

2.1

Conceptualizing M&E Systems Strengthening: A Scoping Review

Douglas Glandon
Guilherme Cunha Lima

Abstract

This scoping review explores how Monitoring and Evaluation (M&E) systems strengthening is defined and conceptualized in scholarly and practitioner literature. It responds to the growing emphasis among governments, development partners, and research institutions on building robust, sustainable M&E systems that support evidence-informed policymaking and accountability. Across the literature reviewed, M&E systems strengthening is predominantly characterized as a holistic, multi-level process encompassing individual, institutional, and systemic capacity building. Core themes include the integration of M&E into national policy and budgeting processes, the institutionalization of M&E through legal mandates and dedicated structures, and the creation of enabling environments supported by political will, financial investment, and stakeholder engagement. This synthesis follows a scoping-review protocol grounded in JBI guidance, with reporting per PRISMA-ScR.

Key conceptual frameworks emphasize capacity building across three interconnected levels – individual, institutional, and country (system) – with individual and institutional capacities grouped at the micro level and country-level capacity at the macro level. Institutionalization is identified as essential, transforming M&E from isolated technical efforts into embedded governance functions. The review also highlights the importance of aligning M&E with national priorities, avoiding donor-driven fragmentation, and building demand for evidence among policymakers and civil society.

Notable gaps include limited frameworks for assessing the long-term outcomes of M&E capacity-strengthening efforts and a persistent imbalance between monitoring and evaluation, with the latter often remaining peripheral. Emerging insights call for a shift toward outcome-based indicators, a stronger evaluation culture, and a balanced supply-demand dynamic for M&E information. Ultimately, the review underscores that strengthening M&E systems requires more than technical inputs – it demands strategic alignment, sustained political commitment, and structures that foster continuous learning and use of evidence.

Introduction

How are M&E systems strengthening approaches defined and conceptualized in the literature? This question has gained increasing prominence as governments, development partners, and research institutions emphasize the importance of robust and sustainable monitoring and evaluation (M&E) practices. Strengthening M&E systems goes beyond technical capacity alone: it involves fostering a culture of evaluation, ensuring political support, and embedding M&E within policy and governance processes to enhance accountability and evidence-based decision-making (CLEAR-AA, 2020, pp. 4, 15, 24). Given the heightened demand for reliable performance and outcomes data – especially in the context of global development goals and public-sector reforms – a clear understanding of M&E systems strengthening is vital to ensure interventions are both context-sensitive and durable.

This review is structured into four main sections. First, it offers a working definition of M&E systems strengthening, drawing on key sources that explicitly address the term. Next, it examines core conceptual frameworks and approaches often discussed in the literature, organized under three themes: (1) multi-level capacity building, (2) institutionalization and policy integration, and (3) enabling environments and contextual factors. Following that, the review explores gaps and emerging insights, shedding light on unresolved challenges and notable shifts in practice. Finally, it concludes with a synthesis of the main findings and reflections on future directions.

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 *definition and conceptualization* sub-topic at the decision-maker/unit level.

