

# Making the Implicit Explicit: A Theory-Informed Reflective Case of Emergent Evaluation Capacity Building in a Public Training Institution

Doreen Otieno  
*Ohio University*

John Malombe  
*Kenya School of Law*

Krisanna Machtmes  
*Ohio University*

Samuel Isiko  
*Ohio University*

Samuel Obudho  
*Kenya Institute of Curriculum Development*

**Background:** Evaluation Capacity Building (ECB) is increasingly understood as a process through which organizations develop evaluative thinking, learning, and evidence-informed decision-making. While ECB is often associated with formal training initiatives, emerging literature suggests that evaluative capacity can also develop through routine organizational practices, leadership behaviors, and governance processes.

**Purpose:** This reflective case study examines how evaluation capacity became embedded in everyday organizational practices within a public legal training institution, despite not being explicitly recognized as ECB at the time. The paper explores how evaluative thinking emerged through governance routines, peer learning structures, and decision-making processes.

**Setting:** A public legal training institution in East Africa responsible for professional legal education, paralegal training, and continuing professional development.

**Intervention:** Not applicable.

**Research Design:** Theory-informed reflective case study drawing on reflective inquiry, organizational artifacts, and retrospective analysis guided by Preskill and Boyle's (2008) Evaluation Capacity Building framework.

**Data Collection and Analysis:** Data sources included organizational documents such as committee minutes, leadership retreat notes, planning documents, monitoring and evaluation reports, and internal memoranda. Analysis involved reflective interpretation of institutional practices, triangulated through documentary evidence and informal peer debriefing.

**Findings:** Three interconnected practices supported the emergence of evaluation capacity: (1) creating supportive structures and a culture of inquiry, (2) fostering peer learning and practice-based evaluation mechanisms, and (3) integrating evaluation into planning and decision-making processes. The analysis proposes an implicit-to-explicit four-stage pathway through which evaluative practices evolved from latent organizational routines to an embedded evaluative culture supported by leadership, feedback loops, and institutional systems.

**Keywords:** *evaluation capacity building; evaluative thinking; evaluation culture; organizational learning; emergent; embedded.*

Evaluation capacity building (ECB) is widely understood as a means through which organizations strengthen their capacity for critical inquiry, organizational learning, and evidence-informed decision-making over time. According to Searle and colleagues (2024), ECB is a process that supports new learning, fosters shared understanding, and recognizes existing evaluation efforts. Traditionally, ECB efforts have emphasized formal training approaches, including workshops, technical skill development, and the use of external expertise (Stockdill et al., 2002). These approaches remain valuable in providing essential tools and foundational knowledge. However, they may not fully capture how evaluation becomes deeply embedded in an organization's culture, an area that recent scholars have laid emphasis on the need to explore (Fierro et al., 2024; Nielsen et al., 2024). A growing body of work highlights how evaluation capacity may emerge and be modeled through leadership decisions, institutional routines, decision-making practices, and feedback loops, even when no formal ECB program is in place (Buckley et al., 2015; Taylor-Ritzler et al., 2013). Scholars are also calling for more attention to how evaluative thinking becomes embedded in daily practice, rather than being delivered through one-time training events (Chaytor, 2023).

This reflection builds on that shift in thinking by revisiting a period of institutional practice in which many core elements of evaluation capacity building (ECB) were already in motion, even though they were not named as such. During this period of reform, the organization did not conceptualize its activities as evaluation capacity building; rather, the practices were experienced as routine and largely oriented toward improvement and meeting institutional demands. These practices align with what Preskill and Boyle (2008) describe as organizational learning strategies for ECB. This reflection focuses on three interrelated practices—creating supportive structures and culture, peer learning and practice-based mechanisms for evaluation capacity, and integrating evaluation into planning and decision-making—situated within a leadership-enabled governance and planning context that shaped how evidence, inquiry, and learning were taken up in practice. Drawing on these practices, the reflection advances an implicit-to-explicit four-stage pathway for ECB, illustrating how latent governance routines can be recognized, scaffolded through structures and feedback processes, and ultimately stabilized into

a more sustained evaluative culture. The analysis is guided by the central question, “How does ECB become embedded in everyday organizational practices when it is not explicitly named or recognized?”

## Methodological Approach

This paper adopts a theory-informed reflective case study approach, drawing on three interrelated methodological traditions: reflective inquiry, case study methodology, and theory-informed analysis. Schön's (2017) work on the reflective practitioner legitimizes the use of professional experience as a source of knowledge. Through reflection-on-action, I, Otieno revisit my role as an evaluation practitioner to surface tacit institutional practices that, while not originally named as ECB, reveal themselves as such through critical, retrospective analysis. The reflective analysis draws on a focused set of organizational artifacts that were used to ground and triangulate my recollection of institutional practice over time. These materials included board and senior management agendas and minutes, leadership retreat program and notes, committee agendas and minutes, planning documents and reporting templates, monitoring and evaluation reports, and internal memos documenting key decision-making processes. These artifacts were not treated as a comprehensive dataset, but as prompts to support reflection-on-action and to corroborate patterns observed through professional experience. Interpretive claims were further refined through informal sense-checking conversations with colleagues involved in the same period of work, helping to surface alternative perspectives. The institution is treated as a bounded case (Stake, 1995), defined by its governance and planning arrangements, allowing for in-depth examination of practices within a leadership-enabled context that oriented organizational attention toward inquiry, evidence use, and learning. The case is instrumental in nature, serving as a lens to explore how ECB can emerge organically through routine organizational processes rather than formalized interventions. Preskill and Boyle's (2008) multidisciplinary model of ECB provides the conceptual scaffolding for this reinterpretation. The framework highlights organizational-level strategies, such as leadership engagement, supportive structures, and peer learning mechanisms, that help build sustainable evaluative culture. Applying this model enables a

theory-informed reading of institutional routines previously viewed as administrative.

## Positionality Statement

Positionality is central to reflective and qualitative research, because knowledge is never produced from a neutral standpoint (Yip, 2024; Holmes, 2020). Researchers must therefore account for how their identities and experiences shape interpretation, the aspects they choose to foreground, and the insights they draw.

This reflection is grounded in the professional lived experience of the main author, Otieno, who previously served as an evaluation practitioner at the case institution. In that role, I led reforms focused on performance, accountability, and systems improvement. Although not framed at the time in the language of evaluation capacity building (ECB), many of these practices reflected principles largely recognized in ECB literature (Norton et al., 2016). I thus bring an insider perspective, both as a designer of the systems under review and as someone whose academic training now enables reinterpretation through a scholarly lens. This dual role creates potential for both insight and bias: Proximity provides contextual depth while requiring critical reflection on how positionality shapes interpretation, example selection, and conclusions. To address this, I employed analytic safeguards to mitigate retrospective bias. Interpretive insights were reviewed through a brief peer-debrief loop with a colleague who was not directly involved in the reforms examined, providing an opportunity to test assumptions and surface alternative interpretations. In addition, key analytic claims were traced to specific types of organizational artifacts (e.g., board and committee agendas, leadership retreat notes, and planning documents), allowing for triangulation between professional recollection, documentary evidence, and colleague input, and supporting transparency and analytic traceability.

