

2.4

M&E Systems Strengthening and Evidence Use: A Scoping Review of the Empirical Evidence

Douglas Glandon
Guilherme Cunha Lima

Abstract

This scoping review investigates the empirical evidence linking efforts to strengthen Monitoring and Evaluation (M&E) systems to changes in the demand for, supply of, and use of evaluative evidence in policymaking. It aims to guide practitioners and policymakers seeking to enhance the quality and utilization of evidence to inform decision-making. This synthesis follows a scoping-review protocol grounded in JBI guidance, with reporting per PRISMA-ScR.

Key findings suggest that demand for evaluation is typically low where M&E systems are fragmented or donor-driven. However, sustained political leadership, deliberate stakeholder coordination, and active civil-society engagement can stimulate domestic demand. On the supply side, national capacity remains uneven; while some countries rely heavily on external actors, others—such as South Africa and Uganda—have institutionalized M&E frameworks and invested in local capacity. The use of evaluation findings is most likely when evaluations are embedded within government planning and budgeting processes, particularly when champions help align evidence with policy priorities and timing. Even under favorable conditions, however, evidence use tends toward incremental operational improvements rather than transformative policy change.

Case examples from Ghana, Zambia, Malawi, Sri Lanka, South Africa, and the Caribbean illustrate both progress and persistent challenges. These cases show that while external support can catalyze reforms, sustainable systems require political will, domestic ownership, and institutional integration. Sri Lanka's Managing for Development Results (MfDR) framework and South Africa's Department of Planning, Monitoring, and Evaluation (DPME) are highlighted in the reviewed literature as notable examples of embedding evaluation in national governance, though challenges in follow-through and uptake persist across contexts.

The literature reviewed here suggests that strengthening M&E systems has the potential to improve evidence-informed policymaking, but that realizing this potential likely requires long-term investments in institutional infrastructure and sustained political engagement. Further research is needed, particularly on causal pathways and long-term outcomes.

Introduction

This review addresses two core research questions: (1) What is the empirical evidence linking M&E systems strengthening to improved evaluative evidence demand, supply, and/or use? and (2) Which case examples best exemplify this evidence? Understanding these questions is critical for practitioners and policymakers who seek to enhance the relevance, quality, and utilization of evaluation findings in shaping policies and programs. If M&E systems are weak, evaluative evidence may remain underused, resulting in missed opportunities for data-driven decisions. Conversely, well-functioning M&E structures may deepen government accountability, stimulate learning, and improve development outcomes—though the strength and consistency of this relationship across contexts remains an open question.

In what follows, the scoping review is organized into two main sections. First, it examines the empirical evidence on how M&E system reforms shape the demand, supply, and use of evaluative evidence, drawing on studies from sub-Saharan Africa, South Asia, and the Caribbean. Next, it presents case examples in which M&E strengthening initiatives have been implemented, illustrating both successes and challenges. The review concludes by integrating key empirical insights with standout case illustrations, thereby clarifying broader implications and identifying gaps in current research.

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 and theory of action 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 *empirical evidence* sub-topic at the decision-maker/unit 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 evidence-oriented and required breadth across heterogeneous, multidisciplinary material — peer-reviewed scholarship, institutional and multilateral evaluation reports, and case-study compilations. Eligibility followed the JBI Population–Concept–Context (PCC) framing, with the review question and inclusion criteria anchored primarily to the concept of the empirical link between M&E systems strengthening interventions and improvements in evaluative evidence demand, supply, and use. Per §11.1.1 of the JBI Manual, no formal critical appraisal or risk-of-bias assessment was applied.