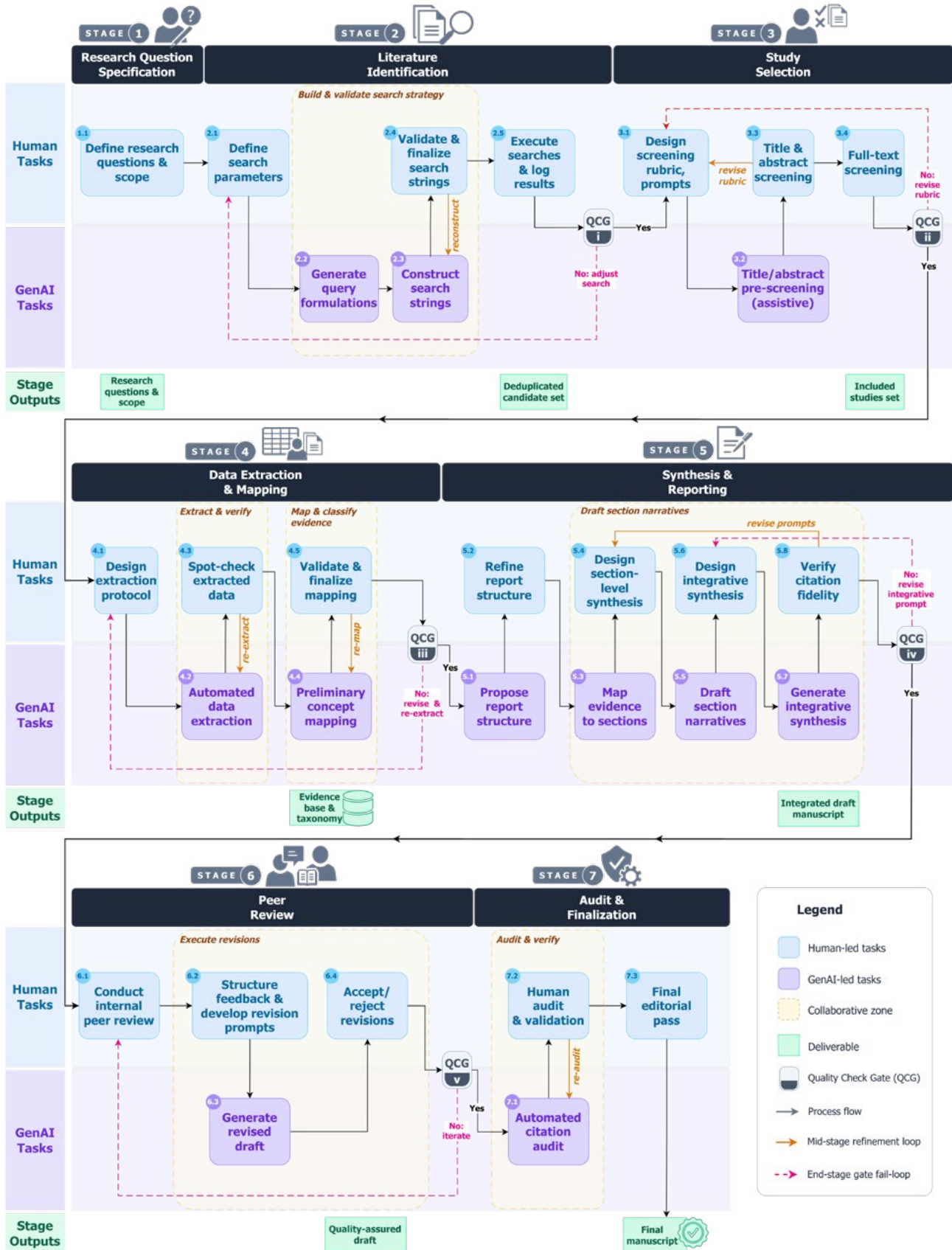
Methods

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 definitional and required breadth across heterogeneous, multidisciplinary material — peer-reviewed scholarship, institutional strategy and guidance documents, and ECD framework documents. Eligibility followed the JBI Population–Concept–Context (PCC) framing, with the review question and inclusion criteria anchored primarily to the concept of M&E systems strengthening as a construct. 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 defined, conceptualized, or offered an explicit framework for M&E systems strengthening — understood as deliberate interventions, approaches, or processes for building, reforming, or improving M&E systems and the capacities that sustain them. The construct was treated as encompassing its commonly used variants, including evaluation capacity development, evaluation system reform, M&E capacity building, and national evaluation capacity development. Eligibility was drawn tightly around sources engaging with how systems are built, changed, or strengthened, rather than sources merely describing standing system features or evaluation practice without an explicit strengthening lens. Eligible source types included peer-reviewed journal articles; institutional strategy and guidance documents; ECD frameworks; systematic and scoping reviews of ECD definitions; and doctoral theses. No a priori restriction was placed on earliest publication date, and no language restriction was applied. Searching concluded in June 2025, which constitutes the publication cutoff for inclusion. Included sources were entirely English-language. At full-text review, sources were excluded for not engaging with the strengthening construct as a construct, falling outside the review's scope, 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) offer an explicit and defended definition, conceptualization, or framework for M&E systems strengthening, rather than invoking the terminology in passing; (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 definitional sources on M&E systems strengthening are disproportionately published as institutional guidance, strategy, and framework documents rather than as journal articles, institutional registers and repositories were central to this review's search architecture. Scopus and Web of Science Core Collection provided peer-reviewed coverage, with Google Scholar searched as a complementary academic search engine. Institutional repositories systematically searched were the World Bank Open Knowledge Repository, the UNEG Publications Repository, the OECD iLibrary, Global Evaluation Initiative (GEI) network publications, and the institutional publication channels of professional evaluation associations (the American Evaluation Association and the International Development Evaluation Association). Targeted searches of the publication channels of regional evaluation partnerships active in ECD work — including CLEAR Anglo-phone Africa and Twende Mbele — 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.

