DOI: 10.1002/ev.20616

ORIGINAL ARTICLE

WILEY

Two decades of evaluating evaluation capacity building: A bibliographic coupling review

Mervyan J. Konjore

Division of Behavioral and Organizational Sciences, Claremont Graduate University, Claremont, California, USA

Correspondence
Mervyan J. Konjore, Division of Behavioral and Organizational Sciences, Claremont Graduate University, Claremont, CA, USA.
Email: mervyan.konjore@cgu.edu

Abstract

Evaluation capacity building (ECB) has emerged as a dominant tool in evaluation. Several reviews have taken stock of ECB over the past 20 years, examining how understanding of ECB has developed over time. By examining this existing research on the evaluation of ECB design, effectiveness, measurement, integration, and sustainability, opportunities to advance the field's existing knowledge base around ECB evaluation forward become apparent. This landscape review uses bibliographic coupling and content analysis to offer a robust analysis of publications on the evaluation of ECB. The results of this review of multidisciplinary ECB publications shed light on areas of strength within the existing literature and identify opportunities to build upon current scholarship. Promising areas for future research include exploring how moderators and mediators influence ECB outcomes, eliminating threats to internal validity in ECB studies, and strengthening the rigor required for external generalizability.

INTRODUCTION

Conceptual models and systematic reviews provide guidelines for designing and implementing evaluation capacity building (ECB) initiatives (Preskill & Boyle, 2008). They have also identified common principles and strategies for both individual and organizational capacity building, and emphasized the importance of collaborative processes as integral to ECB practice (Labin et al., 2012). Reviews of published literature have found successful strategies for implementing and sustaining ECB, including conducting needs assessments, reinforcing organizational commitments to ECB and evaluation, integrating practical learning strategies, and providing follow-up workplace support (Norton et al., 2016).

I would like to thank Bradley Cousins, Becky Reichard, Jaymes Rombaoa, and Prina Patel for reviewing earlier drafts of the article.

However, there is still much to learn about the landscape of ECB's impact and how to go about evaluating ECB initiatives.

In a recent review, Bourgeois and colleagues (2023) assert that in practice, ECB practitioners' implementation of ECB is characterized by reliance on a limited number of ECB strategies, the outcomes of which are rarely described. Evaluations of ECB are rarely published or shared publicly, making for a limited understanding of what works when it comes to strengthening people's evaluation capacity. Some have made observations about the effectiveness of ECB initiatives. Bourgeois and colleagues reported that changes in individual knowledge, skills, and behavior materialize only half of the time, and that there appear to be mixed results when ECB lacks a practical component (Bourgeois et al., 2023). Suarez-Balcazar and Taylor-Ritzler (2014) have argued that measuring ECB does not always lead to gains in programming, improved services, better outcomes for beneficiaries, or sustainable capacity.

This presents an opportunity to better understand the existing literature around the designs and findings of ECB evaluation initiatives. Greater knowledge about evaluations of ECB can lead to better understanding of how to evaluate ECB initiatives and guide future decades' research agendas. It can also reveal areas of knowledge that can further be strengthened. By "greater knowledge about evaluations of ECB" I specifically mean research on the evaluation of ECB design (Preskill & Boyle, 2008), effectiveness (Struyk et al., 2011), mainstreaming (Labin, 2014), sustainability (Suarez-Balcazar & Taylor-Ritzler, 2014), and measurement (LaFond et al., 2002).

In this article, I add to the literature by reviewing publications on the evaluation of ECB initiatives. By coupling articles that draw on similar sources, I offer a nuanced assessment of peer-reviewed evaluations of ECB over the past two decades, and by using technology to leverage the "big data" of citation indices I offer a robust analysis of the evaluation of ECB. I systematically review the multidisciplinary evaluation literature to identify and map the relationships among peer-reviewed articles that discuss the evaluation of ECB initiatives over the past two decades (Cobo et al., 2011; Zupic & Čater, 2015). I then identify themes across the articles to clarify how evaluations on ECB are being conducted, acknowledge current findings, and highlight opportunities for future scholarship.

PURPOSE AND SCOPE

In this article, I aim to take stock of current knowledge about the evaluation of ECB, highlight gaps in current knowledge, and identify promising directions for further scholarship. My review of multidisciplinary evaluation literature sheds light on areas of strength within the existing literature and identifies opportunities to build upon current scholarship. Since current literature uses multiple definitions of ECB (Milstein et al., 2002; Stevenson et al., 2002; Stockdill et al., 2002), I have refrained from adopting a single definition of ECB, instead using a broad set of inclusion criteria in order to maximize the number of relevant studies included in the analysis.

Research questions

The following questions guided this study:

1) What patterns exist within the publications on the evaluation of ECB?

2) What are the strengths revealed by publications on the evaluation of ECB? Where are the opportunities to continue to build upon the field's knowledge of effective ECB?

METHODS AND PROCEDURES

This study had two phases. The first phase identified relevant articles to include in the review and analyzed relationships between them. The second phase reviewed and synthesized findings from the included articles.

Phase 1

This phase included two steps. First, to identify relevant articles, I conducted a systematic search of evaluation literature. Second and to answer Research Question 1, I used a form of bibliometric analysis called bibliographic coupling. Bibliometric analyses handle large amounts of literature to provide a nuanced summary of a research field (Donthu et al., 2021) by uncovering a broad spectrum of themes indicative of emergent topics and potential future developments (Vogel et al., 2021). Specifically, they map relationships among citing publications to reveal thematic clusters that underpin a field's knowledge structure and its development through time (Donthu et al., 2021). Khare and Jain (2022, p. 568) define knowledge structure as the "hidden patterns of an academic field" that might relate to authorship or publications on a topic, themes or patterns discussed, or interactions and collaborations among authors or researchers.

Bibliographic coupling, specifically, analyzes relationships among citing publications (Donthu et al., 2021). Bibliographic coupling examines primary documents that cite the same secondary documents, assuming that two documents that share common references are also similar in their content (Zupic & Čater, 2015). Primary documents are those returned from key searches, whereas secondary documents are the references used within the primary documents (Bronk et al., 2023).

Donthu et al. (2021) differentiate bibliographic coupling from other bibliometric analyses, which analyze relationships among the most influential publications to understand the foundation of a field (co-citation analysis), explore relationships among topics by focusing on the written content of publications (co-word analysis), or assess social interactions and relationships among authors by analyzing author affiliations (co-authorship analysis). As a science mapping technique, bibliographic coupling operates on the assumption that two primary documents that cite the same secondary document (Figure 1) share a common interest and are more connected and more similar in their content the more references they have in common (Donthu et al., 2021; Vogel et al., 2021; Zupic & Čater, 2015). The coupling strength between primary documents is determined by how frequently they cite the same secondary documents (Vogel et al., 2021), with overlapping bibliographies indicating stronger similarities between documents. Thus, the coupling strength increases as two primary documents have more secondary documents in common. A higher coupling strength is, therefore, indicative of a more important relationship between two primary documents (Ma et al., 2022).