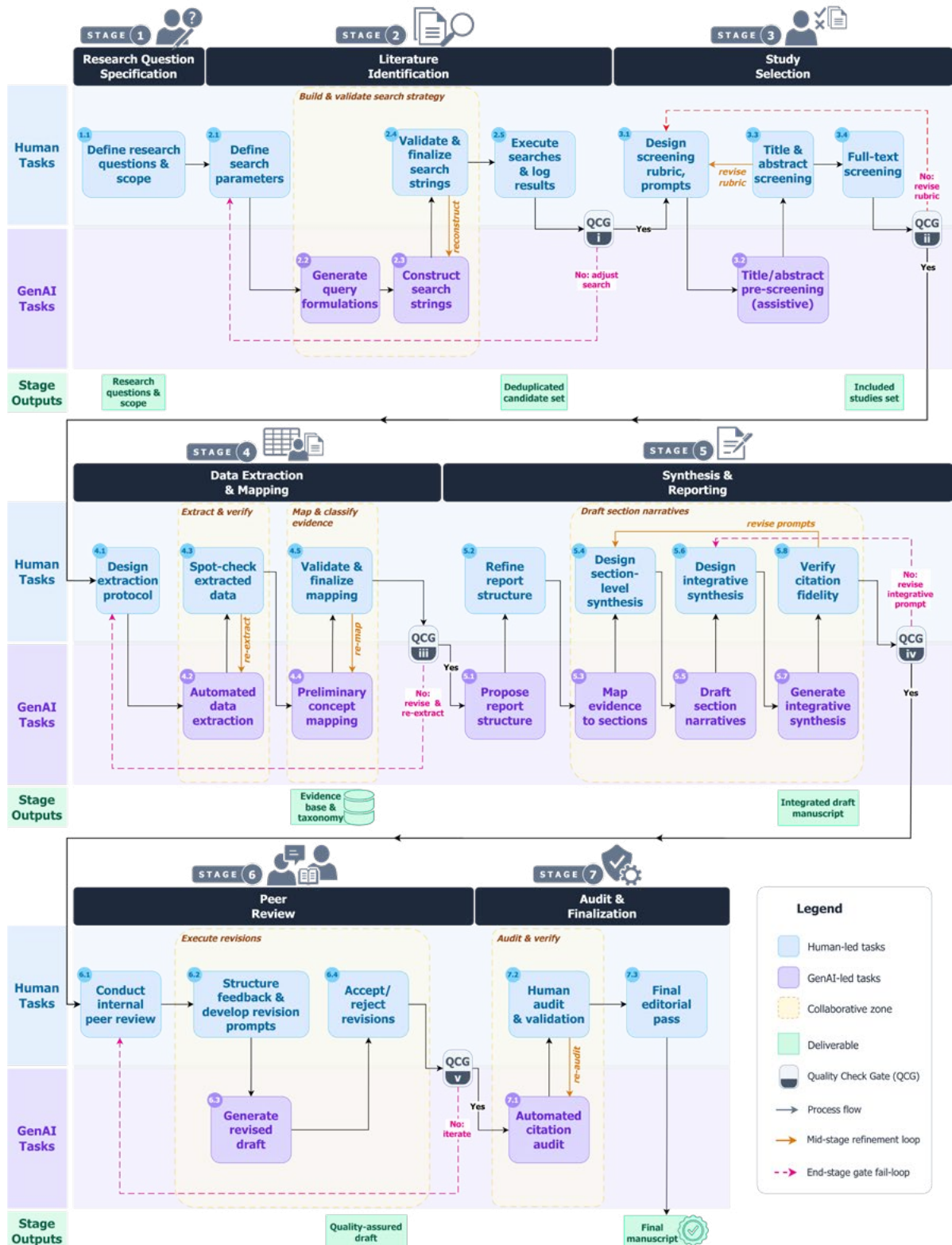
The review followed the seven-stage human-in-the-loop workflow shown in **Figure 1**, beginning with human-led specification of the research question and scope (Stage 1, Task 1.1). Across all stages, LLM tools supported defined tasks under human-designed prompts; two human reviewers retained authority 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.

Eligibility criteria

Sources were eligible if they presented empirical evidence — quantitative, qualitative, or mixed — or detailed case-study material linking M&E systems strengthening interventions to changes in the demand for, supply of, or use of evaluative evidence. Eligible source types included peer-reviewed journal articles; institutional and multilateral evaluation reports; case-study compilations and country-impact studies; doctoral theses; and learning products from regional evaluation partnerships. No a priori date or language restriction was applied; included sources were entirely English-language. At full-text review, sources were excluded for not establishing an evidence link between strengthening interventions and demand/supply/use outcomes, addressing only descriptive system characteristics, 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) present empirical evidence or detailed case-study material linking strengthening interventions to demand/supply/use outcomes, rather than describing strengthening interventions without an outcome link; (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).

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



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). Scopus and Web of Science Core Collection provided peer-reviewed coverage, with Google Scholar searched as a complementary academic search engine. Five institutional repositories were systematically searched: the World Bank Open Knowledge Repository, the UNEG Publications Repository, the 3ie Development Evidence Portal, the OECD iLibrary, and the institutional publication channels of regional evaluation partnerships active in this review's source base, including CLEAR Anglophone Africa and Twende Mbele. Given the prominence of African country cases in the strengthening-impact evidence base and the frequency with which such case evidence appears in institutional channels rather than through general database indexing, targeted searches of these institutional channels were conducted as a structured supplementary stream. Coverage was further extended through backward and forward citation chasing and expert recommendation.

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 57 identified through backward and forward citation searching and expert recommendation and known-source inclusion (**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. Twenty records advanced to full-text screening (Task 3.4); four could not be retrieved, leaving 16 assessed, of which four were excluded with coded reasons. In parallel, 55 of the 57 supplementary-stream records were retrieved and assessed (two were not retrievable), and 50 were excluded with coded reasons. A total of 17 sources were included in the review.