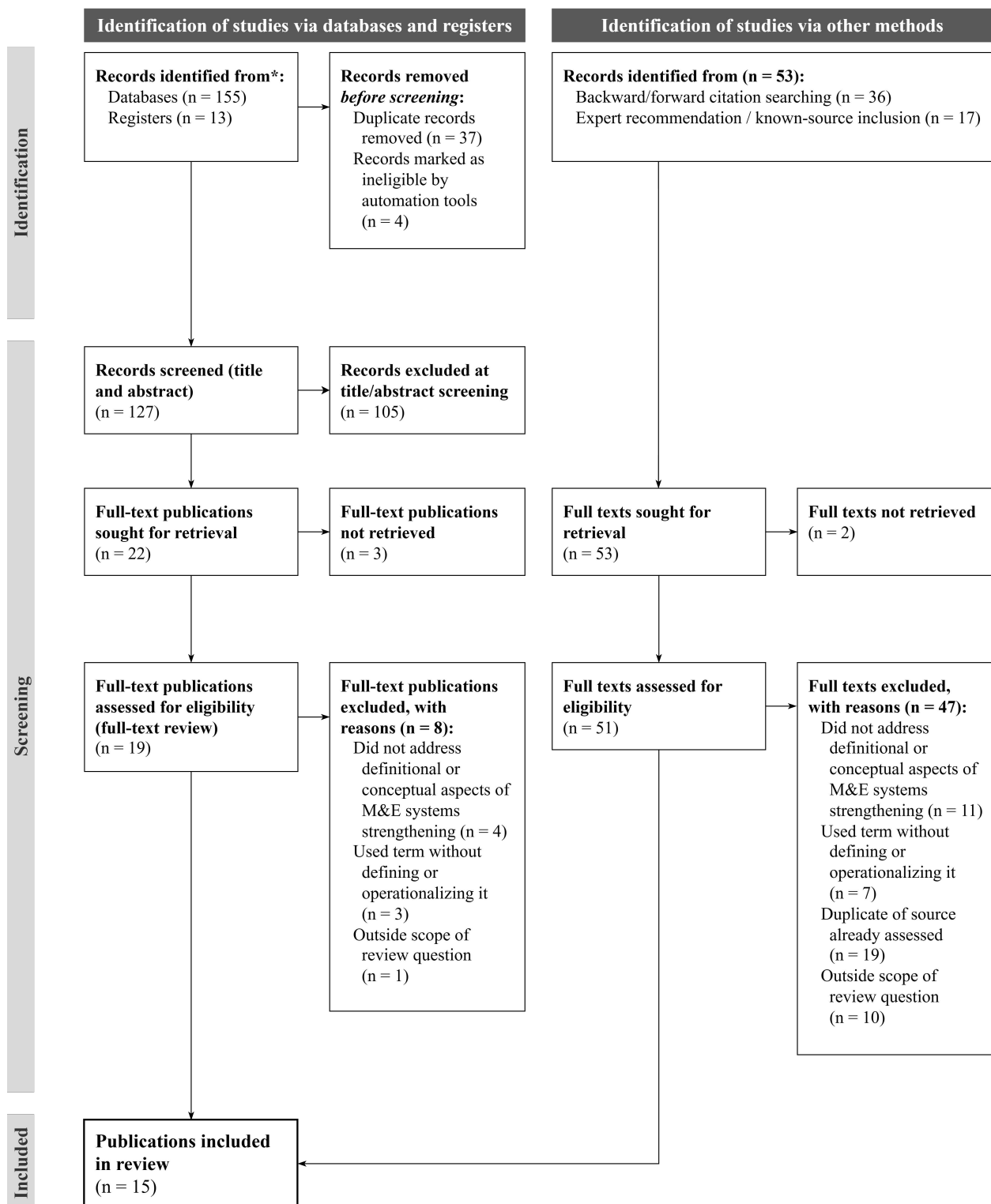
Source selection

Study selection (Stage 3) operated under a human-designed screening protocol (Task 3.1). A total of 168 records were identified through databases and registers (155 from databases; 13 from registers), with a further 53 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, 127 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-two records advanced to full-text screening (Task 3.4); three could not be re-trrieved, leaving 19 assessed, of which eight were excluded with coded reasons. In parallel, 51 of the 53 supplementary-stream records were retrieved and assessed (two were not retrievable), and 47 were excluded with coded reasons. A total of 15 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.

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 definition or conceptualization of strengthening offered; the framework advanced (where applicable); the components and dimensions identified; the underlying theoretical lineage; and any stated limitations. 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 definitional statements and framework 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), clustering the included sources along two axes: the definitional contribution offered (definition only, framework advanced, or boundary-condition clarification) and the conceptual lineage invoked.

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 three-part analytical structure comprising an opening definitional discussion, a section on the principal frameworks and approaches that have been advanced, and a closing section identifying gaps and emerging insights. 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.

Defining M&E Systems Strengthening

Although "M&E systems strengthening" can encompass various organizational, institutional, and national-level processes, the literature converges on a definition that emphasizes holistic capacity development, integration into public policy, and sustained demand for evidence. CLEAR-AA (2020, p. 2) defines M&E capacity strengthening as "the process through which individuals, institutions and countries develop, enhance and organise their M&E systems to effectively monitor the performance and evaluate development interventions at both state and institutional (non-state/state) levels." This definition underscores three essential dimensions: (a) the multi-tiered nature of capacity (individual, institutional, systemic), (b) the importance of a cohesive, organized approach to M&E, and (c) the end goal of facilitating better monitoring and evaluation of development outcomes.