Search process

To identify publications related to research on or evaluations of ECB, I conducted a Web of Science search using the following search phrases: "evaluat* capacity," "evaluat* capacity build*," "build* evaluat* capacity," "evaluat* capacity develop*," OR "develop evaluat* capacity." The search occurred on March 30, 2023, and yielded 549 records.

Of these primary records, 63% (315) were from five top Web of Science categories: the social sciences; health; education research; engineering; and electrical, electronic, and

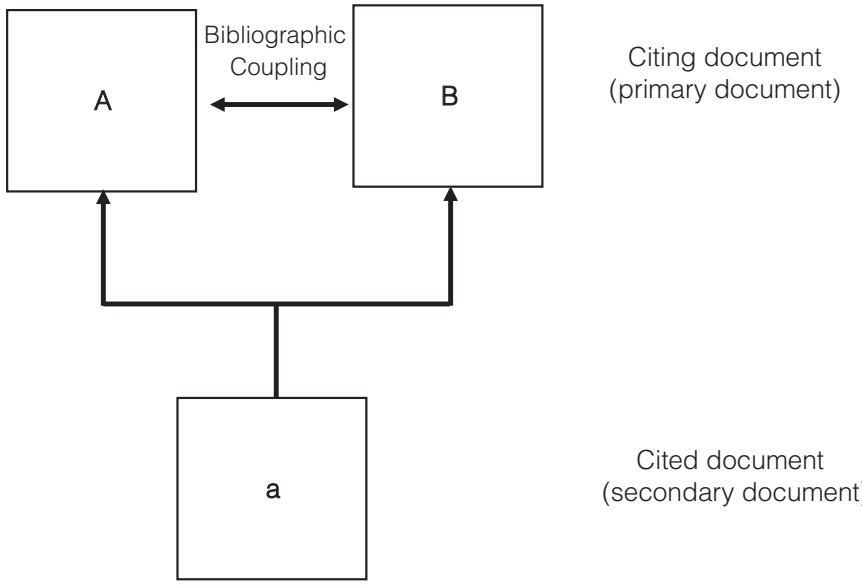

FIGURE 1 Bibliographic coupling method.

telecommunications fields. Four top evaluation publications accounted for 27.1% (149) of the records: *American Journal of Evaluation*, *Evaluation and Program Planning*, and *Canadian Journal of Program Evaluation*, and *Evaluation*.

Analysis process

I imported these search results into VOSviewer to produce a two-dimensional map of the nodes (documents) in the network (Cobo et al., 2011; Van Eck & Waltman, 2014). I then set a minimum citation threshold, meaning the lowest number of citations or secondary documents that coupled primary documents would have in common (Nájera-Sánchez et al., 2020). After testing various thresholds, I selected a threshold of one. I chose this low citation threshold to ensure the review captured the emerging nature of the literature on the evaluation of ECB, while also emphasizing documents that were more influential in the literature (Zupic & Čater, 2015).

This process produced 372 primary documents that met the minimum citation threshold, of which only 277 were connected. I excluded those that were isolated and consequently less connected to evaluating ECB (Cobo et al., 2011). I reviewed the connected documents to ensure they aligned with the inclusion criteria that isolate publications on the evaluation of ECB. I excluded those that did not include content related to the evaluation of ECB, leaving 53 articles organized into nine clusters. Since the purpose of clusters is to offer an overview of the attributes and relationships among documents within a cluster (Van Eck & Waltman, 2014), I also excluded three clusters that contained only one secondary document each; they did not contain enough documents to make the interpretation of relationships within the cluster possible. This resulted in 50 publications being retained for analysis and included in the coupling map. Figure 2 shows a flow diagram of the included studies.

To account for primary documents with extensive secondary sources, I constructed the coupling map using the fractional counting method. I gave each citing document equal weight in the review, regardless of the number of its references. Thus, secondary sources are

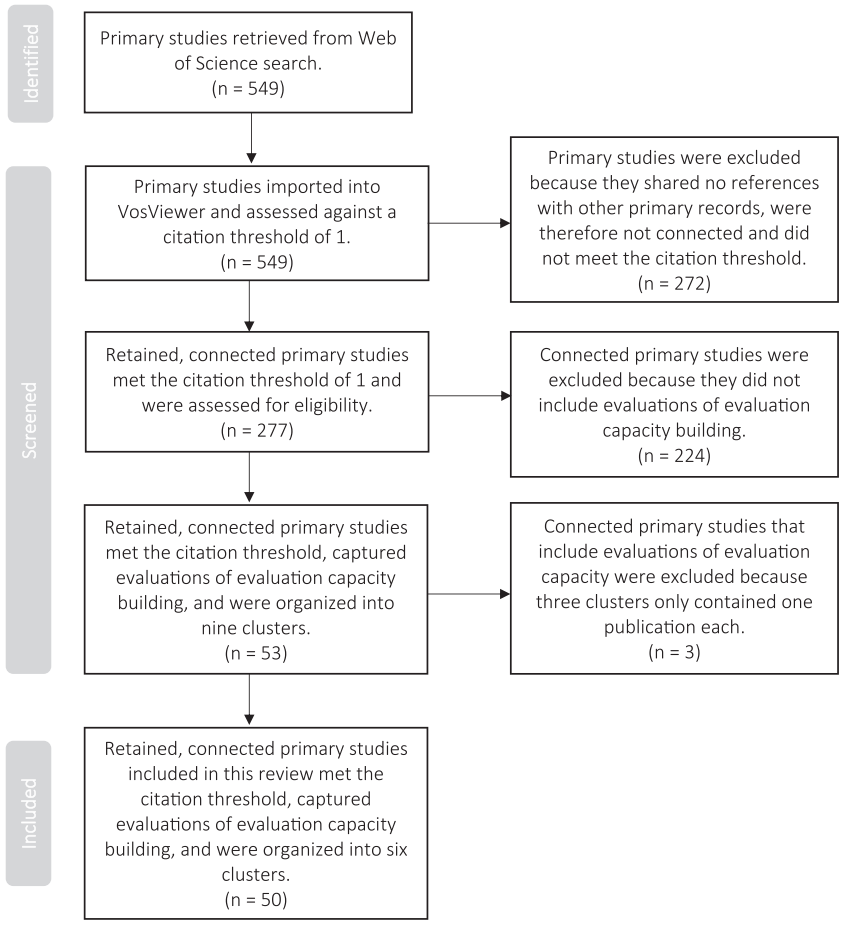

FIGURE 2 Flow diagram of included studies.

weighted equally, irrespective of whether they were highly cited or not (Perianes-Rodriguez et al., 2016). For example, a secondary paper that has been cited by five other primary documents carries the same weight as one that has been cited by 500 other primary documents, which means that the former will have a 1/5 and the latter a 1/500 fractional counting weight.

