

Toward a systemic approach to evaluating evaluation capacity development

Bob Williams¹ | Elah Matt² | Sarah Klier³  |
Scott G. Chaplowe⁴ | Stefanie I. Korswagen⁵ 

¹Evaluation Consultant, Wellington, New Zealand

²Evaluation Consultant, Norwich, UK

³DEval, German Institute for Development Evaluation, Bonn, Germany

⁴Evaluation Consultant, Madrid, Spain

⁵Heat International, Königstein im Taunus, Germany

Correspondence

Bob Williams, Evaluation Consultant,
Wellington, New Zealand.
Email: bob@bobwilliams.co.nz

Abstract

In this article, we contribute to evaluation capacity development (ECD) research and practice by demonstrating how a systems-based evaluation approach can be applied to a complex ECD initiative. Since 2013, the German evaluation institute, DEval, has implemented an ECD program focused on Latin America and the Caribbean (LAC). This is a diverse, emergent program developed using a bottom-up approach. DEval takes a facilitating rather than managerial role, supporting ECD among a range of stakeholders. The article discusses several issues related to evaluating this complex ECD program. The evaluation viewed the program as a system comprising two sub-systems: one focused on improving the knowledge and skills of evaluators and stakeholders (capacity), and the other focused on fostering an enabling evaluation environment (capability). It drew on an evaluation tradition primarily focused on the consequences of evaluations. A focus on the consequences of evaluations is essentially one step beyond evaluation use (an instrumental orientation) toward usefulness (a value-based orientation). The evaluation adopted a systemic approach to define the core boundaries of evaluation focus, acknowledging that various stakeholders perceive the evaluation differently based on their interests or stake. By aligning the evaluation with these perspectives, it could focus on what matters most to different stakeholder groups. Four key perspectives or framings were identified: the *program* itself, the *transferability* of the LAC lessons to ECD in other parts of the world, *collaborations* with global agencies undertaking ECD at a global level, and *institutional* practices within DEval.

INTRODUCTION

To set the stage for the discussion in this article, it is important to first establish our use of evaluation capacity development (ECD) rather than evaluation capacity building (ECB). Over the past few decades, both ECB and ECD have gained traction among evaluation scholars and practitioners alike. While these terms are often used interchangeably, they can also have distinct connotations. ECB typically refers to enhancing the abilities of individuals and organizations to conduct evaluations through teaching and learning strategies (Preskill & Boyle, 2008). ECD expands on this to include building sustainable systems and fostering an evaluation culture, typically in the international development context (e.g., DEval, 2024a; IEG, 2024). Given the international development focus of this article and the characteristics of the evaluand, we use the term ECD. Further, this terminology aligns with our distinction of capability development encompassed by ECD, versus a focus on capacity development that typifies ECB, which we elaborate later in this article.

The focus of ECD practice and research has shifted over time. ECD practice has evolved from developing individual capacities to enhancing organizational and governmental capabilities to support evaluation. On the research side, there have been efforts to develop frameworks for conceptualizing ECD (e.g., Bourgeois et al., 2023; Preskill & Boyle, 2008). However, papers reporting on applied evaluations of ECD initiatives remain sparse. Due to the complex nature of these initiatives, it can prove difficult to determine which evaluation approach will best serve the purpose of evaluating an ECD initiative in a given context.

This article presents a systems-based approach for evaluating an ECD initiative. We developed this approach to evaluate an ECD initiative implemented by the German Institute for Development Evaluation (DEval) in Latin America and the Caribbean (LAC) (Klier et al., 2022). In doing so, we aim to contribute to ECD research and practice by demonstrating how a systems-based evaluation approach can be applied to a complex ECD initiative.

We continue this article by describing the complexities of ECD systems, and then introduce DEval's ECD program and our approach to ECD evaluation. Our discussion then lays out our reflections on the value of our approach and how it addresses some ECD challenges.

Due to the complex nature of these initiatives, it can prove difficult to determine which evaluation approach will best serve the purpose of evaluating an ECD initiative in a given context.