The study did not involve human subjects research. The analysis draws exclusively on organizational records and professional reflection on leadership and organizational practices that oriented organizational attention toward inquiry, evidence use, and learning. While grounded in practitioner experience, the reflection was further enriched through collaboration with colleagues who contributed complementary theoretical, evaluative, and institutional perspectives, thereby strengthening reflexivity and analytic balance.

## Case Context

This study focuses on a public legal training institution in East Africa with a statutory mandate to train law graduates for admission to professional practice. In addition to its core postgraduate training programs, the institution offers paralegal certification and continuing professional development for practitioners. For the purposes of this inquiry, the institution is conceptualized as a bounded case, with attention directed to its governance and planning processes as sites in which evaluative capacity may be observed and interpreted.

## Framing ECB in This Reflection

While the practices discussed in this reflection were not originally labeled as evaluation capacity building (ECB), looking back through a scholarly lens invites a deeper question: What counts as ECB, and who gets to name it as such? To meaningfully reinterpret these institutional routines, it is first necessary to situate them within the broader discourse of how ECB has been conceptualized in the literature.

A recurring theme in the literature is the absence of a singular, universally accepted definition of ECB, something that some scholars view as problematic (Compton & Baizerman, 2007), while others see it as contextually necessary (Taut, 2007). As the field has evolved, so too have its conceptual boundaries, resulting in a wide range of definitions that reflect differing priorities, disciplinary influences, and contextual realities. Stockdill and colleagues (2002) describe ECB as a context-dependent and intentional set of actions aimed at making evaluation a routine organizational practice. Cousins and colleagues (2014) emphasize the importance of developing evaluative knowledge and technical competencies as part of building sustainable evaluation capacity, while McDonald and colleagues (2003) argue for broader notions like evaluation capability, which incorporates the long-term cultural sustainability of evaluation practices. Lee and Cousins (2004) further connect ECB to organizational learning and development, emphasizing that capacity building must align with the institution's capacity to learn. Preskill and Boyle (2008) describe it as teaching and learning strategies aimed at developing sustainable evaluation practice. Labin and colleagues (2012) define it as an intentional process to increase individual and organizational ability to conduct and use evaluation. Across

these definitions, common threads are evident. Most definitions highlight intentionality, capacity for use, and the need for systems and culture that support evaluation over time.

These definitional emphases are further developed in collaborative, participatory, and empowerment-oriented approaches to evaluation capacity building, which conceptualize ECB as an embedded, relational, and practice-based process. Empowerment evaluation explicitly positions evaluation as a means of building the capacity of individuals and organizations to assess their own activities, reflect on evidence, and use findings for improvement and accountability through ongoing participation (Fetterman, 2023). Collaborative and participatory evaluation approaches emphasize stakeholder involvement, shared inquiry, and collective sense-making as central mechanisms through which evaluative thinking and use develop over time (Cousins & Whitmore, 1998; Fetterman et al., 2018). Consistent with these participatory and empowerment-oriented perspectives, evaluation capacity is understood to develop incrementally through practice, as stakeholders repeatedly engage in evaluative activities embedded in planning, implementation, and decision-making processes, rather than solely through formal training interventions (Cousins et al., 2014; Taylor-Ritzler et al., 2013). Synthesis and review studies further underscore that ECB develops through multiple pathways and is shaped by organizational context, relationships, and leadership (Labin et al., 2012; Konjore, 2024).

Organizational learning and institutional theories further demonstrate how evaluative capacity can develop through routine organizational practice over time. Argyris and Schön's (1997) distinction between single- and double-loop learning illuminates how repeated governance and planning activities may initially support compliance-oriented learning but, when accompanied by inquiry and reflection, can enable deeper questioning of goals and assumptions. Feldman and Pentland's (2003) view of organizational routines as generative systems further suggests that stability in routine practice can create opportunities for variation and learning. The durability of such evaluative practice is shaped by processes of institutionalization and culture; as Schein (2010) argues, practices endure when artifacts, espoused values, and underlying assumptions align. Greater analytic precision is also gained by attending to how evaluation is used over time. Drawing on Weiss's (2021) typology, evidence use

in ECB may initially be conceptual, shaping understanding and sense-making, before becoming more instrumental as evaluative practices are embedded in decision-making processes. Accordingly, in this reflection, ECB is understood as an emergent process rooted in context that becomes embedded in routine practice over time, through which individuals and organizations develop the evaluative knowledge, behaviors, systems, and cultures that support not only the conduct of evaluation, but its sustained use for learning and decision-making.

## Theoretical Framework: Reframing Institutional Practices through ECB Theory

To make sense of the evaluative practices described in this reflection, this paper draws on Preskill and Boyle's (2008) multidisciplinary model of evaluation capacity building. This theoretical lens enables a reinterpretation of routine institutional actions, not merely as managerial tasks, but as implicit ECB strategies that contributed to the development of a more evaluative culture over time. Applied retrospectively, the framework helps to make implicit practices explicit and demonstrates how evaluation capacity building can emerge through everyday institutional routines. This reinterpretation encourages re-examination of ordinary organizational routines through a new lens, one that surfaces the evaluative potential embedded in institutional culture and structure. In this light, what may have appeared as administrative formalities take on deeper significance as drivers of evaluative thinking and practice. Guided by Preskill and Boyle's (2008) framework, this reflection identifies three interrelated domains—creating supportive structures and culture, peer learning and practice-based mechanism, and integrating evaluation into planning and decision-making—all situated within a leadership-enabled governance and planning context.

## Leadership as a Foundational Enabling Condition for ECB

ECB literature extensively identifies leadership as a critical enabling condition for the institutionalization and sustainability of evaluative practice (Preskill & Boyle, 2008; Labin et al., 2012; Taylor-Ritzler et al., 2013). In this

case, leadership played a foundational role in the emergence and sustainability of evaluation capacity within the institution. While structures, systems, and peer learning mechanisms enabled evaluative routines, leadership signaled the importance of evaluation, set the tone for learning-oriented governance, and created the conditions under which evaluative practices could take root and endure. This framing situates leadership as a crosscutting enabling condition shaping how the domains of practice that follow were enacted and sustained.

## Creating Supportive Structures and Culture

ECB is most sustainable when reinforced by organizational structures and leadership styles that normalize the use of evaluation across decision-making levels (Preskill & Boyle, 2008; Labin et al., 2012). In this case, routine governance mechanisms, including quarterly board-level and senior management meetings and annual leadership retreats, established a predictable cadence for engaging with evidence, serving as dynamic forums in which evaluative thinking was embedded into institutional practice. This regular rhythm reduced reliance on ad hoc or crisis-driven data use and created recurring, expected moments for interpretation, questioning, feedback and decision-making consistent with Feldman and Pentland's (2003) view that routines generate stability while enabling adaptation, as actors repeatedly reinterpret and act on evidence within familiar governance structures.