To enhance understanding and interpretation of the coupling map, I merged small clusters based on a minimum cluster size of one, set the map's visualization to total link strength, and left the default resolution parameter unchanged at 1.00 to avoid higher resolution values that would result in more clusters (Van Eck & Waltman, 2014).

Phase 2

In the second phase of this study, I conducted a conventional content analysis of the 50 included articles to identify themes and allow for new insights to emerge (Hsieh & Shannon, 2005). To further answer Research Question 1, I assessed the bibliographic data of the included studies. I reviewed the included articles' abstracts and coded the ECB focus of each study to create a label for each bibliographic coupling map cluster.

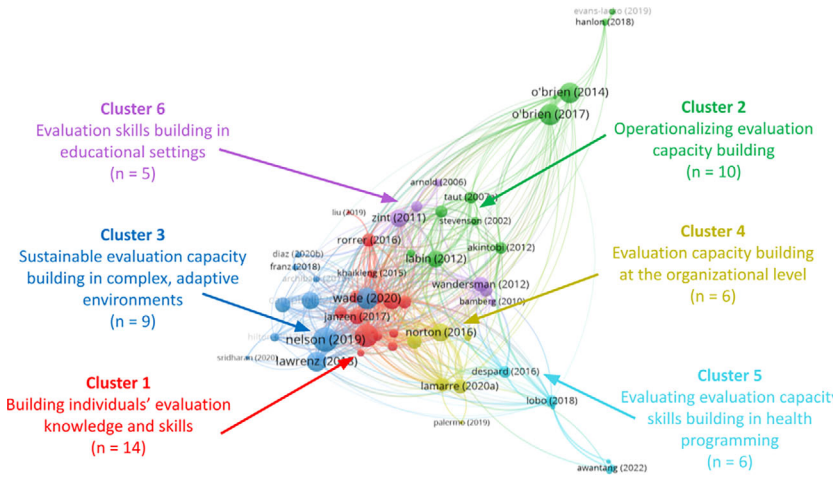

FIGURE 3 Bibliographic coupling map of evaluations on ECB over the past two decades. *Note*: Node size corresponds to total link strength. With each cluster, larger nodes represent stronger total link strength. Nodes that are closer to each other are more connected or related. Node colors signify six distinct clusters of documents on evaluating ECB.

To answer Research Question 2, I systematically recorded pertinent information from the included 37 articles that dealt with the evaluation of ECB initiatives, excluding those that focused on the measurement of evaluation capacity. For each article, I recorded: descriptive information (i.e., year of publication, source of publication, country of implementation); methodological content (i.e., preassessments, operationalization of variables, evaluation designs and methods); types of ECB initiatives (i.e., objectives, definition adopted, intervention, study population, and who provided the ECB); and reported outcomes. I collected this information in an Excel spreadsheet. I then conducted a content analysis to reveal strengths and opportunities in the literature and identify directions for future scholarship on the evaluation of ECB.

RESULTS

Patterns of publications in ECB evaluations

The bibliographic coupling map shows the relationships among publications discussing similar and different aspects related to ECB evaluation. It also suggests the extent to which various aspects are discussed, referred to here as importance and influence. The bibliographic coupling map consists of six clusters and provides a high-level overview of the patterns of evaluations that have assessed ECB over the past two decades (Figure 3).

Cluster 1: Building individuals' evaluation knowledge and skills

Cluster 1 is the largest, meaning it contains the most documents. With 14 articles, Cluster 1 publications focus on building individuals' evaluation knowledge and skills (i.e., ECB outcomes; Table 1). With the highest number of citations (391), that is, the most secondary documents or references that primary documents have in common, Cluster 1 represents the most influential focus (Zupic & Čater, 2015) of published literature on the evaluation of

TABLE 1 Bibliographic coupling overview by cluster.

Cluster	Cluster title	*N*	Number of citations	Total citation strength	Median year (range) of publication
Cluster 1	Building individuals' evaluation knowledge and skills	14	391	103	2017 (2014–2022)
Cluster 2	Operationalizing evaluation capacity building	10	201	276	2012 (2002–2020)
Cluster 3	Sustainable evaluation capacity building in complex, adaptive environments	9	219	45	2018 (2014–2020)
Cluster 4	Evaluation capacity building at the organizational level	6	140	52	2019 (2008–2022)
Cluster 5	Evaluating evaluation skills capacity building in health programming	6	57	38	2017 (2014–2022)
Cluster 6	Evaluation skills building in educational settings	5	83	113	2010 (2006–2012)

ECB. However, its total link strength, indicative of the strength of the relationship between primary documents and their references, is 103—lower than that of Clusters 2 and 5. This suggests that documents in Cluster 1 are not as related and may have less in common when compared to those in Clusters 2 and 5. Primary documents that have fewer references in common are less connected.

Cluster 2: Operationalizing ECB

Cluster 2 is the second-largest cluster, with 10 documents. It has 201 citations and a total link strength of 276. This makes Cluster 2 the third most influential area of interest for evaluating ECB, with its documents being the most connected and similar in content when compared to any of the other clusters. Conversations in Cluster 2 focus on operationalizing ECB, and an examination of their dispersion shows two distinct subclusters that operationalize individual and organizational outcomes in schools and organizations.

Cluster 3: Complex contexts

Cluster 3 contains nine documents that discuss sustainable ECB in complex, adaptive environments such as Extension settings, research networks, and multi-site programs. With 219 citations, Cluster 3 is the second most influential cluster. However, the size and proximity of the nodes indicate conversations are related but not closely connected. At 45, Cluster 3's total link strength is low relative to the other clusters.

Cluster 4: ECB at the organizational level

Cluster 4 has six articles, is the fourth most influential focus for evaluating ECB, and emphasizes ECB at organizational levels. The distribution and proximity of the nodes indicate two subclusters, emphasizing ECB for programs and within organizations. Cluster 4 has a total link strength of 52.

Cluster 5: Evaluating ECB in health programming

Nodes that are further away from each other are less related. With six documents in Cluster 5, the distance between the nodes suggests that the documents are somewhat isolated even though they all discuss ECB in health programming. Cluster 5 has both the lowest citation score (57) and the lowest total link strength (38), which may indicate this cluster's emerging nature.