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.

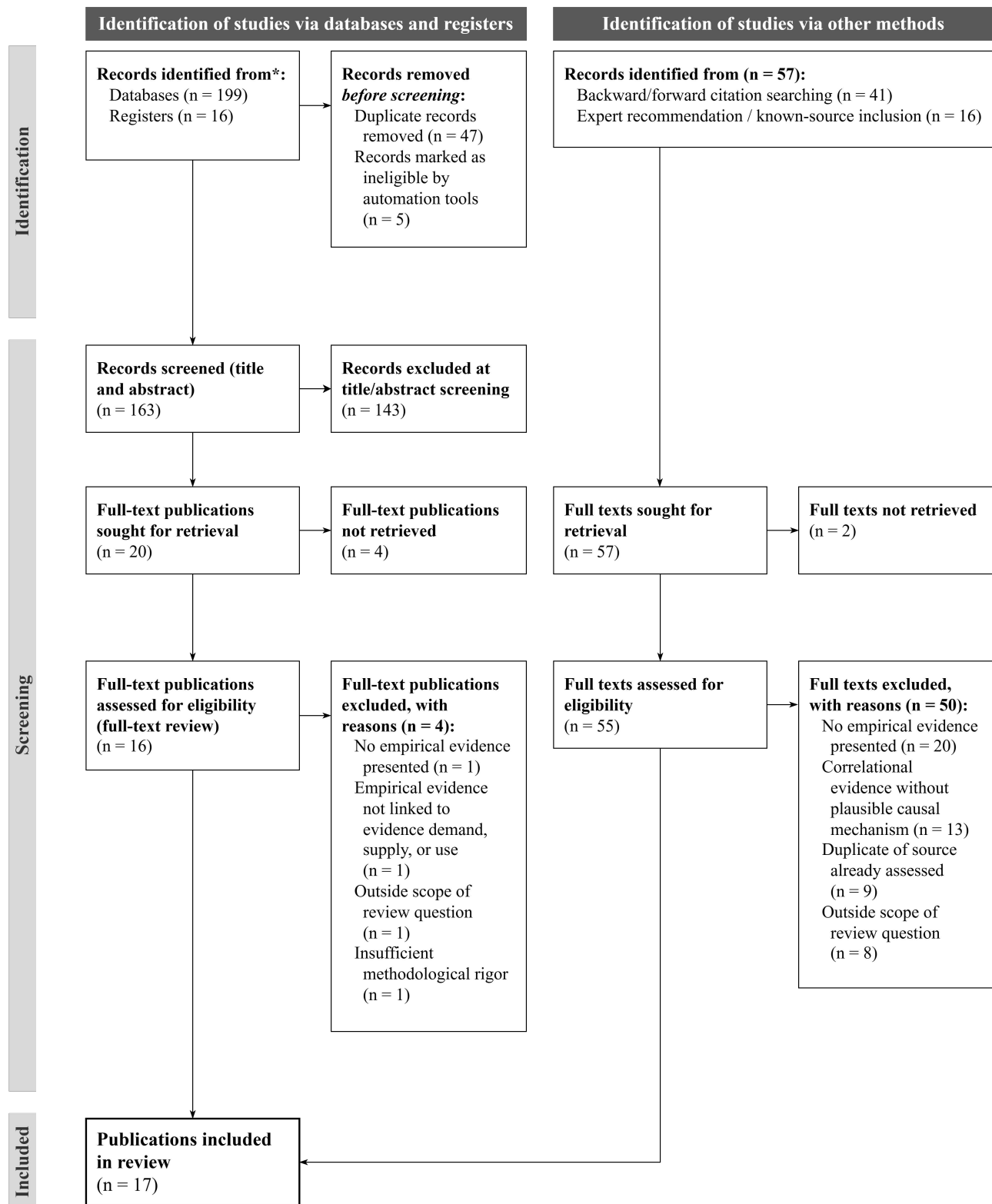
Data extraction and mapping

Data extraction and mapping (Stage 4) proceeded under a human-designed extraction protocol (Task 4.1). A structured codebook captured bibliographic metadata; the strengthening intervention at issue; the demand/supply/use outcome reported; the design and rigor of the underlying evidence; the country or regional context; and the strength and directionality of the reported link. 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 evidence-link statements and outcome attributions 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), organizing the

included sources along the demand/supply/use outcome dimension and across the case material, motivating an analytical structure spanning empirical evidence on demand, supply, and use, followed by case examples illustrating M&E systems strengthening impact.

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.

Figure 2. PRISMA Flow Diagram



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 four-part analytical structure spanning empirical evidence on demand, supply, and use of evaluative evidence, followed by case examples illustrating M&E systems strengthening impact. 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.