THE COMPLEXITIES OF ECD SYSTEMS

Over the past few decades, the supply of and demand for ECD has increased among evaluation scholars, practitioners, and users alike. On the supply side, there has been a response to the industrialization of evaluation as a commercial field and the economic interest of national and international evaluation agencies and universities to provide evaluation training as an income source. On the demand side, there is increasing pressure, especially in the international development field, for evaluations to be designed, driven, and conducted by people who really understand the local or national context and apply evaluation methods and approaches that are appropriate to local values, traditions, and circumstances. Calls for the decolonization of evaluation are getting louder, and many agencies are wrestling with the implications of their practice. While the "helicopter" approach of hiring an evaluation team external to the country or region has been challenged for many years, the calls

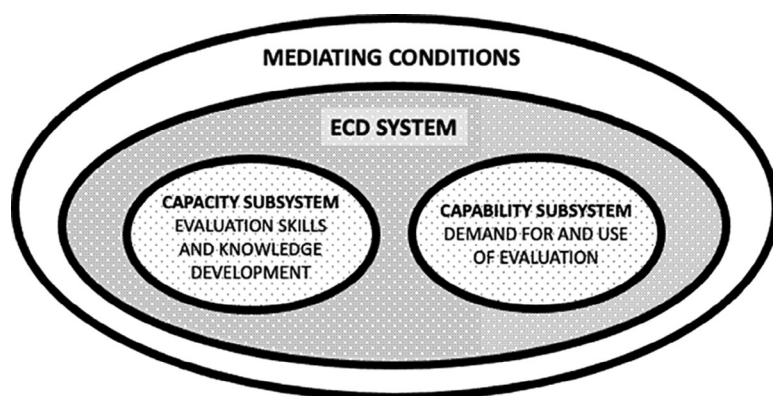


FIGURE 1 Structure of ECD system and sub-systems. ECD, evaluation capacity development.

are getting louder and more influential. The implication of these trends is that there is both increased demand for and supply of locally sourced evaluation skills.

In between supply and demand are the evaluation facilitators. These are funders, donors, foundations, development agencies, government departments, and nongovernmental organizations (NGOs) that commission and use evaluations. The level of understanding, ownership, and support for the practice of evaluation and its capacity development is not a given but will vary according to institutional context. In some instances, institutional support is forthcoming, but it is not uncommon for evaluation to be consigned to a “descriptive, tick-box, accounting exercise that steers clear of judgment rather than providing judgment that steers decision making” (Chaplowe & Hejnowicz, 2021).

It is useful, therefore, to consider the evaluation system we have described as comprising two sub-systems, as illustrated in Figure 1.

The capacity subsystem contains all the activities, products, values, and results to develop the necessary technical, organizational, political, and social skills and knowledge to *design and undertake* evaluations. The capability subsystem contains all the activities, products, values, and results to develop the necessary technical, organizational, political, and social space that allows the *application and use* of these skills and knowledge. Outside the ECD system is an environment that mediates (i.e., helps and hinders) the evaluation system in doing its job.

Developing and coordinating both evaluation capacity *and* capability is a massively complex task. This raises some important challenges about *evaluating* capacity/capability development. As observed by John Morgan:

“We need to approach the evaluation of capacity with a sense of modesty, curiosity, and patience. It is not an easy thing to get right for a whole series of reasons. The balance between product and process is not easy to strike. Intangibles, ghosts, and hidden agendas in the countries proliferate. The time and timing issues of all the various participants are hard to match up. And untangling the behavior of complex systems change issues takes patience and contextual knowledge, both of which are in short supply everywhere. I still believe that effective evaluations of capacity can be produced.” (Morgan, 2013, p. 76)

Before diving into our systems-based evaluation approach, we outline the complexity of the ECD system the evaluation needed to manage.