Cluster 6: Evaluation skills building in educational settings

Cluster 6 contains five documents. With a total link strength of 113, documents in this cluster are the second most connected and similar in their content (after Cluster 2), and this cluster's total of 83 citations demonstrates its recent focus on evaluation skills building in educational settings. A low citation count is also indicative of the newness of a topic.

Reflecting on pattern trends

Overall, publication trends show how conversations developed within the field and shed light on when various conversations became more outwardly prevalent. Publication trends also point to areas for continued or future focus within the evaluation field. In Table 1, the range of publication years emphasizes that publications have predominantly focused on operationalizing ECB, followed by the evaluation of ECB at the organizational level. Further analysis of the bibliographic coupling map offers additional insights into publication patterns related to evaluation of ECB over the past two decades. For example,

- In Cluster 2, publications focus on operationalizing ECB and represent the earliest published work as well as the most extended period of attention (see Table 1). With 2012 as the median publication year, these publications cover a period of 18 years from 2002 to 2020, which peaked around 2011–2013 (Figure 4).
- Emphasizing the evaluation of ECB at the organizational level, Table 1 shows that publications in Cluster 4 have a median year of 2019 and span a period of 14 years, from 2008 to 2022. Cluster 4 is also the second-oldest cluster. Figure 4 portrays a linear increase from 2005 to 2010, a decline in 2011–2013 before it increases and eventually plateaus from 2017.
- Clusters 1 and 5 both have 2017 as their median publication year (Table 1), and publications in both clusters have extended over a period of 8 years from 2014 to 2022, respectively, focusing on building individuals' evaluation knowledge and skills and evaluating evaluation skills capacity building in health programming. Figure 4 shows that publications in both clusters peaked in 2014–2016 to decline sharply by 2020–2022.

Overall, patterns of the publications on the evaluation of ECB show a steady increase from 2002–2004 to 2011–2013, followed by a sharp rise in 2014–2016 as conversations in the evaluation of ECB proliferated and plateaued by 2017–2019.

Strengths suggested by the patterns in ECB evaluations

The results presented in this section represent a broad overview, summarizing what the analyzed articles suggest about evaluation approaches and findings of ECB activities. These

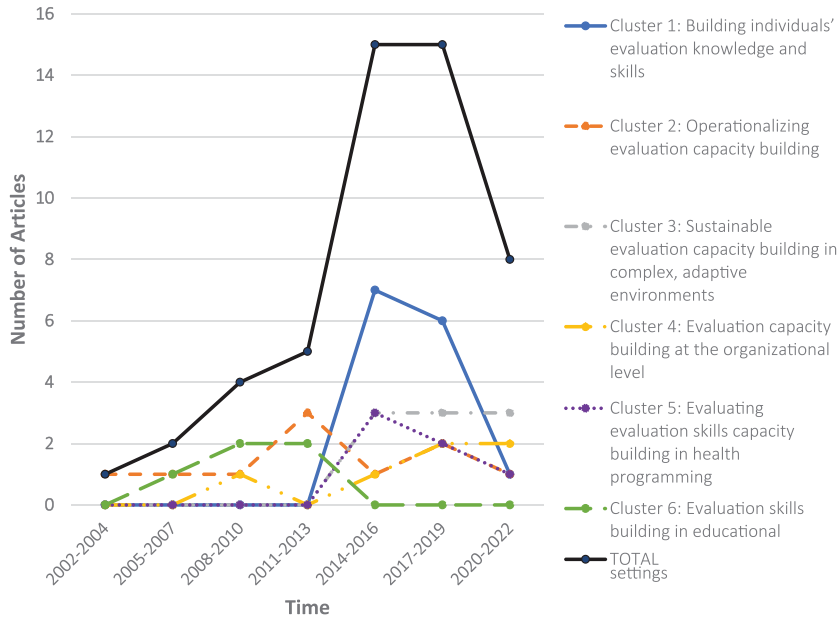

FIGURE 4 Publication dates of articles on evaluating evaluation capacity building from 2002 to 2022 ($n = 50$).

results point to evidence-based or best practices while also highlighting areas where further exploration would be helpful. While the field continues to build knowledge about the evaluation of ECB, this section provides broad guidance to consider when practitioners and researchers implement or evaluate ECB, respectively. Although this study's analyses include publications that examine many topics, this section predominantly focuses on those related to evaluating ECB activities, including training activities, technical assistance, mentorship, and hands-on involvement in evaluations.

Evaluation designs and methods

Evaluations of ECB activities typically use case studies (Janzen et al., 2017; Taut, 2007) or mixed-methods approaches measuring change at multiple time points (Akintobi et al., 2012; Lettenmaier et al., 2014; Ploeg et al., 2008; Stevenson et al., 2002). Given the context of most ECB initiatives, where participation in the initiative or in the evaluation is not mandatory, these evaluations see low response rates (Kaye-Tzadok & Spiro, 2016; Taut, 2007), participant self-selection (Kaye-Tzadok & Spiro, 2016; Lindeman et al., 2018), and small sample sizes (Bakken et al., 2014; Kaye-Tzadok & Spiro, 2016; Taut, 2007). For example, Bakken et al.'s evaluation of community health programs comprised six university students, each evaluating one program. In Taut's study, 47 staff participated in the workshop but only five completed follow-up self-evaluation projects. Additionally, most studies rely on self-report measures of evaluation capacity (Lindeman et al., 2018). While these challenges are not at all surprising given the systems ECB initiatives operate in, they may result in biased responses and difficulty identifying statistically significant relationships or generalizing to broader contexts (Kaye-Tzadok & Spiro, 2016; Taut, 2007).

Most published evaluations of ECB are unable to justify causal or attributional relationships given the reliance on post hoc designs and lack of established comparison groups. Even when more rigorous designs are used, recall bias resulting from retrospective designs

without multiple measurement points (Bakken et al., 2014); social desirability bias springing from working in close collaboration during the ECB process (Lindeman et al., 2018); and small samples (Kaye-Tzadok & Spiro, 2016; Lindeman et al., 2018) continue to threaten internal and external validity. More intentional, a priori designs for evaluations of ECB initiatives would provide more robust and trustworthy findings, allowing the field to coalesce around effective ECB practices.

Training workshops, courses, and seminars

Publications on the evaluation of ECB-focused training workshops, courses, and seminars identify several themes around aspects associated with the effectiveness of evaluating ECB.

From the publications it appears as though evaluation workshops (Janzen et al., 2017; Stevenson et al., 2002), courses (Khaikleng et al., 2015), and university seminars (Bakken et al., 2014; Kaye-Tzadok & Spiro, 2016) are effective modes of training that result in increased individual knowledge, attitudes, and skills around evaluation of ECB. However, an assessment of dosage is seldom included when ECB initiatives are evaluated, resulting in great variability in the duration of training, which ranges from hours (Khaikleng et al., 2015; Stevenson et al., 2002) to days (Akintobi et al., 2012; Janzen et al., 2017; Taut, 2007) and weeks (Bakken et al., 2014) to years (Kaye-Tzadok & Spiro, 2016). Such variability makes it hard to determine the dosage at which training is most effective.