Empirical Evidence on Demand, Supply, and Use

Empirical Evidence on Demand

Demand for evaluative evidence refers to the degree to which government officials, development partners, and civil-society actors actively seek out and value evaluation findings as inputs to decision-making. Rather than arising solely from individual requests for data, demand appears to be heavily shaped by institutional incentives and high-level organizational mandates that determine whether M&E evidence is actively sought and utilized. Chirau et al. (2022, p. 10) note that the nature of incen-

tives for a country depends on how it envisages using M&E information, while Kanyamuna (2021, p. 93) emphasizes that demand is strengthened when high offices such as the presidency actively champion results-based management, establishing M&E as an organizational priority rather than a discretionary activity. Where such institutional drivers are absent, demand tends to remain low, sporadic, or donor-dependent—a pattern well documented in sub-Saharan Africa.

Empirical studies from Ghana and Zambia illustrate this pattern. Adams et al. (2013, pp. 19, 27) report that in Ghana, actual demand for evaluation from the executive has been very limited, with evaluations being donor-led and evaluation accounting for only 3% of M&E expenditure in 2009 and 1% in 2010. This low demand stemmed partly from limited awareness among executives and parliament that evaluation can provide relevant evidence. Similarly, Mulenga and Porter (2013, p. 32) observe that in Zambia there is very little actual demand from government principals outside of development partners, who serve as the main commissioners of evaluations.

Nevertheless, several authors argue that political will and top-down endorsement of M&E can generate robust domestic demand. Uganda exemplifies how high-level offices within the executive branch can champion evaluation to track development outcomes and inform budget performance. Niringiye (2018, p. 11) reports that the Office of the Prime Minister collates information from other departments and produces bi-annual and annual sector performance reports, with annual Cabinet retreats reviewing government performance and issuing recommendations to inform budgeting processes. The Government Evaluation Facility operates under the Office of the Prime Minister to provide technical support for evaluations (Niringiye, 2018, p. 11). Despite these institutional arrangements, monitoring remains the dominant feature of Uganda's national system, and the Office of the Prime Minister's mandate to oversee service delivery reforms through institutionalized M&E and results-based budgeting has not translated into strong demand from decision-makers for deeper evaluative evidence (Niringiye, 2018, pp. 10–13).

Civil-society engagement can also boost demand by advocating for greater transparency around public expenditures and by providing constructive input for the budget process, as civil-society organizations function as checks and balances on government (Kanyamuna, 2021, pp. 94–95). Even so, fragmented stakeholder coordination remains a barrier; Kanyamuna (2021, p. 83) documents how Zambian government agencies and civil-society organizations "remained fragmented and uncoordinated in the manner they demanded for M&E information," a lack of unified arrangement that limits the collective pressure required to incentivize the supply and use of evaluative evidence.

Beyond sub-Saharan Africa, comparable patterns of donor-initiated demand evolving toward institutionalized domestic demand emerge in other regions. In Sri Lanka, donor funding from the Asian Development Bank and UNDP initially drove M&E improvements during the 1990s, strengthening the Post Evaluation System and the Results Based Monitoring and Evaluation System. Subsequently, the government adopted an MfDR framework that institutionalized high-level political commitment: the Cabinet of Ministers directed the Ministry of Plan Implementation to monitor all major development projects and submit quarterly progress reports, while the President chairs regular progress review meetings to assess development projects, programs, and sectors (Sivagnanasothy, 2009, p. 3). In the Caribbean, the Evidence Informed Decision Making Network (EvIDeNCe) utilized a portfolio of interventions—including evidence briefs, stakeholder dialogues, and training—to strengthen research uptake; following these interventions, participants reported high self-reported "intention-to-use" scores of 6.4 for dialogues and 6.3 for training on a 7-point scale (Yearwood, 2018, pp. 1, 5). While these self-reported intentions do not guarantee behavioral change, they signal a shift in attitudes and should be interpreted alongside objective uptake measures when available. Taken together, the evidence across regions suggests that demand for evaluative evidence is most likely to strengthen when institutional mandates, political championship, and coordinated stakeholder engagement converge.

Empirical Evidence on Supply

Whereas demand reflects whether actors want evaluative evidence, supply concerns who produces it, how it is generated, and whether it meets quality standards. Organizational resources, technical expertise, and the presence of standardized procedures all shape supply at systemic—rather than merely individual—levels, requiring country-based programmes that institutionalize national policies, coordinate organizational resources via strategic planning, and leverage the technical expertise of local institutions (Chirau et al., 2020, pp. 1–2).