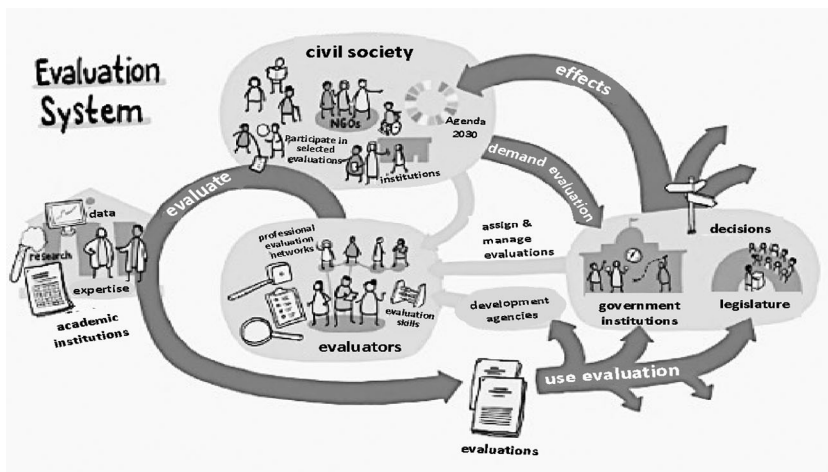


FIGURE 2 The evaluation system.

DEval's ECD PROGRAM

Funded by the German Federal Ministry of Economic Cooperation and Development (which we'll refer to from here on as "the funder"), DEval conducts strategic evaluations of German development cooperation. The funder's flagship ECD program in LAC began in 2012 and was transferred to DEval in 2013. The program is regional in scope, with Costa Rica serving as the focal point for knowledge creation, transfer, and dissemination. The Costa Rican Ministry of Planning (Mideplan, 2024), is the key implementation and knowledge transfer partner.

While this project has gone through several phases, the purpose of this work has remained fairly constant. The intent is to develop evaluation capacity and capability that enable "individuals, organizations and society as a whole to commission and implement evaluations and put them to systematic use" (DEval, 2024a).

To this end, DEval has adapted a systemic approach focused on developing both evaluation capacity and capability throughout the evaluation system (Figure 2), rather than just prescriptive quality assurance of evaluation activities (Klier et al., 2022).

The program promotes the evaluation capacities and capabilities of public, private, and civil society stakeholders at individual and institutional levels. Further, it aims to nurture an enabling environment for evaluations, with country, regional, and international scope. The program promotes partnerships, knowledge sharing, and cooperation. Its objectives and activities include (DEval, 2024b):

- Developing evaluation capacities in public institutions. For example, supporting country-led evaluations of Sustainable Development Goals in Costa Rica and Ecuador.
- Supporting inclusive evaluation processes. For example, DEval and the National University of San Juan in Argentina established the EvalParticipativa platform (<https://evalparticipativa.net>). The 4000-member-strong web-based platform serves as a community of practice and learning for participatory evaluation in LAC.
- Promoting capacities of evaluation experts. This includes training, developing, and disseminating knowledge products and networking for evaluators, including young and emerging evaluators.
- Enabling knowledge exchange among evaluation stakeholders. Activities here include the multi-actor national evaluation platforms in Costa Rica and Ecuador, as well as