Training has been shown to solidify individuals' knowledge and understanding of evaluation theory (Janzen et al., 2017) but to be less likely to have lasting effects on practice unless supported by technical assistance (Akintobi et al., 2012; Lettenmaier et al., 2014); mentorship (Janzen et al., 2017; Lettenmaier et al., 2014); or participation in evaluations that put knowledge into practice, transfer content to work routines, and build staff confidence (Janzen et al., 2017; Khaikleng et al., 2015; Taut, 2007).

When training is supported by technical assistance, obtaining buy-in is fundamental in building evaluation capacity within organizations (Akintobi et al., 2012). Where participation in evaluations follows training, Bakken et al. (2014) emphasize the roles of educators who model the requisite skills for conducting effective evaluations and the roles of their protégés acting as evaluation champions in ensuring evaluative efforts lead to organizational capacity. Other factors influencing training in ECB include staff and organizational resources (Akintobi et al., 2012).

Numerous opportunities remain to grow the field's understanding about the role of training in ECB. For example, future evaluations and research could examine lines of inquiry around adequate dosage, duration, and lasting effects of training events that are required for optimal effectiveness of ECB initiatives. While the literature has demonstrated that knowledge of evaluation gained through training activities is most effective when augmented by other strategies, questions remain about how these other ECB strategies mediate the effects of training, and about the effect and size of this mediating relationship. Additionally, questions remain about the influence of context and how the quality of interpersonal relationships between those supporting ECB and those being supported could influence ECB outcomes. The answers to these questions remain elusive because such effects are seldom systematically captured (Bourgeois et al., 2023) or measured.

Technical assistance

Technical assistance as an ECB activity refers to the provision of targeted assistance or support by an evaluator to a nonevaluator or emerging evaluator (Preskill & Boyle,

2008). Examples of technical assistance include onsite technical assistance and technical assistance rendered by phone. Technical assistance appears to be a promising area for further examination, with publications reporting a relationship between technical assistance and changes in evaluation capacity (Lettenmaier et al., 2014; Stevenson et al., 2002). Akintobi et al. (2012) and Lettenmaier et al. (2014) further observed that capacity outcomes were more promising when technical assistance was accompanied by skills-building in training or participation in evaluations.

Lettenmaier and colleagues (2014) also found an association between the provision of technical assistance and changes in confidence around evaluations. Through pre–post comparisons, they reported changes in knowledge and attitudes up to 6 months postattendance when technical assistance underscored training (Lettenmaier et al., 2014). When accompanied by program evaluations, technical assistance improved information sharing, problem diagnostics, and building staff relationships within organizations (Lettenmaier et al., 2014). But one area warranting further examination is the role sound interpersonal skills play in providing technical assistance. Those supporting ECB rely on interpersonal skills to maintain a careful balance between providing technical assistance and executing the tasks they are meant to support (Stevenson et al., 2002).

Technical assistance appears to support building evaluation capacity when activities respond to individuals' and organizations' evaluation needs and readiness for change and when technical assistance activities are provided as follow-on support (Akintobi et al., 2012; Lettenmaier et al., 2014; Stevenson et al., 2002). Lettenmaier and colleagues (2014) argue that technical assistance should be driven internally to be effective and enhance organizational ownership. Another area requiring further examination is that integrating the outcomes of evaluation technical assistance throughout organizations requires organizational and staff resources, in addition to staff involvement and buy-in (Akintobi et al., 2012). Unless supported with adequate resources, supportive evaluation policies and strategies (Lettenmaier et al., 2014), and commitment from leadership, sustaining evaluation capacity through technical assistance may be challenging at organizational levels (Stevenson et al., 2002).

Mentorship

Preskill and Boyle (2008, p. 447) define mentorship as "individualized technical and professional support" provided in relationship with evaluators. In the reviewed articles, mentoring as an ECB activity is generally paired with other activities, such as training, technical assistance, or participation in evaluations.

Literature has shown that in combination with other ECB activities, mentoring builds recipients' confidence in evaluating programs (Janzen et al., 2017) and facilitates the application of evaluation knowledge (Lettenmaier et al., 2014; Ploeg et al., 2008). Adequate resources, leadership, managerial systems, and shared values are required to effectively design and provide ECB mentorship (Lettenmaier et al., 2014). Other forms of organizational support that influence mentorship outcomes are infrastructure, policies, and practices that support evaluation, and training to mainstream evaluation skills within organizations (Ploeg et al., 2008).

Of all the examined publications, Ploeg and colleagues (2008) have the strongest focus on evaluating ECB mentoring. Through their evaluation, they found that attention to matching mentors and mentees, fostering relationships between them, and regularly reviewing mentorship relationships is important to ensure effectiveness (Ploeg et al., 2008). In their experience, providing mentorship also requires buy-in from mentees, involving

managers to clarify goals and expectations, and having organizational leadership invest in increased evaluation capacity (Ploeg et al., 2008). Allowing mentees to participate in evaluations and linking mentorship to time-bound evaluations relevant to their work contexts also promoted the mainstreaming of ECB efforts (Ploeg et al., 2008). Ploeg and colleagues did acknowledge that ECB mentoring develops an individual's knowledge and skills, putting ECB outcomes at risk in contexts of staff turnover. Therefore, mainstreaming evaluation capacity within the organizational structure and planning for staff turnover helps to sustain evaluation capacity.

As with other ECB activities, there remains an opportunity to understand more about the effects of relational dynamics between mentor and mentee, dosage (i.e., the number and duration of sessions), and the extent to which mentorship should either be incorporated into other ECB strategies or remain a standalone strategy to be effective in building ECB. There are also questions about whether different contexts moderate mentorship relationships and effectiveness differently.

Participation in evaluation

Building a culture for evaluation entails the organization taking ownership of evaluations (Bakken et al., 2014; Janzen et al., 2017; Taut, 2007). Participating in evaluations has been shown to improve comprehension of evaluative processes, engage evaluation partners (Janzen et al., 2017; Rorrer, 2016), and create value for the practice of evaluation (Bakken et al., 2014). Individuals engaging in evaluations acquire evaluation knowledge and improve skills in program planning and evaluative thinking (Taut, 2007).