Prior research has long characterized boardrooms as politically charged spaces where evidence is often used symbolically rather than reflectively (Weiss, 1979; Pettigrew & McNulty, 1995; Boswell, 2009). This skepticism toward evaluation was initially evident in this case, where data presentations functioned more as performances of control than as prompts for inquiry. Instead of attempting to counter this skepticism through the introduction of additional data or stronger claims, governance routines in some contexts appeared to evolve such that evidence was increasingly used to support collective questioning and forward-looking judgment, rather than primarily serving as a tool for justification or compliance. Central to this was the reframing of the summary report as a concise, high-level interpretive document structured around recurring learning questions—for

example, what the data was indicating, where assumptions might need revisiting, and what the implications were for future strategy. As part of this structure, meetings routinely included a brief, focused examination of key deviations between expected and observed results, emphasizing interpretation and future implications. Attention to patterns, deviations from expectations, and emerging risks, rather than performance alone, steered discussion toward collective interpretation and forward-looking judgment. These design choices created conditions for double-loop learning, moving discussion beyond whether targets were met (single-loop) to why assumptions, priorities, or strategies needed revisiting (Argyris & Schön, 1997).

Through these adaptations, board and senior management meetings evolved into structured grounds for data interrogation, collective sense-making, and evidence-informed action aligning with empowerment evaluation principles that emphasize integrating evaluation into ongoing planning, management, and decision-making processes (Fetterman, 2001; Fetterman & Wandersman, 2005; Patton, 2011). These deliberations were subsequently reflected in board-level strategic resolutions, which extended beyond formal decisions to articulate strategic direction informed by what the evidence was communicating about risks, trade-offs, and future trajectories. In several instances, evaluative insights informed both the setting of strategic priorities and the movement of resources toward areas of greatest emerging need or risk. In this way, evaluation was repositioned from an occasional activity to a foundational governance practice.

These regular meetings closely align with Preskill and Boyle's (2008) strategy of establishing formal structures that support evaluation use, as they provided consistent opportunities for leaders to engage with data, interrogate findings, and apply evaluation insights directly in governance decisions. Further, the durability of these practices can be understood through Schein's (2010) model of organizational culture. Initially operating as visible artifacts, such as reports, meetings, and review routines, evaluative practices gained traction as leaders came to treat them as resources for inquiry and learning. Over time, repeated enactment within governance and planning systems helped align espoused values with underlying assumptions about the role of evidence, contributing to evaluation becoming a

routine and expected feature of institutional practice.

#### Box 1. Five Micropractices for Reflective Governance Meetings

**Agenda-embedded learning questions:** Board and senior management meeting agendas were intentionally structured around recurring learning questions (e.g., what the data was indicating, where assumptions might need revisiting, and what the implications were for future strategy), orienting discussion toward inquiry rather than performance defense.

**Interpretive summary reporting:** Summary reports were reframed as concise, high-level interpretive documents that foregrounded patterns, deviations from expectations, and emerging risks, rather than exhaustive performance metrics, prompting collective sense-making rather than compliance reporting.

**Focused variance analysis ritual:** Meetings routinely included brief, structured examination of key variances between expected and observed results, emphasizing interpretation and future implications rather than attribution of blame or performance justification.

**Evidence-informed decision articulation (decision log):** Strategic deliberations and board resolutions explicitly documented how evaluation evidence informed decisions, articulating strategic direction in relation to risks, trade-offs, and future trajectories signaled by the data.

**Evidence-linked prioritization and follow-through ownership:** Insights from evaluative discussions informed the setting of strategic priorities and, in several instances, the reallocation of resources toward areas of emerging need or risk, with responsibility for follow-through embedded within governance and management structures.

These governance practices were, however, not embraced without tension. As board-level discussions became more interpretive and evidence was increasingly used to surface risks, capacity gaps, and implementation constraints, moments of unease emerged, particularly among management actors who experienced this shift as increased exposure of departmental performance in a forum traditionally oriented toward approval and oversight. In early instances, management responses reflected a defensive orientation, with evaluative data interpreted as implicit judgment rather than as a resource for collective learning. These dynamics reflect well-documented tensions between boards and management in governance settings, where evidence use can disrupt established accountability relationships and reconfigure authority over how performance narratives are constructed (Weiss, 2021; Boswell, 2009).

Further, the annual leadership retreat, held off-site to facilitate uninterrupted reflection and dialogue between the board, executive, and management functioned as a reflective learning space, reinforcing what Preskill and Boyle (2008) identify as *promoting a culture of inquiry* and *creating ongoing feedback loops*. Prior scholarship has emphasized that evidence use is most effective when data is engaged with collectively through facilitated sense-making rather than consumed individually through

reports (Weick, 1995; Patton, 2011; Johnson et al., 2007). Strategic retreats and off-site workshops have been shown to create conditions for such reflection by temporarily suspending routine pressures and enabling shared interpretation of complex information (Argyris & Schön, 1997). These retreats became intentional spaces where evaluation reports were critically examined, activating feedback loops between strategic oversight and operational execution. They were structured around pre-prepared evidence packets developed by task forces and validated by implementers, combining synthesized evaluation findings, trend data, and selected performance indicators intentionally curated to prompt interpretation. Board members, executives, and department leads engaged in probing discussions that illuminated institutional blind spots, surfaced learning, and prompted adaptive decisions, including difficult conversations that often unfolded outside formal agendas, in what many referred to as “off-agenda dialogues.”

Such conversations emerged when evidence highlighted a growing gap between strategic priorities and available operational capacity, prompting candid exchange about the feasibility of proposed commitments within existing resource constraints. While board members emphasized the strategic risks of inaction signaled by the data, management drew attention

to the practical implications for staffing, budgets, and delivery timelines. Rather than resulting in immediate resolution, the dialogue clarified trade-offs and informed how strategic directions were subsequently taken forward into planning cycles, shaping prioritization, implementation sequencing, and, in some cases, procurement and resource-allocation decisions. Such practices align with what Mayne (2008) describes as fostering an evaluative culture, one in which evidence use, and reflection are embedded in institutional practice.

In retrospect, routine meetings and leadership retreats may be viewed as embedded evaluation capacity-building (ECB) strategies. These structures provided consistent opportunities for reflection, inquiry, and the use of evaluation findings, key components of organizational-level strategies outlined by Preskill and Boyle (2008) and further reinforced by Taylor-Ritzler and colleagues (2013), who emphasize leadership engagement and structural alignment as important predictors of sustained evaluation capacity. As such, this interpretation helps make implicit evaluative practices analytically visible, illustrating how institutional routines can evolve into mechanisms that support the development and maintenance of a culture of evaluation.