Building on this multi-level framing, Chirau et al. (2022, p. 2) define a national M&E system, drawing on their earlier work (Chirau et al., 2020), as "the collection of M&E frameworks that guide the commissioning and undertaking of M&E and the individual and institutional M&E capacity to undertake and use M&E information to inform development planning, policymaking and decision-making." Their emphasis on both the conduct and use of M&E situates system strengthening as a process aimed at enhancing "supply and demand" — ensuring that credible, high-quality data are produced (supply) while fostering strong demand for such data among policymakers and stakeholders.

A closely related concept in the literature is the "institutionalization" of M&E, sometimes used synonymously with M&E systems strengthening. Makadzange (2022, p. 3), for instance, adopts the International Atlas of Evaluation Framework developed by Furubo, Rist, and Sandahl to assess the institutionalization of national M&E systems, establishing that for a system to be fully institutionalized, it must fulfill multiple domains including pervasive M&E practice across the public sector, institutional arrangements in parliament for M&E support, and policies or regulations that govern M&E practices. Conceptualized in this way, M&E systems strengthening moves beyond sporadic or donor-driven projects and becomes woven into the daily functioning of public-sector organizations. Lázaro (2015, p. 16), drawing on the same Furubo and Sandahl tradition, offers a complementary, process-oriented characterization: an evaluation system exists when evaluations are no longer commissioned on an ad hoc basis but through permanent arrangements that guarantee both the supply of evaluative information and its appropriate use. This framing foregrounds the dynamic, ongoing nature of institutionalization — it is not merely the presence of structures and regulations, but the establishment of routine, sustained evaluation practice integrated into decision-making cycles.