Kaye-Tzadok and Spiro (2016) found that the perceived impact of participation in evaluation as an ECB activity on organizational evaluation capacity was stronger for organizations with no prior experience of evaluation, compared to those with previous experiences of evaluations. They also found that perceived ability was higher after implementing evaluations compared to only planning evaluations (Kaye-Tzadok & Spiro, 2016). Janzen and colleagues (2017) point to the evaluation timing influencing the organizational uptake of evaluations because programs in the process of implementation may be constrained by limited timeframes or forced to continue using preexisting tools to meet donor requirements. Mainstreaming participation in evaluations within organizations requires continuous implementation support (Taut, 2007) during both the planning and implementation of evaluations (Lindeman et al., 2018).

Supportive organizational cultures, structures, and processes that are characterized by trust, transparency, and learning are prerequisites for participants to learn from evaluations (Janzen et al., 2017; Taut, 2007). For example, a seminar supported by the implementation of an evaluation had a greater impact on an organization's interest and capacity when individuals felt supported by their organizations, consulted superiors and colleagues, and presented their evaluation plans and findings to them (Kaye-Tzadok & Spiro, 2016). Constraints that hamper the implementation of evaluations are lack of resources, skills, and incentives; staff workloads; perceptions that evaluation is not part of their duties; inadequate understanding of evaluation's role in learning; and the evaluator's status and credibility (Stevenson et al., 2002; Taut, 2007). Opportunities for future inquiry should consider examining the influence of supportive environments on evaluations of ECB by probing how organizational cultures characterized by trust, transparency, and learning moderate outcomes from evaluation of ECB.

DISCUSSION

This landscape review examines patterns in the published literature on the evaluation of ECB over the past two decades, and summarizes both the evaluation designs and methods used to assess ECB initiatives and the primary outcomes across ECB activities.

Patterns within ECB evaluations

Patterns within the publications on evaluation of ECB show that operationalizing ECB has been the most important area of interest. Published literature on operationalizing ECB contains some of the earliest published works, and has captured the attention of the field for the longest time. Building an individual's knowledge and skills has been the most influential focus of evaluations of ECB with publications on this topic being the most connected and similar in their content given the references they have in common.

Despite their relatively short, 6-year time span, publications in the cluster on evaluating skills building in educational settings are the second most connected and similar in their content, after the publications on operationalizing ECB. As Bourgeois and colleagues (2023) note after analyzing publication trends in their review: Half of the evaluations of ECB the authors studied took place in areas related to education and public health, suggesting that the evaluation of ECB is more prolific in these two areas. The finding from Bourgeois et al. contradicts my study's findings that evaluation of ECB in health programming are emergent in nature and not connected to one another, even though these publications spread over a period of 8 years.

Evaluations of ECB in complex, adaptive environments have a 6-year publication spread and represent the second most influential topic, but these documents are not as connected. This finding supports Bourgeois et al.'s assertion that the rising trend in evaluations of ECB beyond health and education holds much promise but requires more comprehension of what works.

The publication trends discussed in this study highlight past areas of investigation into the evaluation of ECB to understand future areas of exploration. Publication trends also summarize past conversations to provide a foundation that supports the evaluation field's understanding of areas where future exploration can propel the field forward.

Understanding relationships between multifaceted ECB approaches and their outcomes

The published literature on evaluation of ECB initiatives repeatedly shows that adopting only one ECB strategy is insufficient. Instead, multifaceted approaches are endorsed (Bourgeois et al., 2023; Norton et al., 2016). Adopting multifaceted approaches comprise strategies that include content to develop new knowledge, attitudes, and skills, as well as mechanisms that apply knowledge, attitudes, and skills to work routines and to build staff confidence. Previous reviews of the evaluation literature have stressed multifaceted approaches that combine experiential learning with a practical component (Norton et al., 2016) reported greater proportional increases in evaluation skills for indirect approaches that require engagement in evaluation processes, compared to direct approaches, such as workshops that change cognition or behavior (Bourgeois et al., 2023).

ECB design and implementation matters and should be considered in evaluative efforts (Preskill & Boyle, 2008). Bourgeois et al. (2023) have pressed for hypothesis testing to surface the nature of ECB relationships resulting from the adoption of multifaceted

approaches. My review underscores the pivotal role of multifaceted approaches and highlights gaps in knowledge about various approaches' optimal dosage, weight, and duration. My findings further suggest a need for research on and evaluation of ECB initiatives that employ multifaceted approaches; such research could shed light on these approaches' directionality, effects, and size.

Interpersonal competencies and relations impact ECB and its evaluation

Building and evaluating evaluation capacity comprises a complex web of activities and relationships (Weiss, 1998). Technical competencies alone are insufficient for providing and evaluating ECB. Social, interpersonal, or cultural competencies too are crucial to the practice of ECB (Stevahn et al., 2020). Employers of evaluators have expressed concerns about the interpersonal skills—such as "communication, negotiation, conflict, collaboration, and cross-cultural skills" (Stevahn et al., 2005, p. 52)—that are perceived to be required to engage in this type of work (Dewey et al., 2008). This hints at an opportunity to explore how interpersonal skills influence outcomes of ECB initiatives and which skills are most impactful.

As evaluation practice progressively involves evaluation partners (Skolits et al., 2009), evaluators' interpersonal skills and relationships with partners increasingly impact both participants' learning and the skills transfer to the workplace (Bourgeois et al., 2023). For example, modeling evaluation skills and championing the need for evaluations are crucial to the success of both training and participation in evaluations (Bakken et al., 2014; Preskill & Boyle, 2008). Likewise, providing technical assistance or mentorship or participating in evaluations requires some form of buy-in, trust, involvement, and cooperation from evaluation partners involved in ECB and evaluation processes. Previous reviews have stressed the importance of evaluators' facilitation, leadership, collaboration, and communication skills, as well as the quality and frequency of evaluation partners' participation (Labin, 2014; Preskill & Boyle, 2008). Despite their critical role in the success of both ECB and evaluation efforts, knowledge and consistent measurement of the mediating roles of interpersonal skills and relationships remain scant and signify a much-needed area for future exploration.

The moderating role of contextual factors in ECB readiness, effectiveness, mainstreaming, and sustainability

Evaluation capacity varies across organizations, interventions, staff positions, and levels of engagement during the course of evaluations (Lindeman et al., 2018). Individual changes in evaluation capacity affect organizational changes, and organizational environments are necessary for individual outcomes to occur and be sustained (Labin, 2014). Supportive organizational cultures, structures, and processes are prerequisites for learning from evaluation (Janzen et al., 2017; Taut, 2007), with previous reviews recognizing the influence of resources, leadership and management, culture and values, systems and work processes, policies and procedures, technology, and communication (Bourgeois et al., 2023; Norton et al., 2016; Preskill & Boyle, 2008). Adding to the list, Norton and colleagues include prioritizing evaluation; staff support in the form of expertise, incentives, and time for evaluation; and timeous use of evaluation findings.