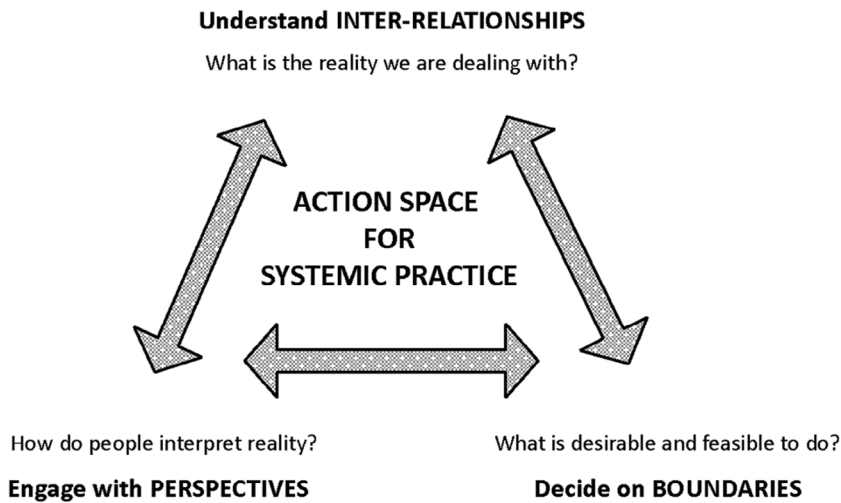


FIGURE 3 A systemic approach to understanding complexity.
Williams, B. (2019).

the National Evaluation Capacities Index (<https://inceval.org/initiative>), which measures national evaluation capacities and practices. These initiatives have fostered the development of domestic and regional communities of practice.

- As described above, the development of national evaluation capacities is of great interest to the international community. Accordingly, DEval cooperates with various bilateral and multilateral organizations including the Global Evaluation Initiative (GEI, 2024) and the World Food Program (WFP).

OUR APPROACH TO EVALUATING A COMPLEX ECD SYSTEM

While the ECD program seeks to enable “individuals, organizations and society as a whole to commission and implement evaluations and put them to systematic use” (DEval, 2024b), its activities are not predetermined or prescribed. Rather, the activities emerged from the ECD system as it evolved. In other words, the ECD program was managed as a complex system.

The evaluation design used a framework developed by one of the authors (Williams, 2019), illustrated in Figure 3. Essentially, this framework focuses attention on three aspects of an evaluation: the need to understand interrelationships, engage with multiple perspectives, and deliberate carefully on important boundary choices.

In terms of understanding the program’s interrelationships, the evaluation’s scope included a wide range of activities at different spatial scales over a 10-year span of action. In addition, the interrelationships between the evaluation *capacity* subsystem and the evaluation *capability* subsystem were intricate and highly dynamic, each influencing the other over time and place.

Furthermore, the program affected and interested a wide range of stakeholders, each bringing a different set of perspectives on the value and worth of these interrelationships and their consequences. Each of these stakeholders held specific perspectives of the program: those of individual evaluation professionals; collective perspectives of organizations and institutions involved in the region; perspectives regarding ECD within DEval; and the perspectives of those who work for a range of regional and international agencies that

have an ECD role. These represented a wide range of views on the value and worth of the program.

In practical terms, this complexity meant that while the evaluation needed to be aware of the large range of real-world activities and the numerous perspectives on what these activities mean, for the evaluation to be feasible not all activities and perspectives could be included.

Indeed, all evaluation systems are partial, because boundaries must be drawn to make the evaluation task feasible and useful. Consequently, we had to turn to the third corner of the triangle and draw some boundaries. We had to exclude parts of the ECD system so that we could evaluate it.

Setting system boundaries

There is nothing unusual about setting boundaries around evaluations. Evaluators do this all the time under a variety of names: evaluation purposes, evaluation questions, evaluation criteria, evaluation methodologies, and statistical techniques. All of these set boundaries around an evaluation. Each of these boundaries means that some things will be included in an evaluation and some things will not. Most attempts to draw boundaries around evaluations are value-based. They are based on beliefs about what evaluation is, what is considered important about the evaluand, who or what should (and should not) benefit from the evaluation, and how we believe phenomena can be observed and measured. Thus, boundary setting needs careful deliberation, as it involves ethical and cultural as well as technical decisions. In practical terms, boundary choices constitute the value-driven task of deciding which perspectives and interrelationships to include and which to exclude.

Given these demands, identifying a desirable and *viable* evaluation system required us to reconsider the original evaluation brief given to us by DEval. In collaboration with DEval, we identified and reflected on these critical boundaries using two approaches: one from the systems field and one from the evaluation field.