In a complementary vein, Goldman et al. (2018), adapting Holvoet and Renard's framework, identify six characteristics of an effective national evaluation system: 1) *policy*, including an evaluation plan, differentiation between monitoring and evaluation, evaluator autonomy and impartiality, and links to use, planning, and budgeting; 2) *methodology*, including indicator selection, evaluation prioritization, programme theory, methods, and data collection; 3) *organisation*, including coordination across central, statistical, line-ministry, decentralised, and project levels; 4) *capacity*, including recognition of capacity constraints and a capacity-building plan; 5) *participation of actors outside government*, including Parliament, civil society, and donors; and 6) the *quality and use*

of evaluation products and processes. This operational framework moves the definition of a strengthened M&E system beyond abstract principles toward a set of concrete institutional characteristics.

Key Conceptual Frameworks and Approaches

Multi-Level Capacity Building

A prominent theme across sources is the notion that capacity building must occur at multiple inter-connected levels—individual, institutional, and country (system)—to enable genuine and lasting M&E systems strengthening. Both CLEAR-AA (2020, p. 2) and Masvaure and Chirau (2020, p. 2) conceptualize M&E capacity as spanning these three levels, with individual and institutional capacities grouped at the "micro" level and country-level capacity at the "macro" level.

At the individual level, building evaluators' and policymakers' technical skills is critical; however, as Stewart (2015, pp. 549–550) argues, one-off trainings are not sufficient. Ongoing mentorship, specialist support, and the development of peer-support systems cultivate sustained changes in how research and evaluation evidence is used. Such sustained engagement facilitates "process use," where active participation in evaluation activities leads to changes in individual thinking and shifts in organizational culture, thereby fostering broader organizational learning (Alkin & King, 2016, pp. 568, 575–576).

At the institutional level, capacity-building initiatives often include forming dedicated M&E units, standardizing data collection processes, and clarifying roles and responsibilities. Several sources suggest these structural investments must be accompanied by attention to relational dynamics: Alkin and King (2017, pp. 445–446) highlight that evaluator credibility and the meaningful engagement of stakeholders are essential determinants of whether evaluation findings are utilized, noting that the development of rapport and a collaborative working relationship between evaluators and clients is key to maximizing use over the long term. Where these processes are well-managed and transparent, the likelihood of organizational ownership and continuous refinement of M&E grows.

At the system-wide level, the development of national M&E policy frameworks signals a government's commitment to evaluation, while integrating M&E into budgeting and development planning processes is consistently emphasized as important for sustainability (Chirau et al., 2022, pp. 7, 9). For example, Kanyamuna (2021) examines Zambia's whole-of-government monitoring and evaluation system, focusing on the demand side and proposing a model for strengthening it. Country-led approaches further emphasize that local contexts, rather than donor imperatives, should dictate the focus and structure of M&E system interventions; as Chirau et al. (2020, p. 2) note, each country programme should be "owned, focused and led by the country itself," with local in-country partners who understand the country's M&E context implementing interventions.

Institutionalization and Policy Integration

Another cornerstone of M&E systems strengthening lies in formalizing M&E within governance structures, often referred to as "institutionalization." This process involves embedding M&E within policy and planning systems so that government agencies not only collect data but also use it systematically in decision-making. Expanding on the framework introduced in the preceding section, Makadzange (2022, p. 3) conceptualizes "fully institutionalised" national M&E systems through a 12-domain framework requiring M&E to be pervasive across the public sector, reinforced by national laws or regulations, and managed by permanent oversight bodies with established structures for conducting and disseminating M&E work. However, the organizational form this institutionalization takes varies considerably across governance contexts. Lázaro (2015, pp. 18, 111), synthesizing Jacob's (2005) comparative analysis, identifies a typology of organizational configurations ranging from mo-

nopoly — where a single body dominates evaluation promotion and practice, typical of early-stage systems — through centralized pluralism, where one coordinating body sets norms while multiple agencies carry out evaluation, to fragmented concurrence, where no coordinating element exists and evaluation proceeds unevenly. This typology suggests that M&E systems strengthening need not follow a single institutional blueprint; rather, the appropriate configuration depends on a country's political structure, administrative traditions, and stage of evaluation development.