A recurring concern is that in many low- and middle-income settings, national or local systems lack the resources to conduct robust evaluations, leading donors or international consultants to fill the gap. Adams et al. (2013, p. 19) report that in Ghana, evaluation accounted for only 1–3% of M&E expenditure in 2009–2010, and that most evaluations were conducted by or on behalf of development partners, with Ghanaian evaluators' involvement remaining "rather symbolic" (p. 17). In Malawi, Kumwenda and Latib (2013, p. ii) observe that evaluation studies are "seen as a tool applied to support accountability for management or political leadership" rather than for internal policy development, and that international consultants usually lead major evidence-gathering exercises while local consultants play a support role. Similarly, Mulenga and Porter (2013, p. 32) observe that in Zambia, development partners are the main managers and commissioners of evaluations, with very little actual demand from government principals. This donor-centric pattern is reinforced when donors maintain parallel M&E systems to satisfy their own reporting requirements, a practice that undermines national ownership and weakens the development of internal, whole-of-government institutional arrangements (Kanyamuna et al., 2020, pp. 26–27).

Yet a lack of local resources is only part of the story; where evaluations are commissioned primarily for donor accountability, even available national expertise may be sidelined. South Africa's DPME offers a contrasting model by commissioning high-quality evidence on domestic programs, exemplified by its large-scale evaluation of the national Early Childhood Development initiative, and leveraging an increasing supply of evaluation evidence from local universities and consultancies (Stewart et al., 2018, p. 242). Uganda has likewise adopted M&E frameworks—including policies, evaluation guidelines, and evaluation plans—that create favorable conditions for institutionalizing both monitoring and evaluation functions and signal that evaluation is not solely a donor-driven activity (Chirau et al., 2022, p. 4).

Targeted professional-development initiatives can complement, but not substitute for, these system-level reforms. A four-day short-course training in Kenya targeting mixed methods research, systematic review, and M&E principles resulted in statistically significant knowledge gains among participants (Wao et al., 2017, pp. 1, 4). Because these interventions operate primarily at the individual level, they are most effective when paired with institutional incentives that enable newly trained staff to apply their skills within government structures.

Despite such progress, challenges persist when evaluation models prioritize rigorous impact techniques, such as randomized controlled trials, which may be ill-suited to the diverse information needs of government ministries and incompatible with rapid decision-making cycles driven by political calendars (Witter et al., 2017, pp. 2, 10). Fragmented data-management systems, limited sharing of evaluation reports, and underutilized local knowledge further restrict supply. Ethiopia's Compassionate, Respectful, and Caring (CRC) program demonstrates how the supply of evaluative evidence falters when a lack of standardized indicators and limited resources is compounded by the failure to integrate program-specific M&E into the main institutional systems (Minyihun et al., 2022, pp. 65, 71). These inconsistencies stemmed in part from separate planning structures that prevented the CRC's M&E functions from integrating with the broader institutional system, highlighting coordination—not merely resourcing—as a critical bottleneck for sustaining evaluative evidence supply.

Empirical Evidence on Use

Use of evaluative evidence concerns the extent to which findings influence policies, budgets, and ongoing programmatic decisions. Studies from multiple countries indicate that even when evaluations are conducted, applying results in practice is far from guaranteed—and that the factors enabling or constraining use are shaped by political contexts, institutional structures, and the nature of evaluation engagement itself.

Political incentives, patronage networks, and vested interests often determine which recommendations are adopted and which are ignored. Adams et al. (2013, p. 29) report that in Ghana, despite numerous donor-commissioned evaluations, utilization by the executives, legislature, and government agents has been limited. This limited uptake reflects the fact that evaluation work in government ministries was predominantly donor-led—designed to comply with donors' requirements rather than driven by domestic demand—and without deliberate efforts to ensure Ghanaian involvement and ownership, such evaluations were unlikely to be used by the country (Adams et al., 2013, pp. 17, 19, 27). Mulenga and Porter (2013, p. 14) similarly observe that in Zambia, transforming latent demand into consistent actual demand is challenging as a variety of other considerations come into play, with the political context often determining whether evaluation findings influence policy decisions. Research on the Farmer Input Subsidies Programme (FISP), for instance, found that the scheme was disproportionately allocated to better-off households and failed to significantly reduce poverty (Mulenga & Porter, 2013, p. 17). Despite these findings, reform efforts were constrained by the interplay of powerful interest groups, leading the government to retreat from proposed subsidy removals and reaffirm support for the program (Mulenga & Porter, 2013, pp. 16–17).

