Grounding evaluation capacity development in systems theory

Sarah D. Klier (iD)
German Institute for Development Evaluation (DEval), Germany

Raphael J. Nawrotzki
German Institute for Development Evaluation (DEval), Germany

Nataly Salas-Rodríguez
German Institute for Development Evaluation (DEval), Germany

Sven Harten
German Institute for Development Evaluation (DEval), Germany

Charles B. Keating
Old Dominion University, USA

Polinpapilinho F. Katina (iD)
University of South Carolina Upstate, USA

Abstract

While "systemic thinking" is popular in the context of capacity development and evaluation, there is currently a lack of understanding about the benefits to employing systems theory in evaluation capacity development. Systems theory provides a useful orientation to the work involved in complex systems (e.g. national evaluation systems). This article illustrates how evaluation capacity development practitioners can use systems theory as a conceptual tool to gain a better understanding of the functional aspects and interrelationships present within a given evaluation system. Specifically, the systems theory perspective can help elucidate the reasons for the success or failure of a given evaluation capacity development program or activity. With the goal of motivating evaluation capacity development practitioners to use systems theory in their

Corresponding author:
Sarah D. Klier, German Institute for Development Evaluation (DEval), Fritz-Schäffer-Straße 26, Bonn 53113, Germany.
Email: sarah.klier@deval.org

work, this article presents a systems theory framework for evaluation capacity development and offers practical examples of how it can be adopted.

Keywords
axioms and propositions, evaluation capacity development, national evaluation systems, systemic approach, systems theory

Introduction

Governments and other organizations are increasingly expressing interest in enhancing the effectiveness of their development interventions (Davies and Pickering, 2015; Global Partnership for Effective Development Co-operation, 2020; Prizzon, 2016). Evaluations are important tools for strengthening transparency, accountability, and learning, all of which are important aspects of good governance. Alarmingly, the Bertelsmann Transformation Index (Hartmann, 2020) now shows that many governments are engaging in a growing number of authoritarian actions and there has been a concomitant weakening of democratic structures throughout the world. These issues have further been exacerbated by the ongoing COVID-19 pandemic. Such conditions emphasize the need for evaluations, which are highly relevant tools that can be used to hold these entities accountable for their actions and spending practices, thereby instilling trust that they are wisely using resources during the current global crisis. However, not all countries and organizations possess the necessary skills or knowledge to perform high-quality evaluations, which makes it critically important to develop evaluation capacities where they are lacking.

Evaluation capacity development is the formal process of establishing both the human capital (skills and knowledge) and organizational structures needed to perform high-quality evaluations (Boyle and Lemaire, 1999; Preskill and Boyle, 2008). International debates on the importance of ownership and effectiveness in the context of development cooperation (e.g. Paris Declaration 2005 and Accra Agenda for Action 2008) have increasingly focused on the importance of using national systems to accomplish monitoring and evaluation. This is also evident when looking at the United Nations 2030 Agenda for Sustainable Development (hereafter, 2030 Agenda). This stresses the importance of systematically monitoring and evaluating the implementation of 17 sustainable development goals (SDGs) and emphasizes the role of country-led evaluations, which entails the need to strengthen national evaluation capacities. The OECD Development Assistance Committee (DAC) EvalNet Task Team for evaluation capacity development envisions that evaluation will be recognized as a pillar of democratic governance and increasingly used to measure the progress toward achieving the SDGs.

Given that programs aimed at developing national evaluation capacities must often operate within highly complex socioeconomic and political environments, systems theory offers a useful orientation to account for the multidimensionality of relevant organizations and institutions while clarifying the dynamic nature of their interactions and connections, both internally and externally (Midgley, 2006; Ortiz Aragón and Macedo, 2010). More specifically, systems theory focuses on processes, patterns, perspectives, boundaries, and interrelationships (Williams and Van't Hof, 2016), which are important components of evaluation capacity development practice.

As a practical example, this article draws on evaluation capacity development work conducted at the German Institute for Development Evaluation (DEval), which is mandated by the German Federal Ministry of Economic Cooperation and Development (BMZ) to evaluate BMZ-funded development cooperation. DEval has been given three core tasks: (1) conduct independent and strategically relevant evaluations of German development cooperation, (2) enhance evaluation methods and standards, and (3) foster evaluation capacity development in selected partner countries. Evaluation capacity development has been undertaken at DEval for nearly 8 years. External evaluations and data from DEval's monitoring and evaluation system show that its evaluation capacity development projects are highly successful. This success can partially be attributed to the implicit use of systems theory, which inform most evaluation capacity development applications at DEval.[1] Systems theory has successfully and more formally been used in the fields of capacity development (e.g. Fisher, 2010; Morgan, 2005) and evaluation (e.g. Hummelbrunner et al., 2015; Reynolds et al., 2016; Williams and Imam, 2006). Moreover, a variety of studies have acknowledged its potential usefulness in the evaluation capacity development context (Krapp and Geuder-Jilg, 2018; Ortiz Aragón and Macedo, 2010; Stockmann, 2018). However, the literature does not currently contain any descriptions of how to systematically embed evaluation capacity development work in systems theory. This is a critical omission, as such information could provide an orientation for evaluation capacity development practitioners to ground their work in theoretical considerations, with many added benefits. This article targets this gap by providing the first theoretical account of how evaluation capacity development activities and program components can be anchored in systems theory, illustrated by practical examples from work at DEval.

Systems theory and national evaluation systems

Systems theory emerged during the 1940s providing an alternative to *reductionism* (Laszlo and Krippner, 1998; Parsons, 1951; Von Bertalanffy, 1949), which holds that a complex organism (system) is the sum of its parts; therefore, understanding can be deduced from its components (Hammond, 2002). By contrast, systems theory is centered on *holism*, which suggests that a system is more than the sum of its parts. Systems therefore produce behaviors that are not directly attributable to individual elements, but which instead emerge through interactions between those elements (Laszlo and Clark, 1972; Von Bertalanffy, 1972). Systems of all types share certain common characteristics, thus permitting a generic description; that is, one which is not inherent to any particular discipline, domain, or sector (Hammond, 2002). Systems can be understood as self-organizing functional units that differentiate themselves from their environments through a unique composition of interactions and relationships (Von Bertalanffy, 1967). Systems theory does not comprehensively represent the complete system in the form of a massive supermodel, but instead derives constructs or abstractions that facilitate an understanding of observed phenomena (Bernard, 2006; Midgley, 2006). In this regard, systems theory permits multiple perspectives, thus acknowledging that different constructs or abstractions may offer a variety of similarly valid viewpoints on the same system (Midgley, 2006).