## Peer Learning and Practice-Based Mechanisms for Evaluation Capacity

The institution supported evaluative practice through the formation of evaluation task forces, technical working groups, and activity-specific committees. Consistent with Preskill and Boyle's (2008) argument that peer learning groups are critical for reinforcing evaluative thinking across organizations, these structures provided recurring opportunities for shared interpretation of data and collective reflection. Although most were initially created to support routine operations and programmatic reforms, over time they evolved into spaces for collaborative inquiry, co-design of interventions, and continuous improvement.

In practice, these task forces and committees brought together staff from across departments, including operational, technical, and management actors to collaborate on activities such as performance monitoring, program review, and institutional planning. Their regular engagement with evaluation data, ranging from student and customer feedback to institutional performance reports, encouraged shared

learning and critical dialogue. As members worked together to make sense of evidence, they built a common evaluative language and contributed to a culture in which reflective inquiry was normalized. These practices align with Preskill and Boyle's (2008) emphasis on organizational strategies that promote internal learning structures, peer engagement, and shared ownership of evaluation processes. Such patterns of collective interpretation are also consistent with Nutley and colleagues' (2007) findings that evidence use in organizations is shaped through dialogue, shared understanding, and social interaction.

These groups functioned as mechanisms through which evaluative thinking was incorporated into everyday organizational work. They played a central role in preparing evaluation summaries and briefing materials for management, while also identifying evolving data needs within their respective departments. Members routinely convened small-group data interpretation sessions in which emerging findings were discussed and contextualized before being shared more widely with staff and senior leadership. Through these activities, formal task structures became vehicles through which evaluative thinking was enacted and sustained in practice.

Beyond their technical functions, these groups increasingly operated as peer learning networks in which evaluative thinking was modeled and shared across units. Staff who participated frequently began initiating evaluative conversations within their own departments, mentoring colleagues, and encouraging data-informed decision-making, effectively serving as informal evaluation champions. As these practices became more routine, participants increasingly drew on evidence to justify positions, frame recommendations, and reason through trade-offs during planning and review discussions. Planning and evaluation instruments and performance data were regularly referenced in organizational conversations, and these groups came to be seen as internal points of reference for evidence use within their units. Over time, repeated engagement with data contributed to greater confidence in organizational dialogue, as staff and managers alike were better equipped to articulate rationales, question assumptions, and negotiate priorities using shared evidence.

The expansion of peer-based evaluative practices was not without tension. As evaluative conversations became more visible and data were increasingly used to interrogate performance and

operational decisions, concerns emerged among some members of management about the exposure of departmental inefficiencies and the potential erosion of managerial control over performance narratives. These concerns were particularly evident when peer learning groups began interpreting data and framing recommendations, activities traditionally associated with managerial authority. This redistribution of interpretive agency challenged established norms regarding who was entitled to define performance narratives and draw evaluative conclusions. These dynamics mirror longstanding observations that learning-oriented evaluation can unsettle established governance arrangements, particularly when interpretive authority is distributed beyond formal hierarchies (Weiss, 2021; Cousins et al., 2004). Efforts to reassert oversight through role clarification or information control reflect ongoing negotiation between evaluative learning and managerial accountability (Boswell, 2009; Patton, 2011).

These peer learning structures illustrate what Cousins and colleagues (2004) describe as context-sensitive and participatory approaches to evaluation, where individuals build capacity through structured collaboration, shared learning, and routine engagement with data. Through regular interaction, collaborative problem-solving, and shared engagement with evaluation data, these groups cultivated collective learning and built evaluative habits across the institution. This recognition enables these spaces to be analytically understood as peer learning networks. Doing so helps surface their evaluative function: They served as relational environments where evaluation capacity was built through social learning. Viewed this way, these networks exemplify the kind of organizational-level strategies outlined by Preskill and Boyle (2008), where evaluative thinking emerges from embedded, practice-based mechanisms sustained by peer networks and institutional routines.

## Integrating Evaluation into Planning and Decision-Making

Consistent with Preskill and Boyle's (2008) multidisciplinary model of evaluation capacity building, and with core principles of empowerment evaluation that emphasize embedding evaluative thinking within ongoing

organizational processes (Fetterman, 2001; Fetterman & Wandersman, 2005), the institution made targeted investments that supported the integration of evaluative practices into planning and decision-making systems. A core element of the institution's organizational-level approach involved the development of structures and processes that supported evaluation use across the institution (see Table 1). In this case, evaluation was advanced through the alignment of systems, policies, and tools that normalized evaluative practice.

For instance, one notable milestone was the integration of a dedicated monitoring and evaluation (M&E) framework within the institutional strategy blueprint. This framework outlined key indicators, assigned responsibilities, established timelines, and introduced learning loops to guide implementation. It also committed the institution to conduct midterm and end-term evaluations, not only of outcomes but of the evaluation process itself. This embedded structure mirrors what Goodyear and colleagues (2024) describe as scaffolding for evaluative practice: institutional mechanisms that reinforce reflection and learning. This formalization of evaluative logic also reinforces Preskill and Boyle's emphasis on embedding evaluation in structures that support long-term use.

System-level coherence was further supported by aligning key planning instruments, including strategic plans, performance contracts, work plans, procurement frameworks, and budgets, with evaluation findings to ensure decisions were grounded in evidence. A performance-based planning model required departments to move beyond listing activities and articulate measurable outcomes and timelines. Procurement and budget proposals were similarly restructured to align with strategic priorities and justify resource allocation using data. These changes resonate with Bamberger and colleagues' (2015) systems approach to M&E integration, where learning and adaptation are built into institutional processes. This approach reflects an emphasis on aligning evaluation tools with organizational systems to strengthen capacity (Suárez-Balcázar & Taylor-Ritzler, 2013) and echoes observations that capacity is more likely to take root when tools are codeveloped with users and grounded in context (Searle et al., 2024).