Contextual factors shape ECB (Bourgeois et al., 2023; Suarez-Balcazar & Taylor-Ritzler, 2014) and therefore play a significant role in determining its readiness, effectiveness, mainstreaming, and sustainability. For example, readiness requires an alignment to organizational and leadership needs (Rorrer, 2016); effectiveness necessitates leadership

and staff buy-in (Lettenmaier et al., 2014; Ploeg et al., 2008); mainstreaming calls for resources, staff buy-in and involvement, and reorganizing tasks to prioritize the application of evaluation skills (Akintobi et al., 2012; Ploeg et al., 2008). And adequate resources, leadership commitment, supportive evaluation policies, and management of staff turnover are vital to sustaining ECB (Ploeg et al., 2008; Stevenson et al., 2002).

In addition to influencing ECB, various contextual factors come into play depending on whether evaluations focus on ECB readiness, effectiveness, mainstreaming, or sustainability. Organizations should be better equipped at understanding and influencing the conditions that support building and evaluating evaluation capacity (Preskill & Boyle, 2008), which will allow them more flexibility to tailor the selection of ECB strategies to their needs (Bourgeois et al., 2023), available resources (Norton et al., 2016) and unique contexts. Yet, knowledge of supportive conditions and their moderating influences remains almost nonexistent. Better information on the size, direction, and effects of these moderating relationships will illuminate the right conditions necessary for building, mainstreaming, and sustaining ECB in addition to highlighting the situations in which such efforts are infeasible.

Limitations to evaluation designs impact validity and generalizability of findings

The types of designs employed to evaluate ECB have implications for both internal validity and external generalizability. The designs adopted by the studies included in this review generally did not control the threats to internal validity, making them susceptible to bias due to low response rates and participant self-selection. Attributing cause and effect between ECB activities and their outcomes is further complicated by reliance on designs such as case studies (Janzen et al., 2017; Ray et al., 2012; Stevenson et al., 2002) and pre–post designs without comparison groups (Arnold, 2006; Bakken et al., 2014; Suarez-Balcazar & Taylor-Ritzler, 2014), which may limit the extent to which causal or attributional claims can be made (Norton et al., 2016). Without ruling out alternative explanations that may account for the observed ECB effects, questions remain about whether these effects can truly be attributed to the ECB. As with internal validity, the designs used also impact the generalizability of the ECB initiatives' outcomes beyond the contexts in which evaluations are conducted. Small sample sizes of the studies included in this review also hampered external generalizability (Bourgeois et al., 2023; Suarez-Balcazar & Taylor-Ritzler, 2014), making it more challenging to identify statistically significant relationships between ECB and its outcomes and generalize results to broader contexts.

STUDY LIMITATIONS

While this study has many strengths, particularly in regard to its internal validity, some limitations may compromise its external generalizability. I conducted the content analysis and qualitative review of articles as a single author. Since I did not use a second coder or reliability checks, the final review of the sample of articles I eventually included may reflect my bias in some way. Although I attempted to review the evaluation of ECB systematically, there was some subjectivity in the selection of studies that met the inclusion criteria. Additionally, this review intentionally did not define ECB nor evaluation, in an attempt to be inclusive and embrace emerging conversations. Both terms are characterized by a proliferation of definitions, the adoption of which would have excluded some studies even though they evaluated ECB.

Finally, evaluating ECB is a practical endeavor. Practitioner peer-reviewed publications of evaluation work are limited (Lobo et al., 2014), and due to constraints in time and resources, many evaluations of ECB do not make it into the peer-reviewed literature. When reflections on ECB do make it to the literature, they take the shape of online resources, technical reports, and conference presentations found in the gray literature (Bourgeois et al., 2023). The findings of this review may therefore also be limited by publication bias and the exclusion of gray literature.

CONCLUSION

Patterns in the published literature on the evaluation of ECB help the evaluation field understand where the field is and where it hopes to go. Here, I have examined and discussed such patterns in the area of the evaluation of ECB. I have highlighted the opportunities for further learning and areas evaluation methods can be strengthened. ECB is an important tool in growing evaluation buy-in and use, which ultimately increases the influence and impact evaluation has on the initiatives it evaluates. Most people conducting ECB are evaluators, and such practitioners sometimes forget to consider ECB activities as interventions themselves, rather than just extensions of evaluation work.

Context plays a significant role in shaping ECB outcomes. Interpersonal competencies affect evaluation processes and increasingly impact both learning and the transfer of skills to the workplace. Understanding how contextual factors and interpersonal skills influence the outcomes of ECB initiatives presents opportunities for further exploration. ECB design and implementation matter and should be considered in evaluating ECB initiatives, because these choices have implications for both internal validity and external generalizability. Stronger evaluation designs lead to a better understanding of what makes for effective ECB. This review and all the other articles in this issue of *New Directions for Evaluation* hope to build on the existing foundation and inspire others for future research and evaluation.

ORCID

Mervyan J. Konjore https://orcid.org/0009-0002-4985-7737

REFERENCES

Akintobi, T. H., Yancey, E. M., Daniels, P., Mayberry, R. M., Jacobs, D., & Berry, J. (2012). Using evaluability assessment and evaluation capacity-building to strengthen community-based prevention initiatives. *Journal of Health Care for the Poor and Underserved*, *23*(20), 33–48. https://doi.org/10.1353/hpu.2012.0077

Arnold, M. E. (2006). Developing evaluation capacity in extension 4-h field faculty: A framework for success. *American Journal of Evaluation*, *27*(2), 257–269. https://doi.org/10.1177/1098214006287989

Bakken, L. L., Núñez, J., & Couture, C. (2014). A course model for building evaluation capacity through a university–community partnership. *American Journal of Evaluation*, *35*(4), 579–593. https://doi.org/10.1177/1098214014523671

Bourgeois, I., Lemire, S. T., Fierro, L. A., Castleman, A. M., & Cho, M. (2023). Laying a solid foundation for the next generation of evaluation capacity building: Findings from an integrative review. *American Journal of Evaluation*, *44*(1), 29–49. https://doi.org/10.1177/10982140221106991

Bronk, K. C., Reichard, R. J., & Qi Li, J. (2023). A co-citation analysis of purpose: Trends and (potential) troubles in the foundation of purpose scholarship. *The Journal of Positive Psychology*, *18*(6), 1012–1026. https://doi.org/10.1080/17439760.2023.2168563

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, *62*(7), 1382–1402. https://doi.org/10.1002/asi.21525

Dewey, J. D., Montrosse, B. E., Schröter, D. C., Sullins, C. D., & Mattox, J. R. (2008). Evaluator competencies: What's taught versus what's sought. *American Journal of Evaluation*, *29*(3), 268–287. https://doi.org/10.1177/1098214008321152