Systems theory acknowledges the vast complexity of the world. It is focused on interrelations, and addresses emergent processes, patterns, and relationships. Importantly, it acknowledges the importance of these concepts without trying to "make them fit" by accepting the significance of uniqueness and constant development. Here, systems theory does not try to

simplify complexity or change, but embraces both. In fact, it is based on the demonstrable existence of system effects in the natural world (Morgan, 2005).

Despite the above, there is currently no consensus on what constitutes systems theory. Different versions have been popularized in various academic disciplines, including sociology (Luhmann, 1995), psychology (Von Bertalanffy, 1967), biology (Pattee, 1978), and management sciences (Kast and Rosenzweig, 1972). Yet, in recent years, systems theory has been employed to explain a wide range of phenomena, including childhood obesity (Allender et al., 2019), cybersecurity (Tarafdar and Bose, 2019), and sustainable development (Eustachio et al., 2019). While there is heterogeneity in the applications and foci, a concise set of concepts and principles has become associated with systems theory. Although a categorically accepted definition of systems theory has not yet emerged, the literature provides substantial evidence on its isomorphic concepts, laws, principles, and theorems, which are applicable to different systems (Adams et al., 2014; Clemson, 1984; Katina, 2016; Mobus and Kalton, 2015; Whitney et al., 2015). In this article, we refer to the formal definition offered by Adams et al. (2014) and define systems theory as "a unified group of specific propositions which are brought together to aid in understanding systems, thereby invoking improved explanatory power and interpretation with major implications for systems practitioners" (p. 113). This definition is drawn from 6 major sectors and 42 individual fields of science, organizing systems theory into a set of 7 axioms and 30 propositions. These elements are inclusive of the laws, principles, and theorems that collectively define systems theory. Here, axioms provide an abstract definition of various important parts of a system, while propositions describe tangible rules for how a system operates (Adams et al., 2014). Table 1 shows an adjusted and modified version of the system construct developed by Adams et al. (2014), which was supplemented and revised by Whitney et al. (2015), including the seven axioms and their propositions. In Table 1, asterisks indicate characteristics that were found to be particularly relevant when applying the systems construct in the context of evaluation capacity development work conducted at DEval.

Researchers in the evaluation capacity development field frequently draw from theories on adult learning or organizational change (Bourgeois and Cousins, 2013; Ortiz Aragón and Macedo, 2010; Preskill and Boyle, 2008), while systems theory usage remains very limited. Nevertheless, a previous review of the published evaluation capacity development literature shows that a variety of researchers have acknowledged the potential usefulness of a systemic perspective (Bourgeois, 2016; Danseco, 2013; Haarich and del Castillo Hermosa, 2004; Preskill and Boyle, 2008; Stockmann, 2018). Indeed, systems theory is noted as a valuable framework for dealing with the complexities of evaluation capacity development (Danseco, 2013), and has even been used as an orientation tool for better understanding evaluation capacity development efforts, both within general evaluation systems (Haarich and del Castillo Hermosa, 2004) and educational networks (Grack Nelson et al., 2019). Previous studies have invoked systems language, and referenced important systems concepts such as levels, interactions, feedback loops (Haarich and del Castillo Hermosa, 2004), and control (Grack Nelson et al., 2019). However, the literature lacks a rigorous and comprehensive application of systems theory to evaluation capacity development. This is unfortunate, given the many benefits associated with grounding evaluation capacity development in systems theory as follows:

- *Understanding complexity*. Systems theory provides an orientation for understanding a highly complex situation encountered by evaluation capacity development programs

Table 1. Adjusted and modified representation of the system construct developed by Adams et al. (2014), as revised by Whitney et al. (2015)*.

Axiom	Proposition
1. Centrality	Communication (Shannon, 1948a, 1948b, Skyttner, 2005)
	Control (Checkland, 1993)
	Emergence (Aristotle, 2002; Checkland, 1993)*
	Hierarchy (Checkland, 1993; Pattee, 1973)*
2. Contextual	Complementarity (Bohr, 1928)*
	Incompressibility (Cilliers, 2002; Richardson, 2004)
	Holism (Smuts, 1927)*
	Boundary (Skyttner, 2005; Von Bertalanffy, 1968)*
3. Goal	Equifinality (Von Bertalanffy, 1950)
	Multifinality (Buckley, 1967)
	Purposive Behavior (Rosenblueth et al., 1943)*
	Satisficing (Simon, 1955, 1956)*
4. Operational	Dynamic Equilibrium (Miller, 1978; Von Bertalanffy, 1968)
	Homeorhesis (Waddington, 1957, 1968)
	Homeostasis (Cannon, 1929)
	Redundancy (Pahl et al., 2011)
	Relaxation Time (Clemson, 1984; Holling, 1996)*
	Self-organization (Ashby, 1947)*
	Sub-optimization (Hitch, 1953)
5. Viability	Circular causality (Korzybski, 1933)*
	Feedback (Wiener, 1948)
	Recursion (Beer, 1979)
	Requisite Hierarchy (Aulin-Ahmavaara, 1979)
	Requisite Variety (Ashby, 1956)*
6. Design	Minimal critical specification (Cherns, 1976, 1987)*
	Power Law (Newman, 2005)
	Requisite Parsimony (Miller, 1956; Simon, 1974)
	Requisite Saliency (Boulding, 1966)
7. Information	Redundancy of Potential Command (McCulloch, 1965)*
	Information Redundancy (Shannon and Weaver, 1949)

Note: Key publications for each proposition are shown in parentheses.
*Republished with permission of Inderscience Enterprises Limited (UK), from Whitney et al. (2015) Systems theory as a foundation for governance of complex systems, in *Int. J. System of Systems Engineering*.

with many interconnections, stakeholders, and interests (Imam et al., 2006). Systems theory does not attempt to reduce the complexities or dynamics of systems, but instead helps to address them;

- *Multilevel and multidisciplinary perspectives.* Systems theory provides a perspective from which to link local and global sectors and disciplines (Imam et al., 2006). This perspective is beneficial, since evaluation capacities exist at various levels (Heider, 2011), including the individual (Tseng, 2011), programmatic (Martin and Carey, 2018), and societal (Boyle et al., 1999);

- *Expect the unexpected*. Systems theory provides a perspective from which it is possible to account for and explain properties that emerge unexpectedly (Imam et al., 2006). It

thereby permits evaluation capacity development practitioners to transition from a reactive to a proactive mode of working;

- *Transferability and learning.* Systems theory allows for a generalization of findings between programs (even if unrelated), since problems will be addressed via abstract relationships (propositions) proposed by the theory. It then allows for a transition from individual cases to generic recommendations for systemic evaluation capacity development programs.