Nonetheless, there are documented successes where institutional frameworks have created more favorable conditions for evidence use. Chirau et al. (2022, p. 4) note that Uganda and South Africa are frequently cited in the literature for adopting M&E frameworks—including policies, guidelines, and plans—that create favorable environments for institutionalizing monitoring and evaluation and for measuring development results. In South Africa, the DPME facilitates the integration of evaluation findings into Cabinet-level policy discussions (Goldman et al., 2015, pp. 1, 6), with evidence showing that evaluations of programs such as Early Childhood Development and rural land reform are reported to have contributed to revised policies and integrated implementation strategies at the highest levels of government (Goldman et al., 2015, p. 5). In Uganda, the Government Evaluation Facility (GEF) has commissioned over 20 evaluations designed to inform decision-making and policy formulation (Niringiye, 2018, pp. 10, 22), although their influence on broad reforms is often limited by weak demand from high-level political bodies like the Cabinet and Parliament (Niringiye, 2018, p. 30). Sri Lanka offers a further example of high-level institutionalization: the President chairs regular progress review meetings coordinated by the Ministry of Plan Implementation as the national focal point for monitoring and evaluation, providing a high-level forum to assess development results with top-level political commitment (Sivagnanasothy, 2009, pp. 2–3).

Even in these more favorable contexts, however, evaluation impacts tend to be modest and incremental rather than transformational. Witter et al. (2017, p. 8) report that the implications of evaluations under the 3DE programme were largely managerial, rather than implying larger changes in programming or resource allocation.

One emerging strategy for strengthening use is the increasing practice of embedding researchers within policy teams to produce more relevant, usable evidence (Williamson et al., 2019, pp. 4, 7). Such partnerships are most effective when they maintain ongoing engagement, including presenting findings as they emerge to ensure research remains aligned with shifting real-world priorities (Williamson et al., 2019, pp. 7, 9). In many settings, however, the structural integration of evaluations into existing systems, such as budgeting or strategic planning, remains weak; in Zambia, despite regular reporting of inputs, activities, and outputs, there is limited evidence of M&E information being fully utilized to inform planning, budgeting, and policy-making processes (Kanyamuna, 2021, p. 93).

Case Examples of M&E Systems Strengthening Impact

The following country illustrations ground the empirical patterns discussed above in concrete experiences, clarifying how reforms unfold in practice and under what conditions they influence the demand, supply, and use of evaluative evidence.

In Ghana, the Joint Assessment for Strengthening M&E and Statistics (JASMES) sought to integrate evaluation into policy cycles. However, domestic funding for evaluation remained extremely low — accounting for only 3% of M&E expenditure in 2009 and 1% in 2010, and this, alongside disruptions to institutional continuity accompanying changes in government and overlapping mandates among oversight bodies, hindered the development of a domestically owned evaluation function (Adams et al., 2013, pp. 19, 27). This underinvestment coincided with latent demand from the executive and parliament, attributed partly to a lack of awareness that evaluation can provide relevant evidence (Adams et al., 2013, p. 27). Adams et al. also note that an evaluation of an improved cook stove programme, conducted on behalf of Plan Ghana by external researchers without substantive involvement of Ghanaian evaluators, produced valuable findings but remained largely unknown domestically—illustrating how the absence of local ownership can limit the dissemination and uptake of evaluation results even when the evidence itself is sound (Adams et al., 2013, pp. 18–19).

Zambia offers multiple examples of the complexities of M&E reform. Efforts to establish a Whole-of-Government M&E System (WoGM&ES) reflect a push to create an integrated framework of M&E practices, principles, and standards across government institutional structures, functioning as an apex-level system that draws from component systems to deliver essential M&E products (Kanyamuna, 2021, pp. 83–84). Within this broader reform context, evaluation findings from the Sixth National Development Plan were used to inform the design and implementation of the Seventh National Development Plan (Chirau et al., 2022, p. 4). Yet many evaluations in Zambia remain donor-centric, a factor that frequently limits domestic ownership and the subsequent utilization of evaluative evidence (Chirau et al., 2022, pp. 3, 11). This pattern of external orientation is further illustrated by the Zambian health sector's Annual Progress Reports (APRs), which are compiled primarily to satisfy the requirements of donors and central ministries rather than for internal utility, illustrating a prioritization of upward accountability over local learning and citizen-centered needs (Kanyamuna et al., 2018, pp. 80–81). This external focus reduces incentives to invest in the analytical depth required for local priority setting and internal policy improvement (Kanyamuna et al., 2018, pp. 80, 83). Beyond issues of demand and supply, the Farmer Input Subsidies Programme case serves as a cautionary example of how actual implementation can be negotiated between interest groups to the benefit of the powerful rather than the public interest, demonstrating that loyalties to informal power networks can derail evidence-based reforms even when evaluations highlight program failures (Mulenga & Porter, 2013, pp. ii, 12, 17).

In Malawi, the National Cash Transfer Programme underscores the role of external evaluations funded by international donors. Findings from these evaluations were used to improve operations and beneficiary identification policies, though the government only partially adopted recommendations, rejecting some based on political or infrastructure constraints (Kumwenda & Latib, 2013, pp. 11–12). The relative absence of a strong, government-led M&E policy structure meant that the impetus for evaluation remained externally motivated, and local capacity to conduct or utilize studies remained underdeveloped.