Integration with budgeting processes is repeatedly cited as crucial for achieving this institutionalization. Chirau et al. (2022, p. 9) identify the poor integration of planning, budgeting, and M&E as a common finding across multiple countries, while Kanyamuna et al. (2018, p. 81) argue that linking M&E to the budget process creates powerful incentives for effective M&E across public agencies. When M&E data inform resource allocations, it provides a tangible incentive for policymakers to value credible, timely information. In contrast, where M&E remains a peripheral activity—requested primarily by donors or external partners—its potential impact on policy decisions is diminished. By linking M&E to planning and budgeting processes, governments can work toward building what Kanyamuna (2021) calls a "functional whole-of-government monitoring and evaluation system," signaling that evaluation and monitoring are integral components of evidence-based governance rather than isolated tasks.

Policy integration also involves creating national M&E frameworks or strategies that guide government institutions (CLEAR-AA, 2020, p. 12). Such frameworks typically outline standards, reporting structures, and capacity-building plans. However, the content and orientation of these frameworks matter as much as their existence. Masvaure and Fish (2022) identify a significant tension in how M&E is practiced, noting that it is frequently treated as a compliance or audit function focused on checking for errors rather than fostering institutional learning. Consequently, they argue that capacity-strengthening efforts must transcend the monitoring of technical skills to prioritize deeper evaluation capacities that enable countries to meet their strategic development goals. Lázaro's (2015, pp. 66–68) comparative analysis of evaluation institutionalization in Europe and Latin America underscores this point: in cases where M&E obligations were introduced through top-down regulation — particularly in Southern European contexts — they were frequently assimilated into existing control and compliance functions, generating passive ritualistic adoption rather than genuine learning. The implication is that legal mandates for M&E, while necessary, are insufficient if they operate within administrative cultures unreceptive to the reflexive, learning-oriented use of evaluative evidence. When laws and policies mandate not only data collection but also structured evaluation of outcomes, a more genuine learning culture can emerge—helping shift the paradigm from accountability alone to both accountability and continuous improvement.

Enabling Environments and Contextual Factors

The success of M&E systems strengthening depends on more than technical or policy mechanisms; it also hinges on the broader political, cultural, and financial context. Drawing on Segone (2009), the CLEAR-AA (2020) report notes that "an enabling environment which includes the appropriate policies, legislation, power relations and social norms, is a prerequisite for strengthening M&E capacity at individual, institutional and system level" (p. 2).

Political buy-in is particularly critical. Chirau et al. (2022, p. 7) argue that securing political and administrative buy-in is crucial for making M&E a valued governance practice, and that high-ranking leaders—commonly referred to as "M&E champions"—are strategically positioned to promote the institutionalization of M&E in planning, budgeting, and decision-making. Conversely, weak demand for evidence, frequent leadership changes, or minimal accountability structures can undermine even well-designed initiatives. Lázaro (2015, pp. 20–21) offers a more granular taxonomy of contextual determinants, drawn from comparative analysis across European and Latin American systems. They argue that evaluation's ability to take root depends on factors far broader than mere political will. These include the strength of scientific traditions, the hybridization of academic and administrative cultures,

and the nature of the political system; they notes that majoritarian systems are often more hospitable to evaluation than consociational ones based on power-sharing. Furthermore, exogenous drivers like EU structural fund obligations or donor requirements play a pivotal role. Lázaro (2015, p. 107) concludes that success relies less on technical or institutional shifts and more on a fundamental cultural reorientation toward reflexive governance, continuous learning, and a tolerance for critical findings.

