinct strategic choices governments make when articulating how M&E can support public value.

Implications for M&E System Strengthening

The descriptive review of six national theories of change suggests several areas that may be relevant for governments and partners involved in M&E system design or reform. These implications are not presented as definitive conclusions but as reflections informed by the common and divergent elements, design gaps, and key assumptions observed across the models:

1. Embedding M&E into core governance functions may support institutionalization.

Rather than treating M&E as a parallel process, some models describe integration into planning, budgeting, and performance management cycles. In contexts where M&E is closely linked to these functions, the theory of change tends to frame evaluation as part of routine decision-making rather than an external requirement.

2. Theories of change may benefit from clearer attention to incentives and follow-through mechanisms.

Many models emphasize leadership commitment and professional norms but do not identify structures—such as mandates, conditional funding, or public reporting—that help translate commitment into consistent use. Some governments may find it useful to consider how formal incentives can reinforce evaluation use at multiple levels of the system.

3. Operationalizing feedback loops is both a design and institutional challenge.

While most models reference adaptation and learning, few specify how evidence is reviewed, acted upon, or tracked over time. Structured opportunities for reflection—such as post-budget reviews, executive dialogues, or learning events—may help bridge the gap between evaluation and decision-making.

4. Knowledge brokering may help address the gap between evidence generation and use.

The reviewed models generally focus more on the production of evaluation than on its dissemination or translation. Dedicated structures or functions that synthesize findings, develop policy briefs, and promote peer learning may support broader use of evaluative evidence.

5. Addressing political and institutional dynamics may increase the plausibility of intended system outcomes.

Theories of change that assume rational uptake of evidence may not fully reflect real-world policy environments. In some contexts, it may be helpful to map stakeholders, engage political leadership, or create space for decision-makers to act on evidence without reputational risk.

6. Tailored reform sequencing may reduce implementation challenges.

Several models envision rapid system-wide transformation, but implementation may be more feasible when reforms are introduced gradually or piloted in data-ready sectors. Sequencing reforms according to institutional readiness may support adaptation and reduce risks of overload.

These reflections aim to support ongoing thinking about the design and implementation of M&E system theories of change. As the six national cases illustrate, there is no single model or linear path. But across contexts, efforts to clarify causal assumptions, identify key enablers, and align design features with institutional realities may increase the strategic value and long-term sustainability of national M&E systems.

References

- Aggrawal, N. (2024). A case study on data governance and quality index framework of NITI Aayog. In *Using strategy analytics for business value creation and competitive advantage* (pp. 372–387). IGI Global Scientific Publishing.
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Australian Government. (2024). *Australian Centre for Evaluation strategy*. Department of the Treasury. https://evaluation.treasury.gov.au/sites/evaluation.treasury.gov.au/files/2024-10/australian_centre_for_evaluation_strategy.pdf
- Australian Public Service Commission. (2023). *Capability review of the Australian Public Service Commission*. Australian Government. <https://www.apsc.gov.au/sites/default/files/2023-07/230630%20-%20APSC%20Capability%20Review%20Report%20-%20FINAL.pdf>
- Consejo Nacional de Evaluación de la Política de Desarrollo Social. (2018). *Memorias del CONEVAL 2006–2018: Sistema de monitoreo y evaluación de la política de desarrollo social en México*. CONEVAL.
- Department of Planning, Monitoring and Evaluation. (2018). *Evaluation of the National Evaluation System*. Government of South Africa.
- Development Monitoring and Evaluation Office. (2021). *Vision, mission, and strategy*. NITI Aayog, Government of India. <https://dmeo.gov.in/content/vision-mission-and-strategy-0>
- Federal Government of Nigeria. (2022). *National monitoring and evaluation policy of Nigeria*. Federal Ministry of Finance, Budget and National Planning.
- Goldman, I., Deliwe, C. N., Taylor, S., Ishmail, Z., Smith, L., Masangu, T., ... (2019). Evaluation² – Evaluating the national evaluation system in South Africa: What has been achieved in the first 5 years? *African Evaluation Journal*, 7(1), a400. <https://doi.org/10.4102/aej.v7i1.400>
- National Planning Commission. (2023). *National monitoring and evaluation policy*. Republic of Namibia, Office of the President.
- Nshimiyimana, B., & Rabie, B. (2024a). A results-based monitoring and evaluation system for the Namibian Child Support Grant programme. *African Evaluation Journal*, 12(1), a716. <https://doi.org/10.4102/aej.v12i1.716>
- Nshimiyimana, B., & Rabie, B. (2024b). Assessment of the results-based monitoring and evaluation policy frameworks in Namibia. *Administratio Publica*, 32(2), 138–163. <https://doi.org/10.61967/adminpub/2024.32.2.9>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Peters, M. D. J., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Chapter 11: Scoping reviews. In E. Aromataris & Z. Munn (Eds.), *JBI Manual for Evidence Synthesis*. JBI. <https://doi.org/10.46658/JBIMES-20-12>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>