Table 1. Flow of Evaluation Evidence Across Planning, Budgeting, Procurement, and Review Cycles

| Step | Institutional stage                                  | Routine mechanism                                                    | Evidence entry point                                                                   | Decision effect                                                                          |
|------|------------------------------------------------------|----------------------------------------------------------------------|----------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| 1 →  | Monitoring and evaluation data generation and review | Routine performance monitoring, evaluations, and feedback collection | Patterns, deviations from expectations, and emerging risks identified                  | Priority issues flagged for planning discussions                                         |
| 2 →  | Work planning                                        | Development of departmental and institutional work plans             | Evidence used to inform prioritization of proposed activities, timelines, and outcomes | Workplans aligned with strategic priorities and learning needs                           |
| 3 →  | Budget formulation                                   | Performance-based budgeting and resource justification               | Evaluation data used to substantiate resource requests and trade-offs                  | Resource allocation decisions informed by evidence of need, risk, and relevance.         |
| 4 →  | Procurement planning                                 | Annual procurement planning and sequencing                           | Evidence used to prioritize and phase procurements                                     | Investments aligned with evaluative priorities                                           |
| 5 ∪  | Quarterly review and adjustment                      | Senior management and board review meetings                          | Results, variance analysis, and lessons reflected upon                                 | Adjustments inform subsequent indicator refinement and performance monitoring priorities |

Although originally developed to ensure coherence across plans, budgets, and institutional priorities, these systems functioned in ways that extended far beyond administrative alignment. The deliberate integration of evaluation findings into decision-making routines, combined with formal structures like strategic plans and policies, helped normalize the use of evidence across the institution. This emphasis on integrating evaluative thinking into ongoing planning and decision-making processes is consistent with core principles of empowerment evaluation, which foreground the incorporation of evaluation into routine organizational systems to support learning, accountability, and self-directed improvement (Fetterman, 2001; Fetterman & Wandersman, 2005). These systems reflect the kind of organizational-level mechanisms described by Preskill and Boyle (2008) that support evaluative thinking through the everyday tools and routines that shape how institutions plan, allocate resources, and learn. This analytic move clarifies evaluative practices that were previously implicit: that evaluation had become a governing logic within the institution's core systems.

This institutionalization also unfolded alongside persistent tensions between evidence-informed priorities and externally imposed compliance and fiscal constraints. As a public

sector institution operating within government planning and budgeting cycles, resource allocation decisions were largely shaped by statutory budget ceilings, centrally determined procurement rules, and competing mandates across government structures. In several instances, evaluation evidence pointed to strategic needs or implementation risks that could not be immediately acted upon due to fixed budget classifications, delayed releases, or externally set expenditure priorities. These are the messy realities of evaluation capacity building in public sector systems, where learning-oriented aspirations must coexist with rigid accountability frameworks, fiscal controls, and multilevel governance mandates. Such dynamics are well documented in the public management and evaluation literature (Lipsky, 2010; Pollitt & Bouckaert, 2017; Bamberger et al., 2015).

### Leadership in Action: Modeling, Enabling, and Sustaining Evaluation Capacity

Across the three domains of practice, leadership functioned as the integrative mechanism through which evaluative practices were enacted, aligned, and sustained (see Table 2). Leadership, including the board, director, and management,

modeled evaluative thinking through their consistent engagement with data and openness to feedback. Board meetings were not merely venues for reporting progress but also spaces where leadership engaged in interrogating findings, requesting deeper analysis, and connecting insights to strategic priorities. This active engagement signaled that evaluation was not a compliance exercise but a decision-support function, integral to governance and strategy. As Taylor-Ritzler and colleagues (2013) argue, when leaders demonstrate data use and reflection in action, they model behaviors that build evaluative culture.

Further, leadership enabled evaluation capacity by investing in structures that normalized its practice. For instance, the decision to institutionalize quarterly monitoring and evaluation reports as a standing agenda item for both management and board meetings signaled that performance and learning were governance priorities. Similarly, the strategic use of annual leadership and management retreats as reflective spaces demonstrated a commitment to learning as a core leadership function. These investments align with the findings of Labin and colleagues (2012), who argue that leadership support is one of the strongest predictors of sustainable ECB outcomes.

Leaders also sustained evaluation capacity by establishing accountability and ownership mechanisms across all levels of the institution. For example, department heads were required to submit quarterly performance updates linked to strategic outcomes and to explain variances using data. These routines supported a culture of inquiry, reinforced expectations for evaluative engagement, and made data use an operational norm rather than a discretionary practice. This reflects the central role of leadership in embedding evaluation into institutional

structures and expectations (Preskill & Boyle, 2008) and complements the view of ECB as a learning-oriented process sustained through reflective practice embedded in routine organizational work (Searle et al., 2024).

Crucially, leadership also demonstrated a willingness to grapple with complexity and engage in adaptive learning. During periods of reform, when decisions carried institutional weight and uncertainty, leaders sought out evidence to guide course corrections, often revisiting earlier assumptions and openly acknowledging missteps. These behaviors exemplify what Patton (2011) describes as leadership that is reflexive, inquiry-driven, and oriented toward learning. Such leadership shapes the conditions under which evaluative thinking becomes embedded and sustained.

Finally, the leadership culture supported mentorship and diffusion of evaluative values across the institution. Senior leaders encouraged mid-level managers and operational staff to lead evaluation-related initiatives, participate in planning reviews, and champion the use of evidence within their units. This distributive leadership approach helped decentralize evaluative responsibility and cultivated a second tier of internal evaluation champions. As Cousins and colleagues (2004) and Bamberger and colleagues (2015) argue, the diffusion of leadership for evaluation across hierarchical levels is key to long-term sustainability. In sum, leadership was not a peripheral enabler but a central force in the institution's journey toward greater evaluative maturity. Through the interplay between leadership styles and institutional routines, evaluation capacity became embedded as a core logic of how the institution understood itself, made decisions, and pursued its mission.

Table 2. Analytical Mapping of Preskill and Boyle's (2008) ECB Components to Institutional Practices and Outcomes

| ECB components                                        | Concrete practices in the case                                                                  | Evidence sources                                      | Observed outcomes                                                       |
|-------------------------------------------------------|-------------------------------------------------------------------------------------------------|-------------------------------------------------------|-------------------------------------------------------------------------|
| Leadership modeling and support                       | Board and senior management interrogated data, posed learning questions, revisited assumptions. | Board minutes, summary reports, retreat deliberations | Evaluation framed as decision support rather than compliance            |
| Creating supportive structures and culture            | Regularized board meetings, learning-oriented agendas, annual leadership retreats.              | Retreat evidence packets                              | Reflective governance normalized; inquiry embedded in leadership forums |
| Peer learning and shared ownership                    | Task forces and working groups interpreted data, prepared briefs, convened data discussions.    | Evaluation summaries, departmental reviews            | Distributed evaluative competence; informal evaluation champions        |
| Integration of evaluation into organizational systems | Alignment of evaluation findings with work plans, budgets, procurement.                         | Strategic plans, budgets, procurement justifications  | Evidence-informed prioritization and resource allocation                |
| Evaluation use for learning and decision-making       | Variance analysis, forward-looking discussion, off-agenda learning dialogues.                   | Quarterly review records, retreat discussions         | Shift from performance signaling to adaptive learning                   |
| Feedback loops and continuous learning                | Quarterly reviews informed adjustments to plans and indicators.                                 | Review cycles, revised plans                          | Iterative learning embedded across cycles                               |

## Surfacing Tensions: Navigating Compliance, Learning, and Institutional Practice

While the institution made meaningful strides in embedding evaluation as a learning-oriented governance tool, this journey unfolded amid a series of tensions that reveal the messy realities of ECB. As Goodyear and colleagues (2024) aptly remind us, tensions between theory and practice are not anomalies to be eliminated but essential features of how evaluation takes root in real-world settings. Within the institution, one of the most persistent tensions was the duality of evaluation as both a mechanism for learning and improvement and a tool for external compliance and accountability.