From the systems field drew on the discipline of soft systems (Checkland & Scholes, 1999; Williams & Hummelbrunner, 2010). This discipline was developed in the 1960s to challenge the then-dominant idea that systems are essentially self-evident phenomena that can be recognized by and behave according to specific mechanical processes (e.g., stocks, flows, levers). In other words, the dominant focus was on the interrelationships corner of the triangle. In contrast, soft systems emphasize the perspective corner of the triangle. Soft systems views systems as human constructs and systems' behaviors as products of human perspectives, behaviors, and values.

Thus, two people may look at the same set of interrelationships and identify them as two different systems. For instance, a program that provides venues where IV drug users exchange dirty needles for clean ones will be viewed by some people as a disease reduction system (reducing blood-borne diseases), by others as a safety system (a place to score drugs without being harassed), and by others as a reputation-harming system (reducing the status of the community and thus its ability to attract services and people to the area). An evaluation of the significance and worth of that program would be different when viewed through each of these framings. These different framings of the observed interrelationships will in turn depend on the values, life experiences, histories, interests, and motivations of different stakeholders involved in or affected by the project. Again, in any complex program, there are many potential framings, and selecting a limited number of them can reduce the focus and scale of an evaluation to more manageable levels.

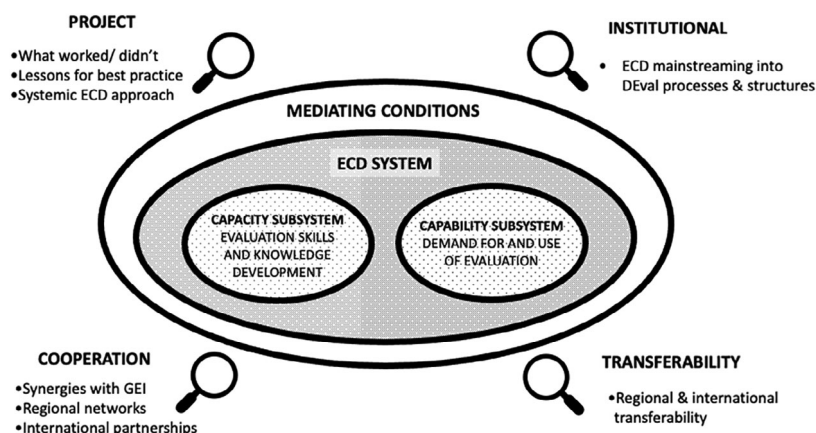


FIGURE 4 Evaluation lenses.

To identify and determine which framings of the ECD program should be the focus of the evaluation, we inserted a scoping stage into the evaluation design. Through a set of interviews and document analysis, this phase focused on exploring the motivations of the various actors in the ECD system and what they might desire to do as a consequence of the evaluation. We explored and analyzed the views and perspectives of key stakeholders within DEval, as the main client and user of the evaluation, as well as those of the funder, some key Latin American stakeholders, and selected stakeholders at the international level.

We used an approach from the evaluation field to help us select and deliberate on a smaller number of framings. This approach was developed in the early parts of this century, primarily by Karen Kirkhart (Kirkhart, 2000) and Mel Mark (Mark, 2024). It is an evaluation orientation specifically focused on the desired *consequences* or outcomes of the evaluation—its *usefulness*. Using this approach, we could narrow the range of framings to those that were likely to lead to the most valuable consequences.

As suggested earlier, the scoping process raised issues that went beyond what happened on the ground in LAC. It included issues relating to collaborations within LAC; between LAC and Germany; within DEval itself; between DEval and its funding agency; and between DEval, its funding agency, and other agencies that also have a stake in ECD inside and outside Germany. It also related to future plans of DEval outside LAC and the implications of demands from countries receiving German development funding to conduct evaluations that serve local interests in addition to German ones.

FRAMING THE EVALUATION

As a consequence of the scoping, four framings were identified and selected as the primary lenses through which the program was evaluated (Figure 4):

- A project framing
- A transferability framing
- A cooperation framing
- An institutional (DEval) framing

Based on these evaluation framings, we devised a set of semi-structured, open-ended interview questions to guide our data collection (Chaplowe & Cousins, 2016). The framings also informed the selection of ECD initiatives and key informants for the evaluation.