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Janzen, R., Ochocka, J., Turner, L., Cook, T., Franklin, M., & Deichert, D. (2017). Building a community-based culture of evaluation. *Evaluation and Program Planning*, *65*, 163–170. https://doi.org/10.1016/j.evalprogplan.2017.08.014

Kaye-Tzadok, A., & Spiro, S. E. (2016). Evaluation capacity building: Can a classroom-based course make a difference? *Research on Social Work Practice*, *26*(5), 565–571. https://doi.org/10.1177/1049731514559521

Khaikleng, P., Wongwanich, S., Sriklaub, K., Ajpru, H., & Smuntavekin, S. (2015). A training module for evaluation capacity building of a health support organisation in Thailand. *Procedia—Social and Behavioral Sciences*, *171*, 1395–1399. https://doi.org/10.1016/j.sbspro.2015.01.259

Khare, A., & Jain, R. (2022). Mapping the conceptual and intellectual structure of the consumer vulnerability field: A bibliometric analysis. *Journal of Business Research*, *150*, 567–584. https://doi.org/10.1016/j.jbusres.2022.06.039

Labin, S. N. (2014). Developing common measures in evaluation capacity building: An iterative science and practice process. *American Journal of Evaluation*, *35*(1), 107–115. https://doi.org/10.1177/1098214013499965

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

LaFond, A. K., Brown, L., & Macintyre, K. (2002). Mapping capacity in the health sector: A conceptual framework. *The International Journal of Health Planning and Management*, *17*(1), 3–22. https://doi.org/10.1002/hpm.649

Lettenmaier, C., Kraft, J. M., Raisanen, K., & Serlemitsos, E. (2014). HIV communication capacity strengthening: A critical review. *Journal of Acquired Immune Deficiency Syndromes*, *66 Suppl*(3), S300–S305. https://doi.org/10.1097/QAI.0000000000000238

Lindeman, P. T., Bettin, E., Beach, L. B., Adames, C. N., Johnson, A. K., Kern, D., Stonehouse, P., Greene, G. J., & Phillips, G. (2018). Evaluation capacity building - Results and reflections across two years of a multi-site empowerment evaluation in an HIV prevention context. *Evaluation and Program Planning*, *71*, 83–88. https://doi.org/10.1016/j.evalprogplan.2018.09.001

Lobo, R., Petrich, M., & Burns, S. K. (2014). Supporting health promotion practitioners to undertake evaluation for program development. *BMC Public Health*, *14*(1), 1315. https://doi.org/10.1186/1471-2458-14-1315

Ma, T.-J., Lee, G.-G., Liu, J. S., Lan, R., & Weng, J.-H. (2022). Bibliographic coupling: A main path analysis from 1963 to 2020. *Information Research*, *27*(1), 918. https://doi.org/10.47989/irpaper918.

Milstein, B., Chapel, T. J., Wetterhall, S. F., & Cotton, D. A. (2002). Building capacity for program evaluation at the Centers for Disease Control and Prevention. *New Directions for Evaluation*, *2002*(93), 27–46. https://doi.org/10.1002/ev.40

Nájera-Sánchez, J.-J., Ortiz-de-Urbina-Criado, M., & Mora-Valentín, E.-M. (2020). Mapping value co-creation literature in the technology and innovation management field: A bibliographic coupling analysis. *Frontiers in Psychology*, *11*, 588648. Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2020.588648

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

Perianes-Rodriguez, A., Waltman, L., & van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, *10*(4), 1178–1195. https://doi.org/10.1016/j.joi.2016.10.006

Ploeg, J., De Witt, L., Hutchison, B., Hayward, L., & Grayson, K. (2008). Evaluation of a research mentorship program in community care. *Evaluation and Program Planning*, *31*(1), 22–33. https://doi.org/10.1016/j.evalprogplan.2007.10.002

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

Ray, M. L., Wilson, M. M., Wandersman, A., Meyers, D. C., & Katz, J. (2012). Using a training-of-trainers approach and proactive technical assistance to bring evidence based programs to scale: An operationalization of the interactive systems framework's support system. *American Journal of Community Psychology*, *50*(3–4), 415–427. https://doi.org/10.1007/s10464-012-9526-6

Rorrer, A. S. (2016). An evaluation capacity building toolkit for principal investigators of undergraduate research experiences: A demonstration of transforming theory into practice. *Evaluation and Program Planning*, *55*, 103–111. https://doi.org/10.1016/j.evalprogplan.2015.12.006

Skolits, G. J., Morrow, J. A., & Burr, E. M. (2009). Reconceptualizing evaluator roles. *American Journal of Evaluation*, *30*(3), 275–295. https://doi.org/10.1177/1098214009338872

Stevahn, L., Berger, D. E., Tucker, S. A., & Rodell, A. (2020). Using the 2018 AEA evaluator competencies for effective program evaluation practice. *New Directions for Evaluation*, *168*, 75–97. https://doi.org/10.1002/ev.20434

Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005). Establishing essential competencies for program evaluators. *American Journal of Evaluation*, *26*(1), 43–59. https://doi.org/10.1177/1098214004273180

Stevenson, J. F., Florin, P., Mills, D. S., & Andrade, M. (2002). Building evaluation capacity in human service organizations: A case study. *Evaluation and Program Planning*, *25*(3), 233–243. https://doi.org/10.1016/S0149-7189(02)00018-6

Stockdill, S., Baizerman, M., & Compton, D. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, *93*, 7–26. https://doi.org/10.1002/ev.39

Struyk, R. J., Damon, M., & Haddaway, S. R. (2011). Evaluating capacity building for policy research organizations. *American Journal of Evaluation*, *32*(1), 50–69. https://doi.org/10.1177/1098214010378355

Suarez-Balcazar, Y., & Taylor-Ritzler, T. (2014). 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

Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 285–320). Springer International Publishing. https://doi.org/10.1007/978-3-319-10377-8_13

Vogel, B., Reichard, R. J., Batistič, S., & Černe, M. (2021). A bibliometric review of the leadership development field: How we got here, where we are, and where we are headed. *The Leadership Quarterly*, *32*(5), 101381. https://doi.org/10.1016/j.leaqua.2020.101381

Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies*. Prentice Hall.

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, *18*(3), 429–472. https://doi.org/10.1177/1094428114562629

How to cite this article: Konjore, M. J. (2024). Two decades of evaluating evaluation capacity building: A bibliographic coupling review. *New Directions for Evaluation*, *2024*, 11–28. https://doi.org/10.1002/ev.20616

AUTHOR BIOGRAPHY

Mervyan J. Konjore is a Rhodes Scholar and doctoral candidate at Claremont Graduate University's Evaluation & Applied Research program. Her research interests are evaluation methods, evaluation capacity, and evaluation capacity building.