Resource allocation further shapes the enabling environment. Kanyamuna et al. (2018, pp. 80–81) emphasize that for a sector M&E function to reach sustainable levels, it requires sufficient funding to overcome chronic deficits, address high levels of understaffing and staff attrition, and improve poor infrastructural capacity. Without sufficient budgets to hire qualified staff, establish reliable data management systems, or fund evaluation studies, M&E efforts risk remaining superficial. Donor-driven fragmentation compounds this challenge: when multiple agencies fund unsystematic, short-term M&E activities that are donor-centric rather than country-led, the result frequently fails to strengthen long-term institutional M&E architecture (Chirau et al., 2020, p. 3).

Finally, stakeholder engagement and a strong evaluation culture are widely cited as enabling factors. Professional evaluation associations facilitate shared learning through dialogue on ethics and standards, while academic programs in M&E expand the supply of trained experts required to sustain enabling environments for evaluation (Makadzange, 2022, pp. 3, 7–8). An absence of such communities can exacerbate capacity gaps and create reliance on external consultants.

Gaps and Emerging Insights

A consistent gap in the literature concerns measuring the outcomes of M&E capacity-building activities. Both Masvaure and Chirau (2020, p. 2) and CLEAR-AA (2020, pp. v–vi) observe that measurement of M&E capacity-strengthening initiatives remains nascent and largely confined to inputs, processes, and outputs at the micro level, with few concrete frameworks for assessing whether such strategies have strengthened systems in sustained ways. Masvaure and Fish (2022, p. 6) highlight a critical need for outcome-based indicators that track changes in how data are generated, used, and integrated into policy decisions. This measurement gap is compounded by a broader imbalance between monitoring and evaluation. Chirau et al. (2022, p. 5) observe that M&E infrastructure in African countries remains biased toward producing monitoring data, with evaluation remaining "on the periphery," and that the overemphasis on monitoring has led to excessive attention on inputs and activities without attention to outcomes and impact. This imbalance can limit organizational learning and hamper the broader objective of evidence-informed policymaking. Strengthening the evaluation function requires stronger institutional and political commitment, not just technical know-how.

Lamarque (2023, p. 163) observes that most frameworks for assessing the institutionalization of evaluation apply primarily to central government and do not account for subnational entities, a significant omission given that the majority of public policies are implemented across multiple levels of government. Beyond gaps in evaluative scope, many sources also point to fragmentation in M&E systems, especially in low- and middle-income countries where donor-funded projects proliferate. Because many M&E systems are currently donor-driven rather than endogenous, they risk relying on discrete interventions that fail to achieve the systemic results necessary to ensure that findings are actually used to inform decisions (Chirau et al., 2020, pp. 1, 3). The shift from short trainings to holistic, long-term strategies – encompassing policy integration, alignment with national priorities, and robust local leadership – is thus an emerging trend in how M&E systems strengthening is conceptualized in the literature.

A further central insight concerns the importance of balancing supply and demand for M&E information. Chirau et al. (2022, p. 5) call for an equilibrium between the supply and demand of M&E information across country M&E ecosystems, while Kanyamuna (2021, p. 84) argues that a functional national M&E system requires both the capacity to supply M&E information and the capacity to demand

and use it. Strengthening the "supply side" involves training evaluators, creating data infrastructure, and establishing robust metrics. Yet the relationship between supply and demand is not merely parallel. Kanyamuna et al. (2018) note that the demand side can affect the supply side of an M&E system and that, when demand and utilization of information by key stakeholders are low, the M&E function may come to be regarded as a bureaucratic burden and compliance with system procedures may deteriorate — illustrating how demand-side weakness can undermine the functioning of the system. Equally essential is energizing the "demand side" by ensuring policymakers and stakeholders actively seek out evaluation findings, value their utility, and incorporate them into decisions. This balance underscores that M&E systems are not merely technical apparatuses but social systems requiring buy-in from diverse actors.