The project framing

Under this framing, we explored lessons from individual projects and initiatives within the ECD program that can be used to improve the ECD program itself. The focus here was to understand what worked well within the individual initiatives, what did not work well, and what could be improved. We also asked key informants which lessons they felt would be useful for other ECD initiatives and stakeholders both regionally and internationally. The intended consequence is relatively conventional; that future projects within the ECD program will be informed by previous experiences.

The transferability framing

DEval conducts evaluations, collects data, and builds partnerships in many countries. These include collaborations in sub-Saharan Africa and Southeast Asia, both of which are German development priority regions. There are therefore discussions about the desirability and feasibility of transferring DEval's ECD approaches and practices to these regions. However, it is well known that projects that are successful in one context may not be as successful when transferred to other contexts.

Consequently, there was much interest in what ideas and practices would and would not contribute to viable and appropriate ECD practices in those regions. This required us to investigate the evaluation context within those regions and assess what might or might not transfer well across cultures and what might help and hinder that. For instance, sub-Saharan Africa and Southeast Asia have more recent experiences of colonization compared with Latin America. The consequence of this evaluation could therefore be that Germany's future ECD activities would need to be aware of and sensitive to local needs and aspirations.

Institutional framing

As mentioned earlier, the primary purpose of DEval is to conduct evaluations of German development cooperation. Mostly, these evaluations are methodologically based on academic social research traditions. DEval evaluations rely on in-house expertise and only collaborate with country-based partners on an ad hoc basis, primarily for data collection. In contrast, DEval's ECD program has developed long-standing, trusting relationships with country-based and international partners, relying on these to support their partners in building evaluation capacity and capability. More recently, the demand from within Germany and the countries receiving German development assistance is shifting toward evaluations that are responsive to specific country needs as well as German agency needs, including greater involvement of agencies and local evaluators in those countries. This raises issues of capacity and capability within those countries to design, undertake, and use evaluations.

Therefore, DEval as a whole, as well as the German Ministry for Economic Cooperation and Development, are interested in how this partnership-oriented approach to ECD can be more firmly anchored, mainstreamed, and effectively managed within DEval. Thus a desired consequence of our evaluation was DEval being able to respond effectively to the demand for more collaborative and developmental approaches to evaluation.

Cooperation framing

ECD is increasingly a global activity, with many agencies undertaking capacity and capability development. Countries and civil society organizations on the receiving end of development cooperation in general, as well as ECD, have complained about the competition, and confusion created by this diversity in ECD approaches and initiatives. There are now regional and international initiatives, such as the Global Evaluation Initiative (GEI), seeking to bring more coherence to the practice of ECD. The German government, through the Cooperation and Development Ministry is a major funder of GEI. Therefore, there is a close relationship between the DEval ECD program and GEI. One of the desired consequences of the evaluation was therefore to show how the ECD work of DEval and GEI can be even better integrated.

CONCLUDING COMMENTS

In this article, we examine what methodological lessons can be learned from this evaluation of a highly complex ECD program.

The 10-year-old program's scope included Latin America and comprised a wide variety of interventions at local and continental scale. The program was bottom-up rather than top-down, in the sense that it was facilitated by DEval in response to local demands and contexts.

Five primary methodological challenges stood out in the evaluation of this program:

1. The evaluand comprised numerous interrelated projects and initiatives covering diverse topics in different contextual environments.
2. The program was multinational and multisectoral in nature.
3. Spanning a 10-year timescale, it required covering a substantial history and adapting to evolving stakeholders.
4. The program employed a "bottom-up" approach, characterized by emergent, developmental, and innovative management.
5. Various stakeholders—including project participants, program managers, supporting staff, funders, and a broader network of organizations providing ECD services—had vested interests in the program and its evaluation.