This tension is well-documented in the literature. Patton (2008) highlights the conflict between developmental learning and bureaucratic accountability, cautioning that evaluation can become hollowed out when driven primarily by compliance mandates. Christie and Fierro (2012) similarly warn against the risk of

evaluation becoming performative, focused more on appearance than substance, when institutional pressures reward favorable reporting over reflective insight. In this case, although institutional leaders framed evaluation as a driver of strategic learning and inquiry, external requirements, such as performance contracting, regulatory audits, and funding justification processes, often created competing incentives.

In practice, this involved differentiating how evaluative information was engaged across institutional forums. Formal performance reports prepared for external accountability emphasized agreed indicators and verified results, while internal spaces, such as management validation sessions, leadership retreats, and learning-oriented review meetings, were explicitly used to interrogate assumptions, surface implementation challenges, and examine emerging risks. Management validation sessions held prior to board meetings provided opportunities to contextualize sensitive findings, align interpretations, and anticipate areas of concern, reducing defensiveness and enabling more reflective dialogue in formal governance

settings. Within meetings themselves, attention was intentionally balanced between decision-making and interpretation, as reflected in agenda design, sequencing of discussions, and facilitation practices. Viewed analytically, this balance resembles facilitation heuristics used in leadership development and organizational learning, where action dominates but is supported by structured interpretation and reflective learning (Argyris & Schön, 1997). Importantly, this was not applied as a formal rule but emerged pragmatically through meeting design and facilitation. These dynamics reflect what Bamberger and colleagues (2015) describe as the “accountability-evaluation dilemma,” where systems designed to promote transparency inadvertently suppress learning by encouraging risk-avoidant reporting. Some reporting practices tended to emphasize achievements while downplaying implementation challenges, an approach shaped less by intent to obscure and more by the institutional pressure to demonstrate performance within externally driven accountability frameworks.

Goodyear and colleagues (2024) underscore that tensions in ECB can be generative when surfaced and negotiated rather than ignored. At the institution, deliberate efforts were made to carve out reflective spaces that allowed evaluation to be used for sense-making. Off-site retreats, informal debriefs, and learning-focused review meetings served as counterweights to compliance-driven structures. These moments provided opportunities for adaptive learning, consistent with double-loop learning (Argyris & Schön, 1996), as applied in Patton’s (2011) work on developmental evaluation.

The ECB literature increasingly calls attention to this tension. Labin and colleagues. (2012) suggest that ECB is most sustainable when it aligns with both technical systems and cultural values, yet achieving this balance requires navigating competing demands. Taylor-Ritzler and colleagues (2013) add that leadership engagement and organizational readiness are strong predictors of sustained capacity, but these

are often shaped by external environments that valorize compliance over reflection. This contradiction was clearly felt in this institutional context, where evaluation was simultaneously framed as a space for learning and a tool for accountability, often within the same meeting. These reflections are not presented as failures but as realities of evaluative practice in public sector systems. Preskill and Boyle (2008) argue that ECB must be contextually situated and flexible enough to adapt to institutional realities. In fact, the presence of tension may signal that evaluation is doing important cultural work, pushing against boundaries, inviting uncomfortable conversations, and creating new spaces for inquiry. Surfacing these tensions, rather than smoothing them over, is itself a capacity-building act, one that affirms the complexity of institutional change and the possibility of evaluation as both a governance and a learning practice.

## The Emergent Four-Stage Pathway to Evaluation Capacity Building

Viewed together over time, and within this leadership-enabled context, an implicit-to-explicit four-stage pathway becomes visible (see Table 3). The evaluative practices—carried through supportive structures, peer learning, and practice-based mechanisms—and planning and decision-making processes, moved from being latent within routine governance processes to being recognized, scaffolded through structures and feedback loops, and ultimately stabilized as part of the institution’s evaluative culture. This progression is summarized as four stages: latent routines, naming and framing evaluative practice, scaffolding and feedback loops, and consolidation and diffusion, each organized around a guiding analytic logic that structures the analysis and supports transferability across contexts.

Table 3. An Implicit-to-Explicit Evaluation Capacity Building (ECB) Pathway

| ECB stage                                 | Analytic description                                                                                                  | Enabling conditions                                                         | Illustrative routines (case)                                                                 |
|-------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| Stage 1<br>Latent evaluative routines     | Evaluative activity embedded in routine governance but primarily oriented toward accountability rather than learning. | Routine reporting requirements; compliance-oriented governance.             | Board and management meetings focused on performance reporting; routine monitoring.          |
| Stage 2<br>Naming and framing             | Existing practices reinterpreted as evaluative; meaning shifts before formal structural change.                       | Leadership orientation toward inquiry; emerging shared evaluative language. | Learning questions introduced in meetings; task forces interpreting findings.                |
| Stage 3<br>Scaffolding and feedback loops | Evaluative practice stabilized through structures and routines.                                                       | Governance structures; explicit feedback loops.                             | Regularized reviews; leadership retreats.                                                    |
| Stage 4<br>Consolidation and diffusion    | Evaluation embedded in core systems and diffused across the organization and organizational routines.                 | Aligned systems; distributed leadership.                                    | Evidence-informed work plans and budgets; procurement prioritization; quarterly adjustments. |

Note. The pathway is iterative; leadership practices, peer learning networks, and tensions between accountability and learning operate across all stages

In its earliest form, evaluation capacity exists primarily as *latent routines* embedded within ordinary governance processes. Attention at this stage is directed toward the evaluative work already taking place through routine organizational practices, even when these activities are not explicitly named or understood as evaluation. Board meetings, management reviews, reporting requirements, and informal discussions create spaces where evidence is generated, interpreted, and used, even though these activities are not recognized as evaluative in intent. At this stage, evaluative work is largely incidental to accountability and oversight functions rather than deliberately oriented toward learning. As the literature suggests, such latent evaluative activity is common in organizations and often goes unrecognized precisely because it is embedded in routine practice rather than formal ECB initiatives (Cousins et al., 2014; Searle et al., 2024).

Over time, shifts in leadership orientation toward inquiry and learning, alongside changes in institutional discourse about the role of evidence in decision-making, enables a transition

toward *naming and framing evaluative practice*. At this stage, analytic attention turns to how existing organizational practices come to be interpreted, named, and framed as evaluative. Routine governance forums begin to be interpreted differently, not merely as sites of compliance but as opportunities for inquiry, learning, and evidence-informed decision-making. This reframing is always evident in how leaders engage with data, the questions posed during meetings, and the emergence of shared evaluative language within task forces and working groups. Importantly, meaning shifts before structure: Practices remain largely the same, but their purpose and value is increasingly articulated in evaluative terms. This stage marks a critical turning point, as recognizing evaluation makes it possible to act intentionally in support of it.