Given these multiple benefits, the following section discusses how the tenets of systems theory apply to practical evaluation capacity development work. At DEval, two large-scale evaluation capacity development projects were jointly implemented with the Costa Rican Ministry of National Planning and Economic Policy (Mideplan). DEval's expertise lies in supporting national evaluation systems (NESs). While most of the existing literature on evaluation systems is situated at the organizational or institutionalized levels (Andersen, 2020; Briceño, 2012; Leeuw and Furubo, 2008; Martinaitis et al., 2019; Raimondo, 2018), the systemic perspective adopted in this article considers all actors, processes, and structures as national evaluation system components that allow for the planning, implementation, and use of evaluations to assess national public policies. As we explain in later sections, it is important to note that the composition of a national evaluation system gradually evolves and changes (Pérez Yarahuán and Maldonado Trujillo, 2015). The examples described in this article refer mostly to the Costa Rican national evaluation system illustrated in Figure 1.[2]

The Costa Rican national evaluation system consists of diverse actors (Serrano, 2020). According to Cunill-Grau and Ospina (2008) as cited by PY and MT (2015), the national evaluation system is under the authority of the executive branch; in this case, Mideplan, through its evaluation unit. Mideplan is at in fact the head of the national evaluation system, and therefore plans, manages, and fosters the employment of evaluation in addition to elaborating and introducing structures to institutionalize evaluation. The ministries provide the primary demand for evaluations, formalized in the National Evaluation Agenda (NEA; Serrano, 2020). External evaluators (some of which are organized within the Costa Rican evaluation network) are contracted to perform evaluations, which are managed by Mideplan. An advisory board comprised of the most relevant stakeholders is tasked with guiding the evaluation process. Academic institutions also help perform evaluations, and even offer relevant postgraduate programs. Thus, evaluators can develop the necessary capacities to implement high-quality evaluations.

Civil society has recently begun to demand and use evaluations, and now takes part in selected participatory evaluations managed by Mideplan. On the contrary, development agencies play minor roles, as they do not demand evaluations in the Costa Rican national evaluation system context, but do support the development of evaluation capacities. In turn, the legislature (especially the technical services office) demands and uses the results of evaluations to adjust policies and laws. The national ombudsman represents civil society in the national evaluation platform (NEP), and may request and use evaluations on public interventions. Finally, the Ministry of Finance provides a budget for public evaluation and oversight of the financial resources available for conducting evaluations. These actors form the Costa Rican national evaluation system and participate jointly in semestral meetings hosted by Mideplan.
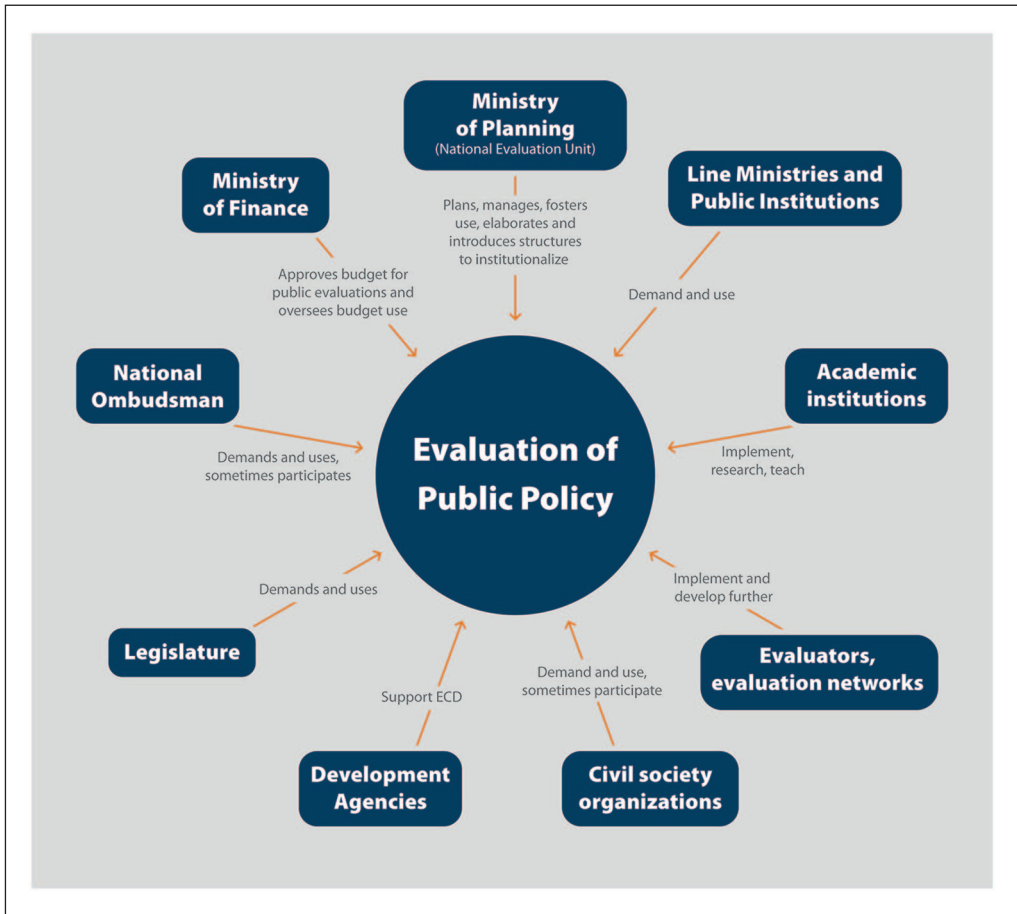
Figure 1. Depiction of the Costa Rican national evaluation system.

Embedding DEval's evaluation capacity development work in systems theory

This section describes the utility of the system construct that Adams et al. (2014) developed, as supplemented by Whitney et al. (2015) for evaluation capacity development practice. We chose to focus on a subset of the propositions that were most relevant to DEval's evaluation capacity development practice (see Figure 2) since a comprehensive examination of all propositions is beyond the scope of this article. Relevance was determined based on expert judgment of DEval evaluation capacity development practitioners in light of the propositions' usefulness and impact on evaluation capacity development projects outcomes. Next, we explain and discuss the selected propositions, and describe how they can be applied in practical evaluation capacity development work, using examples from DEval's evaluation capacity development practice. As an overview, we provide a short description of the full set of axioms and their propositions in the Supplemental Table S1. Figure 2 provides a visual representation of the systems construct applied to national evaluation systems.