Furthermore, it was an evaluation focused on assessing a program for evaluation capacity and capability development, adding the responsibilities and challenges of evaluating our own professional practice.

The evaluation design had to resolve three further challenges that are not uncommon in the evaluation of complex interventions:

1. How to simplify the evaluation without being simplistic.
2. How to focus the evaluation while remaining aware of the evaluand's scope and complexity.
3. How to meet the evaluation needs of multiple stakeholders.

These challenges were addressed by incorporating three key design elements drawn from the systems and evaluation fields:

1. The evaluation viewed the program as a system comprised of two sub-systems: one focused on improving the knowledge and skills of evaluators and stakeholders

(capacity), and the other focused on fostering an enabling evaluation environment (capability).

2. We drew on an evaluation tradition that is primarily focused on the consequences of evaluations. A focus on the consequences of evaluations is essentially one step beyond evaluation use (an instrumental orientation) toward usefulness (a value-based orientation).
3. The evaluation adopted a systemic approach to define the core boundaries of evaluation focus, acknowledging that various stakeholders perceive the evaluation differently based on their interests or stake in the evaluation. By aligning the evaluation with these perspectives, it could better prioritize what matters most to different stakeholder groups. Four key perspectives or lenses were identified: projects, collaboration, transferability, and institutional change.

There is no straight-line connection between the above concepts, challenges, and dilemmas, yet these three design elements collectively provided a practical way to navigate them during this exercise.

This sounds good in theory, but what practical insights or advantages has this approach provided that could be useful for other evaluations of ECD or similar complex capacity development interventions? The answer, as often seen in evaluation, is a combination of yes, no, and maybe.

Capacity and capability

The distinction made between capacity (i.e., skill development) and capability (i.e., opportunity to use those skills) is not new. The notion that context matters has been a mantra in evaluation for decades. However, it is still common for capacity development processes and their evaluations to focus on “training” and whether that training was delivered “right,” rather than assessing whether training or broader capacity development was the right thing to do in the first place. The distinction between capacity and capability also helps evaluators to focus on *whose* capacity has been developed and how. For instance, it can highlight the importance of developing the capacity of those who can enhance the capability of organizations to support and benefit from evaluations, thereby creating spaces that are able to support evaluation politically, culturally, and administratively.

Evaluation consequences

It is a curious feature of the evaluation field that while evaluations often focus on intervention outcomes, they rarely focus on the outcomes of the evaluation itself. By identifying what stakeholders wanted the evaluation to achieve (i.e., its consequences, its usefulness), we were able to design the evaluation backward from those potential consequences. We did not develop a theory of change for the evaluation. We felt that this evaluation, like the intervention itself, occurred in too complex a setting for a theory of change to be a worthwhile tool (Williams, 2024). But at the least we were able to ask stakeholders what they would like the evaluation to achieve rather than what data they would find interesting.

Framing and boundary decisions

All evaluations make boundary decisions, often implicitly. Decisions on what data to collect, how much is necessary, and the analysis processes are often seen as technical rather

than value-based choices. It is not uncommon for evaluations to focus on the objectives set by the commissioner, leaving critical boundary decisions to this single stakeholder. In contrast, the systemic approach employed in this evaluation involved exploring different framings or perspectives of what the program is about, providing a more inclusive understanding of different stakeholder viewpoints. This enhanced the ways in which the evaluation's results can be useful for a range of desirable consequences. This approach also made explicit those framings or perspectives that were excluded, why, and who or what may be affected by those choices.

What makes the evaluation of ECD important?

As stated earlier, many of these conclusions apply to many, if not most, evaluations. What makes them particularly relevant to evaluating ECD is that we are evaluating our own profession and practices. That requires us to set an example when deciding what aspects of the evaluation are most likely to have an impact. It also suggests that we consciously acknowledge that many of the boundary choices we make are more than technical. It is important to acknowledge that there will be people and ideas that benefit from our evaluations as well as people and interests that are marginalized by them. Hopefully the approaches we have described in this article reflect these obligations and will help inform them in future ECD evaluation.

