# What are the key measures used to characterize levels and changes in evaluative evidence demand, supply, and use?

## Executive Summary

This literature review examines how evaluative evidence demand, supply, and use are measured across diverse policy and organizational settings. It synthesizes academic and practitioner insights to clarify which indicators are most commonly used, the strengths and limitations of current measurement practices, and the contextual factors that shape how these dimensions are operationalized.

Measures of demand often include government or donor requests for evaluations, budget allocations for monitoring and evaluation (M&E), and formal policy frameworks mandating evidence-informed decision-making. Studies further distinguish between actual, potential, and latent demand, underscoring the role of institutional structures and donor influence, particularly in low-income settings.

Supply is gauged by evaluating the number, qualifications, and organizational affiliations of evaluators, as well as the strength of national evaluation systems. Training programs, evaluator certification, and institutional networks serve as key indicators, though concerns persist over the sustainability and depth of capacity-building initiatives.

Use is the most complex to assess, encompassing instrumental, conceptual, and symbolic dimensions. Common indicators include the citation of evaluation findings in policy documents, stakeholder involvement in evaluations, and evidence of learning or behavioral change. However, many metrics remain process-focused and fail to capture long-term policy influence.

Cross-cutting themes include inconsistent measurement standards, donor-driven priorities, and the pivotal role of institutionalization in sustaining meaningful evaluation practices. The review concludes by calling for greater methodological standardization, more outcome-focused indicators, and longitudinal approaches that reflect the political and organizational realities influencing evidence use in policymaking.

## Introduction

This literature review addresses the research question: “What are the key measures used to characterize levels and changes in evaluative evidence demand, supply, and use?” In exploring this question, the review synthesizes scholarship on how “demand,” “supply,” and “use” of evaluative evidence are conceptualized and measured in varied organizational, governmental, and sectoral contexts. The discussion proceeds through four major sections. First, it examines “Measures of Demand,” focusing on institutional markers such as budget allocations, national policies, and formal requests for evaluations. Second, it explores “Measures of Supply,” emphasizing indicators related to evaluator capacity, training programs, and institutional frameworks. Third, it addresses “Measures of Use,” describing how the integration of evaluation findings in policy and decision-making is tracked. Fourth, it identifies cross-cutting insights and thematic observations related to data collection methods, standardization gaps, and the influence of institutional and political factors on all three domains. The review concludes by summarizing key findings and discussing implications for advancing measurement practices in the field of evaluation.

## Measures of Demand

Scholars across multiple contexts highlight that “demand” for evaluative evidence can manifest in different ways, including explicit government or donor requests for evaluations, national budget allocations earmarked for evaluation, or institutional mandates requiring evidence-based decision-making (Adams et al., 2013; Mulenga & Porter, 2013). One frequently cited indicator is the existence of formal policy frameworks or guidelines that institutionalize monitoring and evaluation (M&E) processes. For example, CLEAR-AA (2020) underscores that a national M&E policy or formal government requirement for evaluations often serves as a proxy for strong institutional demand. Where national policies clearly mandate evaluations – by specifying, for instance, a fixed percentage of ministry budgets for evaluative activities – researchers treat these provisions as tangible measures of demand. In Ghana, Adams et al. (2013) noted a fluctuation in budget allocations for M&E (3% in 2009 down to 1% in 2010), illustrating how shifts in government support can be measured numerically and used to infer political and institutional appetite for evaluative evidence.

Studies also differentiate among “actual,” “potential,” and “latent” demand, as introduced by Adams et al. (2013) and further elaborated by Kumwenda and Latib (2013). Actual demand refers to real, immediate calls for evaluations (e.g., ministerial requests or parliamentary inquiries). Potential demand might arise if certain enabling conditions materialize, such as new performance management policies or reforms that incentivize the use of evaluation. Latent demand encompasses unrecognized needs – cases in which decision-makers are unaware of or uninterested in commissioning evaluations even though substantial benefits might accrue. Many scholars note that donor-driven mandates often shape actual demand in lower-income countries. Mulenga and Porter (2013) confirm that in Zambia, donors repeatedly commission or require evaluations as part of accountability requirements, making external impetus a strong driver of demand. Similarly, Kumwenda and Latib (2013) in Malawi find that while government initiatives for performance assessment are budding, donor pressure remains a more consistent source of formal evaluation requests.