Figure 2. Depiction of the systems construct applied to national evaluation systems.

Centrality axiom

The centrality axiom states that all systems are driven by two pairs of propositions that are central to their existence: first *emergence* and *hierarchy*; second as *communication* and *control* (Adams et al., 2014; Whitney et al., 2015). We find *emergence* and *hierarchy* to be relevant propositions for evaluation capacity development work.

Emergence. The concept of emergence can be traced to famous philosophers and poets, including Aristotle and Goethe (Blitz, 1992). Specifically, the concept implies that a system can only be distinguished as an entity given the existence of certain properties that are not found by merely looking at its individual parts (Ortiz Aragón and Macedo, 2010; Wan, 2011). Rather, these properties emerge from interactions and interrelationships between components (Morgan, 2005), which thus make the whole system greater than the sum of its parts (Checkland, 1999).[3] In other words, emergence describes the continuous process of the unpredictable and spontaneous appearance of behavior, structure, or performance. Systems must be capable of

accommodating wide variabilities in emergence (robustness). They must also effectively adapt and adjust to emergent conditions (resilience), even though these cannot be known beforehand (Keating, 2009).

Accordingly, evaluation capacity can be seen as an emergent property of an evaluation system (Land et al., 2009) as well as the ultimate goal of evaluation capacity development projects. As an example from DEval's evaluation capacity development work, the unpredictable emergence of evaluation capacity in the Costa Rica national evaluation system is evident through the formation of an EvalYouth (EY) country chapter (see EY global initiative) in 2017. EY is an organization comprised of young and emerging evaluators (YEEs), all of whom are professionally trained, but have fewer than 5 years of evaluation experience (Brenes-Alfaro and Montero Corrales, 2020; EvalPartners, 2016). Members of the Costa Rican EY chapter founded a new Voluntary Organization for Professional Evaluation (VOPE), titled the Evaluation and Monitoring Network of Costa Rican (RedEvalCR, by the Spanish acronym), which quickly gained a positive reputation and influence. Due to the importance of recognizing the emergence of new evaluation capacity within the Costa Rica national evaluation system, both the EY chapter and VOPE were incorporated into evaluation capacity development projects. Specifically, institutions were encouraged and financially supported to include EY members in their evaluations, which had the dual benefit of providing organizations with highly motivated young staff and allowing young evaluators to gain valuable practical work experience. Thus, the emerging evaluation capacity could be channeled and strengthened within the Costa Rican national evaluation system. Moreover, due to the quick and strong emergence of the EY chapter, Mideplan included YEEs in developing its national evaluation policy.

Hierarchy. Systems can usually be disaggregated into subsystems, thus showing a hierarchical (Wilby, 1994) or layered (Checkland, 1999) structure.[4] Between the different layers or levels of the system, "surfaces and filters" (Wilby, 1994: 662) may influence information flow. In an evaluation system, individuals perform evaluations (individual level), while institutions strategically plan and finance evaluations (organizational level), and the environment influences the evaluation through culture, politics, and resources (enabling environment or contextual level) (Heider, 2011). Alternatively, evaluation systems can be divided into layers of demand, intermediation, and supply, which reflect the usage, management, and implementation of the evaluation (Feinstein, 2011). However, the different layers and levels are not independent; rather, they are interconnected. For example, removing skilled labor (sometimes referred to as a brain drain) at the individual level decreases capacity at the organizational level (Fisher, 2010). In the context of evaluation capacity development work, it is essential to strategically analyze all layers and levels of the national evaluation system in addition to determining where the work is most relevant and/or how to generate the best results. It is particularly useful to focus on the flow of information between layers and identify obstacles that may obstruct those flows. For example, after developing a set of guidelines on how to manage evaluations and publishing the National Evaluation Policy, Mideplan realized that specialized training should be offered to national institutions with an associated role in the NEA. With support from DEval, training programs were developed to build capacities at various organizational levels, including those pertaining to office staff, managers, and institutional leaders (individual level). In conjunction with organizational guidelines, the practice of building capacity at the individual level within different parts of the organizational hierarchy ensured support

throughout the organization (organizational level). Finally, the National Evaluation Policy and Agenda assured the sustainable use of evaluation practices (enabling environment or contextual level).

Contextual axiom

The contextual axiom draws attention to the importance of both the environment and the circumstances of a given system that enable or constrain its functions (Adams et al., 2014). The propositions of *complementarity, holism*, and *boundary* were found to be the most relevant in DEval's evaluation capacity development work.

Complementarity. This proposition originally stems from quantum mechanics, and suggests that two methodologically different observations may produce different results, but at the same time may belong together and complement one other (Bohr, 1928). Each new perspective on a system adds important information. However, it can be challenging to accept the existence of different and potentially conflicting "truths." Within evaluation systems there are different realities. For example, when facing an evaluation task, academic actors frequently see only the methodological approach, while government actors see deadlines and policy implications. However, both views are important for the national evaluation system. When considering a problem from various perspectives, product quality also dramatically increases. Consistent with the notion of *complementarity*, DEval works with evaluation platforms (multi-stakeholder groups) that are comprised of different representatives of the national evaluation system, including those from academia, civil society, and both the public and private sectors. These evaluation platforms create a space for connecting, exchanging thoughts, and joint planning, thus leveraging the full strength of *complementarity*. The evaluation capacity development project in Costa Rica created a successful NEP[5] that continues to be in operation 3 years after the project's completion and plays a key role within the country's evaluation system.

Holism. A system is not merely the sum of its parts, but must be treated as a whole (Smuts, 1927). For a system to exist, the *holism* proposition highlights the fact that the interactions and interrelationships are equally important as the individual parts (Morgan, 2005). A system is often unique, in which case its results may not be generalizable. As such, recommendations for best practices or caveats may only be applicable at a high aggregation level (Morgan, 2005). DEval applied the proposition of *holism* in its evaluation capacity development practice. For example, a manual to address participation in evaluations within public institutions was developed in collaboration with Mideplan. Recognizing the uniqueness of interrelationships of certain evaluation systems, DEval drew on practical evaluation examples to adjust parts of the manual to contextually reflect local conditions in Costa Rica and Mexico.