As expectations for evaluative practice grow, leadership responds by putting in place *scaffolding and feedback loops* that support and sustain evaluation over time. The analysis at this stage focuses on how evaluative practice is supported and stabilized through organizational

structures and routines. Regularized governance meetings, leadership retreats, and standing reporting requirements provide the infrastructure necessary to stabilize evaluative practice. These structures reduce reliance on individual champions, create predictable spaces for reflection, and link evidence more systematically to planning and decision-making. Feedback loops among data, discussion, and action become more explicit, reinforcing evaluation as a normal and expected part of institutional life.

Finally, through continued use and reinforcement, evaluative practices move toward *consolidation and diffusion*. Here, attention shifts to how evaluative practice becomes stabilized, integrated, and diffused across the institution over time. Evaluation becomes embedded within core systems, strategic planning, budgeting, procurement, and performance management, such that it functions as a governing logic rather than a discrete activity. Evaluative thinking spreads beyond initial sites and actors, supported by practice-based learning arrangements and supportive leadership practices. At this stage, evaluation is no longer dependent on a particular individual or moment; it has become part of how the institution understands itself, makes decisions, and navigates competing demands for accountability and learning.

Across all four stages, leadership operates as a critical enabling condition shaping how routines are interpreted, which practices are supported, and how tensions between compliance and learning are navigated. Likewise, the tensions that surface between accountability and reflection, performance signaling and adaptive learning do not interrupt the pathway but actively shape it. As the ECB literature suggests, such tensions are not aberrations but constitutive elements of capacity-building processes (Goodyear, 2024). While the four-stage pathway is not intended as a linear maturity model, it can be read heuristically as representing indicative levels of evaluative culture maturity. To support practical application, a simple ECB culture maturity rubric is provided in Appendix A.

## Looking Forward: Insights, Implications, and Future Directions

This analysis moves beyond retrospective interpretation to offer forward-looking insights into how evaluation capacity building (ECB) can

be reimagined when rooted in everyday organizational practice rather than formalized strategies alone. One key contribution is the demonstration that ECB does not always have to begin with an explicit capacity-building plan; instead, it can emerge incrementally through routine governance practices, leadership behaviors, peer interactions, and decision-making spaces that normalize evaluative thinking over time.

It therefore supports a reconsideration of how evaluation capacity is conceptualized. Dominant models tend to emphasize technical skills, training workshops, and formal evaluation roles. While these remain valuable, they risk overlooking the less visible but equally powerful forms of capacity that reside in norms, relationships, and everyday governance practices. In this case, it was the routinization of reflection, through board meetings, retreats, collaborative planning, and data-informed dialogue, that gradually embedded evaluation as an institutional habit. These practices were not initially named as ECB, yet they carried its essence.

For practitioners and institutional leaders, the implications for practice are evident: Fostering evaluation culture does not always require starting from scratch. Instead, it may involve recognizing, legitimizing, and strengthening what is already happening. Leadership, in particular, plays a catalytic role by making evaluative spaces routine, valued, and psychologically safe. Additionally, institutions must be prepared to navigate the dual demands of accountability and learning without allowing one to crowd out the other. This balance requires intentionality, especially in public sector environments where external reporting can easily dominate the agenda.

For scholars, this case contributes to the growing body of literature that calls for more grounded, context-sensitive understandings of ECB. There is a need for research that explores how evaluation capacity unfolds in different types of institutions, under varying pressures, and through nonlinear, emergent pathways. Future studies might examine how frontline staff experience these embedded practices, or how evaluative cultures are sustained through leadership transitions and policy changes. Methodologically, there is also space for more reflexive, insider-informed approaches, especially in contexts where formal ECB language may be absent, but evaluative thinking is nonetheless alive. Future research could also empirically examine the implicit-to-explicit

pathway proposed here, testing its applicability, sequencing, and adaptability across institutional contexts, governance arrangements, and sectors.

## Conclusion

This reflection advances an implicit-to-explicit pathway for evaluation capacity building (ECB), demonstrating how ordinary governance routines can be progressively named, scaffolded through organizational structures and feedback processes, and stabilized into a sustained evaluative culture. Rather than treating ECB as a discrete intervention or training activity, the pathway illustrates how evaluative capacity can emerge through what is practiced, repeated, and reinforced in agendas, conversations, decisions, and institutional expectations.

A central insight from this analysis is that clarifying and naming implicit evaluative practices is a critical step in building evaluation capacity. In this case, many practices that contributed to learning and evidence use were not initially labeled as ECB, yet they reflected principles long emphasized in the literature (Preskill & Boyle, 2008; Patton, 2011; Labin et al., 2012). In institutional contexts shaped by time constraints, hierarchical authority, and strong accountability demands, such implicit pathways may represent a pragmatic and efficient starting point for ECB, rather than an alternative to participatory or empowerment-oriented approaches.

This paper therefore contributes a theory-informed account of how evaluation capacity can emerge under real-world constraints, and how such emergence can create institutional foundations, shared language, routines, and expectations that make more explicit and participatory forms of ECB increasingly feasible over time. Under this interpretation, implicit ECB is not a substitute for formal capacity-building initiatives but a complementary and often necessary precursor within complex governance environments.

## AI Disclosure

Portions of this manuscript, particularly in the structuring and refinement of sentences, benefited from the use of OpenAI's ChatGPT. The tool was employed to assist with clarity, conciseness, and style, but all substantive ideas, arguments, and interpretations are the author's own. The authors reviewed, edited, and takes full responsibility for the final text.

## References

- Argyris, C., & Schön, D. A. (1997). Organizational learning: A theory of action perspective. *Reis*, 77/78, 345–348.
- Bamberger, M., Vaessen, J., & Raimondo, E. (Eds.). (2015). *Dealing with complexity in development evaluation: A practical approach*. Sage Publications.
- Boswell, C. (2009). *The political uses of expert knowledge: Immigration policy and social research*. Cambridge University Press.
- Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*, 36(3), 375–388. <https://doi.org/10.1177/1098214015581706>
- Chaytor, K. (2023). Building evaluation culture—The missing link. *Canadian Journal of Program Evaluation*, 37(3), 422–438. <https://doi.org/10.3138/cjpe.75431>
- Christie, C. A., & Fierro, L. A. (2012). Evaluation policy to implementation: An evaluation of scientifically based research in practice. *Studies in Educational Evaluation*, 38(2), 65–72. <https://doi.org/10.1016/j.stueduc.2012.05.003>
- Compton, D. W., & Baizerman, M. (2007). Defining evaluation capacity building. *American Journal of Evaluation*, 28(1), 118–119. <https://doi.org/10.1177/1098214006298172>
- Cousins, J. B., Goh, S. C., Clark, S., & Lee, L. E. (2004). Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base. *Canadian Journal of Program Evaluation*, 19(2), 99–141. <https://doi.org/10.3138/cjpe.19.006>
- Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. *New Directions for Evaluation*, 141, 7–23. <https://doi.org/10.1002/ev.20076>
- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, 80, 5–23.
- Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, 48(1), 94–118. <https://doi.org/10.2307/3556620>