Moving beyond Africa, Sri Lanka's experience with the MfDR approach and a web-based electronic Project Monitoring System (ePMS) suggests that embedding evaluation in high-level political review can raise demand and expedite use. Progress review meetings are chaired by the President on a regular basis to assess the performance of ministries and institutions, while the Cabinet of Ministers directed the Ministry of Plan Implementation to monitor all major development projects and submit quarterly progress reports (Sivagnanasothy, 2009, p. 3). According to Sivagnanasothy, these high-level forums reflect strong top-level political commitment and sustained leadership, though the source was written

early in the reform process and evidence of downstream impact on policy outcomes was not yet available (Sivagnanasothy, 2009, pp. 3, 15). Yet the effectiveness of these mechanisms also depended on ongoing technical support from donors and an emerging local culture of accountability.

In the Caribbean, the EvIDeNCe Network addressed the research-policy gap by providing contextualized evidence briefs and hosting stakeholder dialogues, an approach designed to align the supply of synthesized research with the political and socioeconomic needs of policymakers (Yearwood, 2018, pp. 2–3). The emphasis on intention-to-use metrics, while not a direct measure of actual policy shifts, signals that such structured interventions can at least generate a stronger climate of openness to research findings.

Finally, South Africa's Department of Planning, Monitoring, and Evaluation stands out as a model for institutionalizing evaluations at scale. The DPME operates within a national framework that includes partner departments and the National School of Government, the latter of which provides mandated training for public servants to build endogenous capacity for evidence-informed decision-making (Stewart et al., 2018, pp. 242, 251). These efforts are further reinforced by regional communities of practice, such as the Africa Evidence Network, which facilitate the relationship-building and knowledge brokerage necessary to bridge the gap between research producers and government users (Stewart et al., 2018, pp. 247, 250). Some findings have informed policy, although the reliance on informal communication channels and the sporadic adoption of recommendations in certain sectors demonstrate that systemic challenges remain.

Conclusion

Several recurring themes emerge across the empirical studies and case examples reviewed here, though the evidence base remains limited and geographically concentrated. Demand for evaluative evidence rises when political champions embed M&E requirements into budgeting and planning processes and when civil-society and legislative actors insist on transparency. Supply improves when governments formalize evaluation policies, invest in local expertise, and create repositories that facilitate access to high-quality studies. Evidence use, in turn, depends on aligning evaluation timelines with policy cycles, employing embedded researchers and knowledge brokers to translate findings, and ensuring that incentive structures—both political and bureaucratic—reward uptake. Even where these conditions are met, however, the evidence reviewed here suggests that evaluation impacts tend to be incremental and operational rather than transformative, reflecting the persistent difficulty of embedding evaluative reasoning into political systems shaped by competing interests and short decision-making horizons.

Case examples reinforce these points: Ghana and Malawi show how donor dependence limits ownership; Zambia highlights the significant role of political economy barriers; Sri Lanka and South Africa demonstrate that routine integration and high-level oversight can translate evaluations into action, even if the resulting changes are often managerial rather than structural. Across all cases, a common thread is that the transition from externally driven M&E to domestically owned systems represents the most critical—and most difficult—threshold for sustaining evidence-informed governance.

Several gaps in the evidence base stand out. First, the reviewed literature offers limited insight into the causal mechanisms through which M&E system reforms translate into changes in evidence demand, supply, or use; most studies describe institutional arrangements and perceived outcomes but do not trace the pathways connecting them. Research that moves beyond describing institutional arrangements to examining how and why reforms lead (or fail to lead) to changes in evidence demand, supply, and use would be a valuable addition. Second, the evidence on how contextual factors—such as regime type, fiscal space, and civil-society density—mediate the impact of M&E reforms remains thin; the cases reviewed here are drawn predominantly from sub-Saharan Africa, with limited representa-

tion from other regions facing distinct governance challenges. Third, while the literature consistently notes the importance of transitioning from donor-driven to domestically owned M&E systems, there is little documented evidence on what this transition looks like over time, what enables or impedes it, and how sustainability can be assessed. Finally, the reviewed studies say relatively little about the role of data infrastructure and digital systems in shaping M&E capacity, an area of growing practical relevance. Addressing these gaps would strengthen the empirical foundation available to practitioners and policymakers seeking to design M&E reforms suited to their institutional contexts.