An additional measure of demand involves specific institutional structures, such as the presence of M&E units or coordinating bodies within government ministries. Adams et al. (2013) reference the Policy Evaluation and Oversight Unit in the Presidency (Ghana) as an example of an organizational anchor for sustaining demand. CLEAR-AA (2020) notes that when government agencies establish and fund dedicated M&E units, it signals an ongoing institutional demand for evidence. A parallel indicator is the scrutiny by legislative or parliamentary committees. Adams et al. (2013) point out that the Ghanaian Public Accounts Committee has occasionally demanded evaluations, even though follow-up can be sporadic. Meanwhile, Niringiye (2018) describes how Uganda’s Ministry of Finance mandates evaluations for large-scale projects, a structural signal of demand measured by the percentage of projects requiring formal evaluation.

## Measures of Supply

Supply of evaluative evidence concerns the availability, capacity, and expertise of individuals and organizations capable of conducting rigorous evaluations. One key measure is the number and skill level of evaluators. As documented by Niringiye (2018), Uganda actively tracks the size of its evaluator pool, initially identifying 75 individuals and 18 firms, later updating to over 150 local experts. This quantification offers a tangible measure of growth in capacity over time. Mulenga and Porter (2013) similarly identify local universities, think tanks, and private firms in Zambia that contribute to the evaluation landscape. Counting these institutions, assessing the qualifications of their staff, and examining their professional affiliations (e.g., membership in Voluntary Organizations for Professional Evaluation [VOPEs]) reveal trends in the national evaluation supply.

Beyond raw counts, researchers gauge evaluator competence by examining specialized training, professional degrees, and proficiency in data analysis software, such as NVivo or Stata (Wao et al., 2017). For instance, Masvaure and Fish (2022) report that capacity-strengthening initiatives in several African countries measure change through pre/post training assessments or by tracking the number of individuals who become certified in advanced evaluation methods. This approach moves beyond headcount to evaluate the depth and breadth of M&E skill sets. Donnelly and Searle (2017) highlight that evaluation capacity building (ECB) is often operationalized by counting workshops conducted or evaluating the presence of formal M&E degree programs, though these efforts sometimes capture only short-term outputs (e.g., number of trainees) rather than the long-term improvements in institutional capacity.

Quality indicators also emerge as measures of supply. Johnson et al. (2009) point to evaluator competence and cultural sensitivity as salient predictors of evaluation outcomes. Various frameworks assess the level of methodological rigor, the clarity of evaluation design, and timeliness of deliverables. Kanyamuna (2021) references LEADS scoring in Zambia (ranging from “Little action” to “Sustainable”) as one structured method to assess dimensions such as organizational capacity, methodology, and participation. Although originally developed to evaluate both demand and use, LEADS can be adapted to measure the degree to which the supply side (e.g., public agencies, private firms) is institutionally ready to produce credible findings.

The distinction between local and international evaluators also serves as a critical supply-side measure. In multiple studies – particularly in Adams et al. (2013), Kumwenda and Latib (2013), and Niringiye (2018) – international consultants are shown to dominate major evaluations, often due to perceived or actual gaps in local expertise. Hence, quantifying the ratio of local to international evaluators in evaluations over time, as well as analyzing daily charge-out rates or labor markets for evaluators, can illuminate the changing supply dynamics. Furthermore, institutional frameworks, such as the presence of recognized M&E networks (e.g., the Ghana Monitoring and Evaluation Forum), are indicators of how robust or weak local supply might be (Adams et al., 2013).

## Measures of Use

In the context of M&E systemstrengthening initiatives, the notion of “use” encompasses not only whether the findings of an individual evaluation inform a discrete decision but also how an organization’s evaluative architecture consistently channels evidence into successive planning, budgeting, and accountability cycles. The “use” of evaluative evidence is frequently subdivided into instrumental, conceptual, and symbolic (or persuasive) use (Johnson et al., 2009). Instrumental use captures situations in which evaluations directly inform decisions, for example, when a policy is revised based on new findings. Conceptual use occurs when stakeholders’ understanding of a problem evolves due to exposure to evaluation results. Symbolic or persuasive use arises when findings are invoked primarily to justify pre-existing positions or satisfy donor or public demands for accountability (Johnson et al., 2009; Worton et al., 2017).