Boundary. According to Von Bertalanffy (1968), as an entity, any system has boundaries, whether they are spatial or dynamic. However, the perimeter that outlines the system is often indistinct making it difficult to identify what belongs to a complex system (Von Bertalanffy, 1972). In national evaluation systems determining boundaries and the resulting clarifications of what is positioned inside and outside is a politically sensitive step (Williams and Van't Hof, 2016). In Costa Rica, practicalities were the main reason why actors from the private sector

only played marginal roles in the national evaluation system. Unlike civil society actors, the private sector was not the focus of either the evaluated policy programs or the evaluations themselves. In addition, the contacted actors showed little interest in the topic of evaluation.[6] Consequently, the private sector has little awareness of the state's evaluation efforts, which is why the evaluation function of accountability to the private sector is less pronounced. The boundaries of the national evaluation system can (and do) shift and are accordingly adjusted on a regular basis.

Goal axiom

The goal axiom states that systems reach certain goals through purposeful behaviors via pathways and means (Adams et al., 2014). In DEval's evaluation capacity development work, the most relevant propositions of this axiom were *purposive behavior* and *satisficing*.

Purposive behavior. Purposive behavior refers to an act or behavior that is intentionally directed toward goal attainment (Rosenblueth et al., 1943). To direct its actions, the system must therefore understand the final intended state (Buckley, 1967). An important first step in DEval's evaluation capacity development projects was to define the goals (desired final state) of the Costa Rican national evaluation system. The actors from the NEP wanted to improve the transparency of public policy via the participation of civil society in evaluation. Thus, actions were set to integrate the ombudsman and a representative of nongovernmental organizations (NGOs) working toward the SDGs into the NEP. Furthermore, a handbook on participation in evaluations within public institutions was developed for the public sector. In this context, goal-oriented clarity resulted in actions that ultimately fostered public policy transparency in Costa Rica.

Satisficing. From the perspective of ethics and rational choice theory, Honderich (2005) describes "satisficing" as choices and actions that achieve sufficient satisfaction, but which are not optimal given particular situational constraints. There are multiple ways to achieve this goal. Due to limited resources, the system cannot discover the "optimal" path, and must "settle" on a path that provides a reasonable degree of satisfaction (Simon, 1956). The notion of *satisficing* became particularly important when selecting evaluation objects within the Costa Rican NEA. While the primary goal of the Costa Rican national evaluation system is to perform and use relevant evaluations of public policies, the number of those that are eligible for evaluation is considerably larger than the number that the available resources allow. It is impossible to know in advance which evaluations will produce the largest impacts or changes (*optimal choice*). It is therefore necessary to select a subset of policies for evaluation, particularly including those that the national evaluation system considers most promising (*satisficing*). Ideally, this selection process is guided by experience, theory, and comprehensive expertise. Evaluation capacity development projects can support national evaluation systems in making more informed choices. In this regard, DEval supported Mideplan in designing a systematic process to select 15 public interventions for evaluation every 4 years within the framework of the NEA. Here, the selection process reflects a tradeoff between the available resources and current political priorities. The choice made every 4 years is likely not optimal, but is still satisfactory given the incomplete knowledge about the future relevance of various public interventions.

Operational axiom

The operational axiom describes the mechanics and internal functions of the system, which determine its unique behaviors and performance. *Relaxation time* and *self-organization* proved to be especially important for DEval's evaluation capacity development work.

Relaxation time. Systems do not usually remain in a state of equilibrium for long periods. Rather, they are frequently disturbed, and will attempt to return to a state of equilibrium over time. However, systems require adequate time to recover from disturbances in an orderly fashion (Clemson, 1984; Holling, 1996). Although disturbances are normal, systems that experience too many disturbances in short succession may become unstable. This makes it important to allow sufficient *relaxation time* between disturbances. Many forces can destabilize a national evaluation system, including a shift in political preferences, economic changes, health crises (e.g. the COVID-19 pandemic), exchanges of key decision-makers, changes in organizational structures, or even the results of a sensitive evaluation. A practical example is the national evaluation system in Ecuador. The Ecuadorian political system and its national evaluation system experienced several disturbances in a short period due to changes in various political institutions (e.g. staff reductions, changes in leadership, and administrative restructuring, including the state entity responsible for evaluation). Since an evaluation capacity development project can also act as a disturbance, DEval postponed some of its evaluation capacity development activities in Ecuador in order to permit sufficient *relaxation time* for the Ecuadorian national evaluation system to return to a state of equilibrium. Stability is a crucial consideration for all evaluation capacity development projects, in which case it is advisable to carefully observe a system to determine disturbances and permit sufficient relaxation time before providing an evaluation capacity development stimulus. National evaluation systems themselves are parts of an overall national political system and can therefore work as destabilizing factors.

Self-organization. To ensure survival, a functioning system can adapt and reorganize its structure to cope with change (Cilliers, 2002). Based on this proposition, we assume that a national evaluation system is inherently capable of dealing with changes and challenges. As such, we are able to recognize the agency of the national evaluation system. Rather than actively directing the national evaluation system, it should be empowered and supported in order to autonomously self-organize to the greatest degree possible. Rather than enforcing practices from powerful Western donor agencies, evaluation capacity development actors should work with national and local methods to foster the culturally sensitive development of national evaluation systems. This approach may reduce efficiency (outputs) in the short term, but is certainly more sustainable in the long run, especially given improvements in ownership and agency. One example of DEval's work is the development of the Costa Rican NEP. Here, NEP stakeholders invited other stakeholders to regular meetings, networked with them, and initiated joint activities to strengthen the evaluation capacity. The evaluation capacity development project was then able to support individual collaborations that formed independently (self-organized). For example, several NEP members (including the National Ombudsman and Mideplan, as well as evaluators) agreed to conduct a participatory evaluation. DEval supported this evaluation project with technical expertise as well as financial and personnel resources. As such, DEval's evaluation capacity development work merely supported the self-organization forces emerging from within the NEP rather than directing or leading the process.

Viability axiom

The viability axiom addresses key parameters within a system that must be controlled to ensure its continued existence. To date, *circular causality* and *requisite variety* are relevant to DEval's evaluation capacity development practices.

Circular causality. This proposition suggests that there is no simple cause–effect relationship. Rather, effects result from multiple causes that often interact, amplify, or cancel each other out. Causes themselves can be influenced by the effects they produce, which results in circular relationships and nonlinear feedback loops. Moreover, causes and effects can operate across levels; indeed, "micro effects can have macro causes and vice versa" (Morgan, 2005: 4). A system is a dynamic entity, such that complex cause–effect relationships may change over time (Morgan, 2005). To promote the use of evaluations (effect), DEval sensitized policymakers to the many benefits and applications of evaluation results (Cause 1). Meanwhile, evaluators were trained (Cause 2) and the development of evaluation standards was promoted (Cause 3) based on the assumption that better evaluation quality would increase use. DEval was also conscious of both nonlinearity and circular causality in its evaluation capacity development program-planning efforts. For DEval's evaluation capacity development projects, theories of change always consider internal and external factors as well as interactions and feedback loops. They are developed in close collaboration with partners and adapted when necessary.