References

- Adams, S., Amoatey, C., Taabazuing, J., & Feinstein, O. (2013). *Study on the demand for and supply of evaluation in Ghana*. Centre for Learning on Evaluation and Results – Anglophone Africa (CLEAR-AA), University of the Witwatersrand.
- 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>
- Chirau, T., Masilela, B., & Dlakavu, A. (2020). *Country-owned, country-led: Strengthening African M&E systems through country-based programmes*. Centre for Learning on Evaluation and Results – Anglophone Africa (CLEAR-AA).
- Chirau, T., Dlakavu, A., & Masilela, B. (2022). Strengthening Anglophone Africa M&E systems: A CLEAR-AA perspective on guiding principles, challenges and emerging lessons. *African Evaluation Journal*, 10(1), a601. <https://doi.org/10.4102/aej.v10i1.601>
- Goldman, I., Mathe, J. E., Jacob, C., Hercules, A., Amisi, M., Buthelezi, T., Narsee, H., Ntakumba, S., & Sadan, M. (2015). Developing South Africa's national evaluation policy and system: First lessons learned. *African Evaluation Journal*, 3(1), a107. <https://doi.org/10.4102/aej.v3i1.107>
- Kanyamuna, V. (2021). Towards building a functional whole-of-government monitoring and evaluation system for Zambia: The demand side. *World Journal of Social Sciences and Humanities*, 7(3), 83–105. <https://doi.org/10.12691/wjssh-7-3-2>
- Kanyamuna, V., Aurick, M., Emily, N., Chinyama, M., & Ackson, M. (2018). An assessment of the demand-side of the monitoring and evaluation system of the health sector in Zambia. *World Journal of Social Sciences and Humanities*, 4(2), 75–86. <https://doi.org/10.12691/wjssh-4-2-2>
- Kanyamuna, V., Phiri, M., Kanenga, H., & Mulonda, M. (2020). Role of actors outside government in strengthening the country monitoring and evaluation system in Zambia. *World Journal of Social Sciences and Humanities*, 6(1), 22–29. <https://pubs.sciepub.com/wjssh/6/1/4/index.html>
- Kumwenda, H., & Latib, S. (2013). *Study on the demand for and supply of evaluation in Malawi*. Centre for Learning on Evaluation and Results – Anglophone Africa (CLEAR-AA), University of the Witwatersrand.
- Minyihun, A., Getnet, M., & Tilahun, B. (2022). How to strengthen the monitoring and evaluation system to improve compassionate, respectful, and caring services in Ethiopia: An implementation study. *Ethiopian Journal of Health and Biomedical Sciences*, 12(Special Issue), 65–76.
- Mulenga, O., & Porter, S. (2013). *Study on the demand for and supply of evaluation in Zambia*. Centre for Learning on Evaluation and Results – Anglophone Africa (CLEAR-AA), University of the Witwatersrand.
- Niringiye, A. (2018). *Diagnostic study on the supply and demand of evaluators in Uganda*. Twende Mbele; CLEAR-AA, University of the Witwatersrand.
- 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.), *JBIM Manual for Evidence Synthesis*. JBI. <https://doi.org/10.46658/JBIMES-20-12>

- Sivagnanasothy, V. (2009). National monitoring and evaluation system in Sri Lanka: Experiences, challenges and the way forward. Ministry of Plan Implementation, Sri Lanka.
- Stewart, R., Langer, L., Wildeman, R., Erasmus, Y., Maluwa, L. G., Jordaan, S., Lötter, D., Mitchell, J., & Motha, P. (2018). Building capacity for evidence-informed decision making: An example from South Africa. *Evidence & Policy*, 14(2), 241–258. <https://doi.org/10.1332/174426417X14890741484716>
- 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>
- Wao, H., Onyango, R., Kisio, E., Njatha, M., & Onyango, N. O. (2017). Strengthening capacity for monitoring and evaluation through short course training in Kenya. *African Evaluation Journal*, 5(1), a192. <https://doi.org/10.4102/aej.v5i1.192>
- Williamson, A., Tait, H., El Jardali, F., Wolfenden, L., Thackway, S., Stewart, J., O'Leary, L., & Dixon, J. (2019). How are evidence generation partnerships between researchers and policy-makers enacted in practice? A qualitative interview study. *Health Research Policy and Systems*, 17, 41. <https://doi.org/10.1186/s12961-019-0441-2>
- Witter, S., Kardan, A., Scott, M., Moore, L., & Shaxson, L. (2017). Generating demand for and use of evaluation evidence in government health ministries: Lessons from a pilot programme in Uganda and Zambia. *Health Research Policy and Systems*, 15, 86. <https://doi.org/10.1186/s12961-017-0250-4>
- Yearwood, A. C. (2018). Applying a logical theory of change for strengthening research uptake in policy: A case study of the Evidence Informed Decision Making Network of the Caribbean. *Revista Panamericana de Salud Pública*, 42, e91. <https://doi.org/10.26633/RPSP.2018.91>