Different studies track use by documenting whether evaluation recommendations are integrated into official policy documents or strategic plans. For instance, Adams et al. (2013) look at the extent to which Ghana’s health and education sector plans reference completed evaluations. Similarly, Niringiye (2018) identifies how the Office of the Prime Minister in Uganda attempts to incorporate evaluation findings but notes that the actual uptake is limited by a lack of systematic feedback loops. Kuchenmüller et al. (2022) propose examining the frequency with which policymaking documents cite knowledge translation products, as well as measuring the participation of policymakers in evidence-related events, as proxies for use.

Another vein of research focuses on process use, i.e., the learning or capacity gains stakeholders experience simply by engaging in an evaluation. Donnelly and Searle (2017) note that stakeholder involvement, typically measured by counting how many and which stakeholders participate in research design or data analysis, can be an important predictor of whether evaluation findings eventually become integrated into decisions. This type of use may not be explicitly visible in final policy documents but can significantly shape organizational culture over time.

Building on these typologies, the literature distinguishes between evaluationspecific indicators of use – such as the direct adoption of recommendations from a particular study – and systemlevel indicators that signal the functionality of the broader M&E ecosystem. Scholars propose measures like the proportion of ministries that prepare annual performance reports drawing on any evaluation portfolio generated by the central unit, the frequency with which cabinet briefs cite the government’s synthesis evaluation reports, or the existence of institutional mechanisms – such as interministerial steering committees or standing parliamentary subcommittees – tasked with reviewing cumulative learning across multiple evaluations (CLEARAA, 2020; Kanyamuna, 2021). By tracing these higherorder feedback loops, researchers assess whether an M&E system has matured from producing isolated studies to sustaining a continuous evidence ecosystem capable of learning and adaptation. Accordingly, when the goal is to strengthen an M&E system rather than to appraise a single evaluation, measures of use must privilege these systemic indicators alongside, not in place of, traditional evaluationspecific metrics.

Several frameworks provide structured checklists or rubrics to capture various dimensions of use. Kanyamuna (2021), using the LEADS scoring system, measures the “Use of M&E Information” in Zambia’s government as a separate dimension, assigning a score based on action taken to integrate findings into practice. Similarly, Kanyamuna et al. (2018) implement a four-component diagnostic checklist to trace how M&E outputs are disseminated and, crucially, whether they guide budget allocation or legislative oversight. Mansilla et al. (2024) focus on categorizing the types of demand-driven questions that policymakers pose and then tracking how those questions get answered or used to inform decisions. These varying approaches underscore the complexity in measuring use: some methods rely on formal citation analyses, others adopt stakeholder surveys or interviews, and still others examine structural indicators, such as references to evaluation findings in newly enacted legislation.

## Cross-Cutting Insights and Thematic Observations

A recurrent theme is the challenge of methodological consistency. Several studies note that while surveys and interviews are common methods for gathering data on demand, supply, and use, the specific indicators vary widely across contexts (Adams et al., 2013; CLEAR-AA, 2020; Donnelly & Searle, 2017). LEADS scoring (Kanyamuna, 2021) and the CKME framework (Worton et al., 2017) exemplify attempts at structured approaches, yet their scoring or categorization schemes differ in scope and application. CLEAR-AA (2020) points out that although national M&E frameworks might standardize reporting procedures, actual measurement of use (and to some extent demand and supply) remains inconsistent and heavily dependent on donor or sector-specific guidelines.

Another cross-cutting challenge involves political and institutional contexts. Donor-driven motivations for evaluation can skew which aspects of demand, supply, and use get tracked, often emphasizing accountability over learning or long-term capacity. Kumwenda and Latib (2013), for example, show that external funders often require robust documentation of “demand” (i.e., government willingness) and “use” (i.e., compliance with donor requirements), yet local stakeholders may not internalize or continue these measurement practices once projects end. Similarly, Witter et al. (2017) find that health ministries often need evidence more rapidly than conventional evaluations can provide, highlighting a tension between the supply side’s rigorous but time-intensive methods and the immediate evidence needs of policymakers.

Regarding data collection methods, both quantitative and qualitative approaches appear across the literature. Budget analyses, rosters of evaluators, and policy document reviews produce numerical indicators (e.g., percentage of projects evaluated, quantity of local experts, ratio of training programs). Meanwhile, interviews, case studies, and focus groups yield deeper insight into how and why evaluation findings drive policy (or not). Johnson et al. (2009) emphasize that reliance on self-reported usage data can introduce bias, while CLEAR-AA (2020) critiques the general dearth of standardized outcome measures for capacity-building efforts, often leading to superficial or process-level assessments.