Requisite variety. A system frequently experiences disturbances in its relationships and interactions. The *requisite variety* proposition states that the variety of controllers must be equal to the variety of the disturbances that need to be controlled (Ashby, 1956). In a national evaluation system, disturbances can be political (e.g. changes in priorities), financial (e.g. loss of funding), or organizational (e.g. changes in staff or consultants). To address these disturbances, the national evaluation system must be as diverse as possible. Initially, the Costa Rican government planned to undertake evaluations solely via Ministry staff. DEval recommended the employment of Ministry staff for oversight and management, but advised that any evaluations should be commissioned to qualified independent evaluators. DEval further recommended the establishment of an advisory board for each evaluation, with the most relevant stakeholders as members. Thus, national evaluation system diversity increased through external evaluation experts who contributed different viewpoints on problems while facilitating the resolution of disturbances due to their varying perspectives and specialized knowledge, skills, and connections.

Design axiom

The design axiom provides information on how a system is planned, instantiated, and developed (Adams et al., 2014; Whitney et al., 2015). *Minimal critical specification* was especially relevant in DEval's evaluation capacity development practice.

Minimal critical specification. The proposition of minimal critical specification assumes that a system works best when only the minimal constraints necessary to support system performance are specified for tasks, jobs, and roles. This situation requires that we first know what is truly essential (Cherns, 1976). While it may be important to describe the output clearly, it is rarely necessary to be precise about how it should be accomplished. Furthermore, rules and

regulations can inhibit adaptation and effective action (Cherns, 1976, 1987). The DEval evaluation capacity development projects defined their objectives in conjunction with the evaluation platform, and generally discussed the activities that can lead to the objectives. Here, time and money were the minimal critical specifications, but projects usually permitted substantial flexibility in how objectives were reached. A practical example of DEval's evaluation capacity development work is the generation of an index to measure national evaluation capacities (INCE, by the Spanish acronym). This idea originated at a regional evaluation platform meeting. A working group of 30 actors was formed, with minimal critical specifications including time, financial costs, and human resources. These collaborative efforts were ultimately successful; after 4 years of work, the index has been measured in nine Latin American countries. We attribute the accomplishment of fruitful cooperation between many diverse actors to the abovementioned *minimal critical specifications*, which evoked ownership among collaborators and permitted substantial flexibility and creativity in reaching the objective.

Information axiom

The information axiom states that systems possess, create, modify, and transfer information. We found *redundancy of potential command* to be especially relevant in DEval's evaluation capacity development work.

Redundancy of potential command. The proposition of *redundancy of potential command* suggests that control is dispersed throughout the system (Beer, 1972), such that the system parts have distributed autonomy and can therefore take autonomous decisions at any time (Bohórquez Arévalo and Espinosa, 2015). As a system grows in size and complexity, it becomes increasingly important to decentralize commands so that decisions can be as close as possible to the supporting actions. For example, if new branches, communities, projects, or groups are added to a system, then efficiency is improved given that these new administrative entities are allowed some form of autonomy. DEval applied this proposition when supporting the establishment of a working group (comprised of the evaluation platform members) to develop a national evaluation policy for Costa Rica. While members belonged to different parts of the Costa Rican national evaluation system, the working group retained its autonomy, which allowed for both efficient performance and the necessary flexibility to adapt to changes.

Conclusion

This article's novel contribution lies in its focus on systems theory. More specifically, using practical examples from DEval, we illustrate how evaluation capacity development practitioners can apply the system construct developed by Adams et al. (2014) and later enhanced by Whitney et al. (2015) to theoretically inform their evaluation capacity development activities. Systems informed evaluation capacity development practitioners provide system actors with technical expertise, but do not direct the system. This requires trustful cooperation. Systemic evaluation capacity development can only be applied if evaluation capacity development practitioners are able to adapt their own agenda to the (constantly evolving agenda of the) evaluation system.

Limitations

This article constitutes a novel contribution due to its focus on systems theory in the context of evaluation capacity development practices. While most examples describe DEval's work with the Costa Rican national evaluation system, first experiences on a regional scale have shown that a systemic approach can also be applied in other countries and other evaluation systems (e.g. Ecuador). This article only described propositions that had the greatest effect on DEval's evaluation capacity development practice and for which particularly illustrative practical examples were found. Future research should apply all propositions to evaluation capacity development practice examples. Thus far, the evaluation capacity development work conducted by DEval has been rooted in a systemic *understanding*, but the system construct was not systematically integrated during project planning. We therefore encourage evaluation capacity development practitioners to consider systems theory already in the planning and design stage to capitalize on the various benefits that a thorough theoretical grounding offers.

Yet, evaluation capacity development practitioners need to be aware of a number of challenges: first, while various systems theory propositions (e.g. *complementarity, requisite variety*) highlight the value of multi-actor (national) evaluation platforms (or other forms of networking), these are not easy to establish. Organizing and managing a (national) evaluation platform requires substantial administrative resources, political sensitivity, and comprehensive knowledge of the sociopolitical context when selecting relevant actors. Second, it may be difficult to sensitize partners to several key systems theory concepts. For example, a systemic evaluation capacity development entails a participatory approach to decision-making and high levels of transparency. As such, a systems theory approach to evaluation capacity development may be particularly challenging in societies with traditional hierarchical and/or autocratic structures. Third, the planning process for a systemic evaluation capacity development project should be flexible and ongoing. However, projects must be formally applied for and agreed upon, with activities and schedules defined in the application before implementation. This makes systemic (flexible) evaluation capacity development practices much more difficult; it is also time-consuming, as actors must build trustful relationships and understand the system, with short project phases often standing in the way. Fourth, systems theory is broad enough to incorporate various interactions and conditionalities, but lacks the detail of more specialized theories (e.g. adult learning theory) (Preskill and Boyle, 2008). For certain specialized aspects of evaluation capacity, it may thus be useful to complement systems theory with theories from disciplines such as political science, sociology, or educational science.