- Fetterman, D. M. (2001). *Foundations of empowerment evaluation*. Sage.
- Fetterman, D. M. (2023). *Empowerment evaluation and social justice: Confronting the culture of silence*. Guilford Publications.
- Fetterman, D. M., Rodriguez-Campos, L., & Zukoski, A. P. (2018). Collaborative, participatory and empowerment evaluation: Stakeholder involvement approaches. *Health Promotion Journal of Australia*, 29(3), 363.
- Fetterman, D. M., & Wandersman, A. (Eds.). (2005). *Empowerment evaluation principles in practice*. Guilford Press.
- Fierro, L. A., Bourgeois, I., Gokiart, R., Searle, M., & Tremblay, M. (2024). Strengthening evaluation capacity building practice through competition: The Max Bell School of Public Policy's Evaluation Capacity Case Challenge. *Canadian Journal of Program Evaluation*, 38(3), 521–531. <https://doi.org/10.3138/cjpe-2024-0012>
- Goodyear, L., Mansori, S., Quinlan, K., McMahon, T., & Kyles, C. M. (2024). Evaluation capacity building in theory and practice: Revisiting models from practitioner perspectives. *New Directions for Evaluation*, 183, 29–41. <https://doi.org/10.1002/ev.20613>
- Holmes, A. G. D. (2020). Researcher positionality—A consideration of its influence and place in qualitative research—A new researcher guide. *Shanlax International Journal of Education*, 8(4), 1–10.
- Johnson, G., Langley, A., Melin, L., & Whittington, R. (Eds.). (2007). *Strategy as practice: Research directions and resources*. Cambridge University Press.
- Konjore, M. J. (2024). Two decades of evaluating evaluation capacity building: A bibliographic coupling review. *New Directions for Evaluation*, 183, 11–28. <https://doi.org/10.1002/ev.20616>
- Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, 33(3), 307–338. <https://doi.org/10.1177/1098214011434608>
- Lee, L. E., & Cousins, J. B. (2004). Participation in evaluation of funded school improvement: Effects and supporting conditions. In *Participatory Evaluation in Education* (pp. 87–100). Routledge.
- Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public service*. Russell Sage Foundation.
- Mayne, J. (2008). Contribution analysis: An approach to exploring cause and effect (ILAC Brief No. 16). Institutional Learning and Change (ILAC) Initiative, Bioversity International. <https://hdl.handle.net/10568/70124>
- McDonald, B., Rogers, P., & Kefford, B. (2003). Teaching people to fish? Building the evaluation capability of public sector organizations. *Evaluation*, 9(1), 9–29. <https://doi.org/10.1177/1356389003009001002>
- Nielsen, S. B., Fierro, L. A., Bourgeois, I., & Lemire, S. (2024). The science and practice of evaluation capacity building. *Canadian Journal of Program Evaluation*, 38(3), 425–432. <https://doi.org/10.3138/cjpe-2024-0011>
- Norton, S., Milat, A., Edwards, B., & Giffin, M. (2016). Narrative review of strategies by organizations for building evaluation capacity. *Evaluation and program planning*, 58, 1–19. <https://doi.org/10.1016/j.evalprogplan.2016.04.004>
- Nutley, S. M., Walter, I., & Davies, H. T. (2007). Improving research use in policy contexts. In *Using evidence* (pp. 231–270). Policy Press. <https://doi.org/10.51952/9781847422323.ch008>
- Patton, M. Q. (2008). *Utilization-focused evaluation*. Sage Publications.
- Patton, M. Q. (2011). *Essentials of utilization-focused evaluation*. Sage Publications.
- Pettigrew, A., & McNulty, T. (1995). Power and influence in and around the boardroom. *Human Relations*, 48(8), 845–873. <https://doi.org/10.1177/001872679504800802>
- Pollitt, C., & Bouckaert, G. (2017). *Public management reform: A comparative analysis—Into the age of austerity*. Oxford University Press.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443–459. <https://doi.org/10.1177/1098214008324182>
- Schein, E. H. (2010). *Organizational culture and leadership* (Vol. 2). John Wiley & Sons.



- Schön, D. A. (2017). *The reflective practitioner: How professionals think in action*. Routledge. (Original work published 1983)
- Searle, M., Fierro, L. A., Pinarski, J., Dixon, L., Tremblay, M., Bourgeois, I., & Gokiert, R. (2024). Meeting the challenge: How the City of Kingston is working to propel evaluation growth. *Canadian Journal of Program Evaluation*, 38(3), 555–565. <https://doi.org/10.3138/cjpe-2024-0006>
- Stake, R. E. (1995). *The art of case study research*. Sage Publications.
- Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, 93, 1–25.
- Suarez-Balcazar, Y., & Taylor-Ritzler, T. (2013). Moving from science to practice in evaluation capacity building. *American Journal of Evaluation*, 35(1), 95–99. <https://doi.org/10.1177/1098214013499440>
- Taut, S. (2007). Studying self-evaluation capacity building in a large international development organization. *American Journal of Evaluation*, 28(1), 45–59. <https://doi.org/10.1177/1098214006296430>
- Taylor-Ritzler, T., Suarez-Balcazar, Y., Garcia-Iriarte, E., Henry, D. B., & Balcazar, F. E. (2013). Understanding and measuring evaluation capacity: A model and instrument validation study. *American Journal of Evaluation*, 34(2), 190–206. <https://doi.org/10.1177/1098214012471421>
- Weick, K. E., & Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3, No. 10.1002). Sage Publications.
- Weiss, C. H. (2021). The many meanings of research utilization. In *Social science and social policy* (pp. 31–40). Routledge.
- Weiss, C. H. (1979). THE MANY MEANINGS OF RESEARCH UTILIZATION. *Public Administration Review*, 39(5). <https://doi.org/10.2307/3109916>
- Yip, S. Y. (2024). Positionality and reflexivity: Negotiating insider-outsider positions within and across cultures. *International Journal of Research & Method in Education*, 47(3), 222–232. <https://doi.org/10.1080/1743727X.2023.2266375>

## Appendix A. ECB Culture Maturity Rubric

| Indicative level | Characteristics of evaluative culture                                                        |
|------------------|----------------------------------------------------------------------------------------------|
| Latent           | Evidence use embedded in accountability routines; evaluative activity implicit and unnamed.  |
| Emergent         | Evaluation recognized as inquiry and learning; shared evaluative language begins to develop. |
| Supported        | Structures and feedback loops stabilize evaluative practice across governance processes.     |
| Embedded         | Evaluation integrated into core systems and functions as a governing logic.                  |