Finally, the role of institutionalization appears repeatedly as a linchpin for sustaining meaningful measurement of demand, supply, and use. Studies indicate that governments with formal M&E policies, dedicated budget lines, professional associations for evaluators, and consistent feedback channels to policymakers are better able to generate reliable data on each dimension (Adams et al., 2013; Mulenga & Porter, 2013). However, even in these countries, authors frequently note fragmentation between governmental and non-governmental M&E activities, underscoring the need for more unified frameworks to capture dynamic changes in the evaluation ecosystem.

## Conclusion

In summary, the literature shows that measuring demand, supply, and use of evaluative evidence involves an array of indicators ranging from budget allocations and the presence of dedicated M&E policies or units (demand) to the capacity and expertise of evaluation professionals (supply) to whether findings meaningfully shape policies or practices (use). Many studies reference donor-driven measurements – such as the requirement for baseline, mid-term, and final evaluations – as key proxies for actual demand, but others emphasize more endogenous signals like parliamentary oversight and civil society advocacy. On the supply side, quantifiable metrics focus on the number and caliber of evaluators, the existence of national or regional M&E associations, and the prevalence of robust training initiatives. Use, meanwhile, is typically gauged through explicit policy uptake, citation in official documents, or evidence of process use via stakeholder learning. Across the source material, the most robust measures appear to be those that blend structured frameworks (e.g., LEADS scoring, CKME) with both quantitative and qualitative insights on how evaluations are commissioned, conducted, and eventually utilized. Critically, where evaluative interventions seek to build durable M&E systems, future measurement frameworks should incorporate systemlevel indicators – such as the regularity of evidence reviews in budget negotiations or the proportion of cabinet dockets accompanied by evaluation syntheses – to complement studyspecific uptake measures.

Looking ahead, these findings suggest a strong need for greater standardization in how demand, supply, and use are tracked, as well as more rigorous methodologies for linking evaluation outputs to longer-term policy outcomes. Notably, many of the documented measures remain process-oriented (e.g., counting evaluations, tracking training outputs, or noting the presence of M&E units) rather than capturing tangible shifts in decision-making culture or policy effectiveness. Future research could develop unified metrics that better account for the political and institutional contexts affecting commissioning behaviors, capacity development, and actual policy influence. Such efforts might include robust longitudinal studies that systematically compare evaluation systems within and across countries, while addressing political economy challenges that constrain the meaningful use of evaluative evidence.

## **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.

Centers for Learning on Evaluation and Results – Anglophone Africa (CLEAR-AA). (2020). *M&E capacity-strengthening approaches and measurement in Africa*.

Donnelly, C., & Searle, M. (2017). Optimizing use in the field of program evaluation by integrating learning from the knowledge field. *Canadian Journal of Program Evaluation, 31*(3), 305–327.

Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation, 30*(3), 377–410.

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.

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.

Kuchenmüller, T., Chapman, E., Takahashi, R., Lester, L., Reinap, M., Ellen, M., & Haby, M. M. (2022). A comprehensive monitoring and evaluation framework for evidence to policy networks. *Evaluation and Program Planning, 91*, 102053.

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.

Mansilla, C., Sweetman, A., Guyatt, G., & Lavis, J. N. (2024). A taxonomy of demand-driven questions for use by evidence producers, intermediaries, and decision-makers: Results from a cross-sectional survey. *Health Research Policy and Systems, 22*, 78.

Masvaure, S., & Chirau, T. (2020). *M&E capacity-strengthening approaches and their measurement in Anglophone Africa: A policy perspective*. CLEAR-AA. [ResearchGate pre-publication]

Masvaure, S., & Fish, T. E. (2022). Strengthening and measuring monitoring and evaluation capacity in selected African programmes. *African Evaluation Journal, 10*(1), a635.

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.

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.

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.

Worton, S. K., Loomis, C., Pancer, S. M., Nelson, G., & Peters, R. D. (2017). Evidence to impact: A community knowledge mobilisation evaluation framework. *Gateways: International Journal of Community Research and Engagement, 10*, 121–142.