Benefits

While keeping the various challenges in mind, this work highlighted many benefits of applying systemic approaches to evaluation capacity development. First, systemic evaluation capacity development practitioners are part of the system with which they work. This entails both accepting that evaluation capacity development practitioners represent only one perspective among many, while also needing to rely on many other perspectives to enhance their work. Systemic evaluation capacity development requires that all actors have space for participation and expression of their expertise. As such, practitioners contribute through their own evaluation expertise while simultaneously benefiting and learning from the evaluation system as a whole. Second, a systemic evaluation capacity development approach accepts that ownership is held by different actors of the national evaluation system. This supports system operations

rather than focusing on goals and results, thus allowing for flexibility and self-determination. By applying a systemic evaluation capacity development approach, the complete evaluation system is strengthened, and not only a single actor. This contributes to a high degree of sustainability and good governance. Third, systems theory provides an analytical lens with which to identify the most relevant leverage effects of an evaluation system and to avoid major political disruptions. It also permits an improved understanding of the reasons for the success or failure of an evaluation capacity development program or component in addition to having implications for program design.

Yet, the employment of systems theory in evaluation capacity development is still nascent. This article offers an orientation for evaluation capacity development practitioners when theoretically grounding their work in systems theory. Governments and organizations are complex entities, and systems theory provides a useful tool to better understand and improve the operation of evaluation capacity development projects and programs therein.

Declaration of conflicting interests

Funding

ORCID iDs

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

Polinpapilinho F. Katina https://orcid.org/0000-0002-2213-4929

Supplemental material

Supplemental material for this article is available online.

Notes

1. While the various activities and program components of DEval's evaluation capacity development work were designed using a systemic understanding, a rigorous grounding of DEval's activities in established systems theory was only conducted retrospectively as part of the present study.
2. Evaluation capacity development efforts in Costa Rica build on a decades-long legacy that can be traced back to the 1980s, leading to the creation of a master's program in programmatic evaluation at the University of Costa Rica in 1995.
3. There is an ongoing philosophical debate on whether "emergence" refers only to novelties that are unexplained and unpredictable or novelties that can be explained or predicted (Wan, 2011).
4. The hierarchy proposition generally assumes a top-down, authoritarian system design (Wilby, 1994). However, an equal or random structure can be present in certain situations. For example, Capra (1997) assumes that all layers are equally important, and that there is no hierarchy within a system.
5. The following institutions form the Costa Rican Evaluation Platform: *Academic organizations* (Public Administration Research and Training Center of the University of Costa Rica; International

Center of Economic Policy for Sustainable Development of National University; Public Administration School of University of Costa Rica; Latin American Faculty of Social Sciences, Costa Rica's campus; Central American Institute of Public Administration; Master in Evaluation of University of Costa Rica); *NGOs* (Laboratory of Public Policy, Costa Rican Investment Promotion Agency); *development agencies* (Inter-American Institute for Cooperation on Agriculture, German Institute for Development Evaluation/Project Focelac, United Nations Children's Fund Costa Rica's office); *national and regional VOPEs* (Evaluation and Monitoring Network of Costa Rica; EvalYouth Costa Rica; Monitoring, evaluation and systematization network in Latin America and the Caribbean); *executive branch entities* (National Planning Ministry, Ministry of Finance; Social Development and Family Allowances Fund, Single System of State Beneficiaries, National Disabilities Office (both dependencies of the Ministry of Labor and Social Security); *legislative branch entities* (Documentary Development and Digital Information Area of the Parliament, National Audit Office, National Ombudsman as representation of non-organized civil society).

6. In Colombia, DEval's evaluation capacity development project has encountered very strong interest in evaluation from some (large) private sector actors. Again, this shows that each national evaluation system is unique (see *holism*), and that the boundaries can be drawn differently.

References

Adams KM, Hester PT, Bradley JM, et al. (2014) Systems theory as the foundation for understanding systems. *Systems Engineering* 17(1): 112–23.

Allender S, Brown AD, Bolton KA, et al. (2019) Translating systems thinking into practice for community action on childhood obesity. *Obesity Reviews* 20(S2): 179–84.

Andersen NA (2020) The constitutive effects of evaluation systems: Lessons from the policymaking process of Danish active labour market policies. *Evaluation* 26(3): 257–74.

Aristotle (2002) *Metaphysics, Book H-Form and Being at Work*, Translated by J. Sachs. Santa Fe, NM: Green Lion Press.

Ashby WR (1947) Principles of the self-organizing dynamic system. *Journal of General Psychology* 37(2): 125–28.

Ashby WR (1956) *An Introduction to Cybernetics*, 1st edn. New York: Springer.

Aulin-Ahmavaara AY (1979) The law of requisite hierarchy. *Kybernetes* 8(4): 259–66.

Beer S (1972) *Brain of the Firm*, 1st edn. New York: McGraw-Hill.

Beer S (1979) *The Heart of the Enterprise*. New York: Wiley.

Bernard HR (2006) *Research Methods in Anthropology: Qualitative and Quantitative Approaches*, 4th edn. Lanham, MD: AltaMira Press.

Blitz D (1992) *Emergent Evolution: Qualitative Novelty and the Level of Reality*. Boston, MA: Kluwer Academic Publishers.

Bohórquez Arévalo L and Espinosa A (2015) Theoretical approaches to managing complexity in organizations: A comparative analysis. *Estudios Gerenciales* 31: 20–9.

Bohr N (1928) The quantum postulate and the recent development of atomic theory. *Nature* 121: 580–90.

Boulding K (1966) *The Impact of the Social Science*. New Brunswick, NJ: Rutgers University Press.

Bourgeois I (2016) Performance measurement as precursor to organizational evaluation capacity building. *Evaluation Journal of Australasia* 16(1): 11–8.

Bourgeois I and Cousins JB (2013) Understanding dimensions of organizational evaluation capacity. *American Journal of Evaluation* 34(3): 299–319.

Boyle R and Lemaire D (1999) *Building Effective Evaluation Capacity: Lessons from Practice*. New Brunswick, NJ: Transaction Publishers.

Boyle R, Lemaire D and Rist RC (1999) Introduction: Building evaluation capacity. In: Boyle R and Lemaire D (eds) *Building Effective Evaluation Capacity: Lessons from Practice*. New Brunswick, NJ: Transaction, 1–19.

Brenes-Alfaro L and Montero Corrales CA (2020) Young and emerging evaluators team experience in a real evaluation process. *Zeitschrift für Evaluation* 19(1): 149–60.

Briceño B (2012) Defining the type of M&E system: Clients, intended uses, and utilization. In: Lopez-Acevedo G and Krause P (eds) *Building Better Policies. The Nuts & Bolts of M&E Systems*. Washington, DC: The World Bank, 33–46.