Conclusion

In direct response to the research question – "How are M&E systems strengthening approaches defined and conceptualized in the literature?" – the literature broadly characterizes M&E systems strengthening as a holistic, multi-level, and contextually embedded undertaking. Sources such as CLEAR-AA (2020, p. 2) and Chirau et al. (2022, p. 7) emphasize capacity building across individual, institutional, and country levels, while also highlighting the central role of institutionalizing M&E within governance structures to ensure it becomes a valued and embedded practice. Multiple definitions converge on the notion that strengthened M&E systems are those which integrate seamlessly with budgeting, policy formulation, and implementation, underpinned by political support and robust legal frameworks. An enabling environment – encompassing political will, financial investment, and stakeholder collaboration – is presented as supporting the continued development of these systems and fostering a culture of evidence use. Thus, M&E strengthening is not merely about technical skills or data collection but about creating sustainable structures, policies, and demand conditions that enable consistent, meaningful use of monitoring and evaluation findings.

These findings suggest several implications for policymakers, practitioners, and researchers. Policy integration emerges as a linchpin, suggesting that without legislative mandates and clear policy linkages, M&E systems may remain peripheral. Equally, capacity measurement stands out as a critical gap, pointing to the need for innovative frameworks and indicators to assess long-term effectiveness beyond counting trainings or short-term outputs. Future research could develop more robust methodologies for assessing how M&E influences government decisions at various levels, as well as approaches for adapting capacity-building models to local contexts. Strengthening the demand side of M&E – ensuring that policymakers, civil society, and citizens value and actively seek evaluative evidence – also warrants further exploration, given that a balanced supply-demand dynamic appears key to sustainability. Such efforts can help refine the conceptual understanding of M&E systems strengthening and contribute to more evidence-informed governance in diverse settings.

References

- Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use. *American Journal of Evaluation*, 37(4), 568–579. <https://doi.org/10.1177/1098214016665164>
- Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation*, 38(3), 434–450. <https://doi.org/10.1177/1098214017717015>
- 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>
- Centers for Learning on Evaluation and Results – Anglophone Africa (CLEAR-AA). (2020). *M&E capacity-strengthening approaches and measurement in Africa* [Report]. <https://unfccc.int/sites/default/files/resource/Clear%20ME%20Capacity%20report-V4b.pdf>
- 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>
- 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).
- Goldman, I., Byamugisha, A., Gounou, A., Smith, L. R., Ntakumba, S., Lubanga, T., Sossou, D., & Rot-Munstermann, K. (2018). The emergence of government evaluation systems in Africa: The case of Benin, Uganda and South Africa. *African Evaluation Journal*, 6(1), a253. <https://doi.org/10.4102/aej.v6i1.253>
- 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>
- Lamarque, D. (2023). Strengthening evaluation capacity in government: Why institutionalisation matters and why it is not sufficient. *Evaluation*, 29(2), 161–167. <https://doi.org/10.1177/13563890231166981>
- Lázaro, B. (2015). Comparative study on the institutionalisation of evaluation in Europe and Latin America. EUROsociAL Programme.
- Makadzange, P. F. (2022). A descriptive narrative on the current situation against the gold standards regarding institutionalisation of national monitoring and evaluation system for Botswana and Zimbabwe. *African Evaluation Journal*, 10(1), a578. <https://doi.org/10.4102/aej.v10i1.578>
- Masvaure, S., & Chirau, T. (2020). *M&E capacity-strengthening approaches and their measurement in Anglophone Africa: A policy perspective*. Centre for Learning on Evaluation and Results – Anglophone Africa (CLEAR-AA).
- Masvaure, S., & Fish, T. E. (2022). Strengthening and measuring monitoring and evaluation capacity in selected African programmes. *African Evaluation Journal*, 10(1), a635. <https://doi.org/10.4102/aej.v10i1.635>
- Matsiliza, N. S. (2019). Strategies to improve capacity for policy monitoring and evaluation in the public sector. *Journal of Reviews on Global Economics*, 8, 490–499. <https://doi.org/10.6000/1929-7092.2019.08.42>

- 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>
- Stewart, R. (2015). A theory of change for capacity building for the use of research evidence by decision makers in southern Africa. *Evidence & Policy*, 11(4), 547-557. <https://doi.org/10.1332/174426414X1417545274793>
- 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>