ORCID

Sarah Klier  <https://orcid.org/0000-0002-0656-2544>

Stefanie I. Korswagen  <https://orcid.org/0000-0003-2306-9329>

REFERENCES

- 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>
- Chaplowe, S. G., & Cousins, J. B. (2016). *Monitoring and evaluation training: A systematic approach*. SAGE Publications, Inc. <https://doi.org/10.4135/9781071878712>
- Chaplowe, S., & Hejnowicz, A. (2021). Evaluating outside the box: Evaluation's transformational potential. *Social Innovations Journal*, 5, 1–19.
- Checkland, P., & Scholes, J. (1999). *Soft systems methodology in action*. John Wiley & Sons.
- German Institute for Development Evaluation (DEval). (2024a). Evaluation capacity development. German Institute for Development Evaluation. <https://www.deval.org/en/evaluation-capacities/evaluation-capacity-development>
- German Institute for Development Evaluation (DEval). (2024b). *Focelac+*. German Institute for Development Evaluation. <https://www.deval.org/en/evaluation-capacities/current-ecd-projects/focelac>
- Global Evaluation Initiative. (2024). *Home | Global Evaluation Initiative*. Global Evaluation Initiative. <https://www.globalevaluationinitiative.org/>
- Kirkhart, K. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 88, 5–24.
- Klier, S. D., Nawrotzki, R. J., Salas-Rodríguez, N., Harten, S., Keating, C. B., & Katina, P. F. (2022). Grounding evaluation capacity development in systems theory. *Evaluation*, 28(2), 231–251. <https://doi.org/10.1177/13563890221088871>
- Mark, M. M. (2024). Evaluation use and influence. *Research Handbook on Program Evaluation*, 14–34. <https://doi.org/10.4337/9781803928289.00008>
- Ministerio de Planificación Nacional y Política Económica (Mideplan). (2024). *Inicio | Ministerio de Planificación Nacional y Política Económica*. Ministerio de Planificación Nacional y Política Económica. <https://www.mideplan.go.cr/>
- Morgan, P. (2013). Evaluating capacity development. In S. I. Donaldson, T. Azzam, & R. F. Conner (Eds.), *Emerging practices in international development evaluation* (pp. 75–104). Information Age Publishing.
- 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>

- Williams, B., & Hummelbrunner, R. (2010). *Systems concepts in practice: A practitioner's toolkit*. Stanford Business Books.
- Williams, B. (2019). *Systemic evaluation design: A workbook*. <https://gum.co/evaldesign>
- Williams, B. (2024). Why do we have theories of change of the programme intervention but not of the intervention that is the evaluation? In A. Koleros, M. Adrien, & T. Tyrrell (Eds.), *Theories of change in reality: Strengths, limitations, and future directions* (pp. 190–196). Routledge. <https://doi.org/10.4324/9781032669618/>
- World Bank Group: Independent Evaluation Group (IEG). (2024). Evaluation capacity development. World Bank Group: Independent Evaluation Group (IEG). <https://ieg.worldbankgroup.org/topic/evaluation-capacity-development>

How to cite this article: Williams, B., Matt, E., Klier, S., Chaplowe, S., & Korswagen, S. I. (2024). Toward a systemic approach to evaluating evaluation capacity development. *New Directions for Evaluation*, 2024, 125–136. <https://doi.org/10.1002/ev.20612>

AUTHOR BIOGRAPHIES

Bob Williams specializes in the application of ideas from the systems field to evaluation and organizational development.

Elah Matt, develops, implements, and evaluates international development projects.

Sarah Klier heads the Evaluation Capacity Development work of the German Institute for Development Evaluation.

Scott G. Chaplowe is an evaluation and strategy specialist with expertise in complex systems analysis for evaluation capacity development.

Stefanie I. Korswagen works on international cooperation for development, with a focus on climate change.