Buckley W (1967) *Sociology and Modern Systems Theory* (Prentice-Hall Sociology Series). Hoboken, NJ: Prentice-Hall.

Cannon WB (1929) Organization for physiological homeostasis. *Physiological Reviews* 9(3): 399–431.

Capra F (1997) *The Web of Life: A New Synthesis of Mind and Matter*. New York: Anchor Books.

Checkland P (1993) *Systems Thinking, Systems Practice*. New York: Wiley.

Checkland P (1999) Systems thinking. In: Currie W and Galliers B (eds) *Rethinking Management Information Systems: An Interdisciplinary Perspective*. New York: Oxford University Press, 45–56.

Cherns A (1976) The principles of sociotechnical design. *Human Relations* 29(8): 783–92.

Cherns A (1987) Principles of sociotechnical design revisited. *Human Relations* 40(3): 153–61.

Cilliers P (2002) *Complexity and Postmodernism Understanding Complex Systems*. London; New York: Routledge, 89–111.

Clemson B (1984) *Cybernetics—A New Management Tool*. Tunbridge Wells: Abacus.

Danseco E (2013) The five Cs for innovating in evaluation capacity building: Lessons from the field. *The Canadian Journal of Program Evaluation* 28(2): 107–18.

Davies R and Pickering J (2015) *Making Development Cooperation Fit for the Future. A Survey of Partner Countries* (OECD Development Co-operation Working Papers 20). Paris: OECD Publishing.

Eustachio JHPP, Caldana ACF, Liboni LB, et al. (2019) Systemic indicator of sustainable development: Proposal and application of a framework. *Journal of Cleaner Production* 241: 118383.

EvalPartners (2016) EvalAgenda 2020. Available at: www.evalpartners.org/sites/default/files/documents/EvalAgenda2020.pdf (accessed 1 December 2020). pp. 63–69.

Feinstein O (2011) On the development of national evaluation capacities. *Paper presented at the workshop evidence—based policy—making and the real world—A difficult match?* Cambridge Judge Business School, University of Cambridge, Cambridge, 2011.

Fisher C (2010) Between pragmatism and idealism: Implementing a systemic approach to capacity development. *IDS Bulletin* 41(3): 108–17.

Global Partnership for Effective Development Co-operation (2020) *How We Partner Together for Sustainable Development (2020–2022 Work Programme)*. Global Partnership for Effective Development Co-Operation. Available at: https://www.effectivecooperation.org/system/files/2020-11/GPEDC_2020-2022_Work_Programme_FINAL_15May%20%281%29.pdf (accessed 20 August 2020).

Grack Nelson A, King JA, Lawrenz F, et al. (2019) Using a complex adaptive systems perspective to illuminate the concept of evaluation capacity building in a network. *American Journal of Evaluation* 40(2): 214–30.

Haarich SN and del Castillo Hermosa J (2004) Development of evaluation systems —Evaluation capacity building in the framework of the new challenges of EU structural policy. In: *2004 regions and fiscal federalism*, Porto, 25–29 August, 1–17. Louvain-la-Neuve: European Regional Science Association (ERSA).

Hammond D (2002) Exploring the genealogy of systems thinking. *Systems Research and Behavioral Science* 19(5): 429–39.

Hartmann H (2020) High vulnerability to crisis: The results of the BTI 2020 in the context of COVID-19. Bertelsmann Stiftung Policy Brief 2020/01 April 2020. [Policy Paper]. Available at: http://aei.pitt.edu/103244/ (accessed 20 October 2020).

Heider C (2011) A conceptual framework for developing evaluation capacities: Building on good practice. In: Rist RC, Boily M-H and Martin F (eds) *Influencing Change. Building Evaluation Capacity to Strengthen Governance*. Washington, DC: The World Bank, 85–110.

Hitch C (1953) Sub-optimization in operations problems. *Journal of the Operations Research Society of America* 1(3): 87–99.

Holling CS (1996) Engineering resilience versus ecological resilience. In: National Academy of Engineering (ed.) *Engineering Within Ecological Constraints*. Washington: National Academy Press, 31–44.

Honderich T (ed.) (2005) *The Oxford Companion to Philosophy*, 2nd edn., new ed. Oxford: Oxford University Press.

Hummelbrunner R, Causemann B, Mutter T, et al. (2015) Systemische Ansätze in der evaluation. In: Wilhelm JL (ed.) *Evaluation Komplexer Systeme. Systemische Evaluationsansätze in Der Deutschen Entwicklungszusammenarbeit*. Potsdam: Universitätsverlag Potsdam, 33–80.

Imam I, LaGoy A and Williams B (2006) Introduction. In: Williams B and Imam I (eds) *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes, CA: American Evaluation Association, 3–11.

Kast FE and Rosenzweig JE (1972) General systems theory: Applications for organization and management. *Academy of Management Journal* 15(4): 447–65.

Katina P (2016) Systems theory as a foundation for discovery of pathologies for complex system problem formulation. In: Masys AJ (ed.) *Applications of Systems Thinking and Soft Operations Research in Managing Complexity from Problem Framing to Problem Solving*. Berlin: Springer, 227–67.

Keating CB (2009) Emergence in system of systems. In: Jamshidi M (ed.) *System of Systems Engineering*. Hoboken: John Wiley & Sons, 169–90.

Keating CB, Katina P, Hodge R, et al. (2020) Systems theory: Bridging the gap between science and practice for systems engineering. *INCOSE International Symposium* 30(1): 1017–31.

Korzybski A (1933) *Science and Sanity: An Introduction to Non-Artistotelian Systems and General Semantics*. Rouben Mamoulian Collection (Library of Congress). New York; Lancaster: Science Press Printing Co.

Krapp S and Geuder-Jilg E (2018) *Evaluation Capacity Development: A Systemic Project Approach by Deval in Latin America*. DEval Policy Brief 7/2018, Policy Brief. Bonn: German Institute for Development Evaluation.

Land T, Hauck V and Baser H (2009) Capacity development: Between planned interventions and emergent processes. Implications for development cooperation. Available at: https://ecdpm.org/wp-content/uploads/PMB-22-Capacity-Development-Between-Planned-Interventions-Emergent-Processes-March-2009.pdf (accessed 31 March 2020).

Laszlo A and Krippner S (1998) Systems theories: Their origins, foundations, and development. In: Jordan JS (ed.) *Systems Theories and a Priori Aspects of Perception*. Amsterdam: Elsevier Science, 47–74.

Laszlo E and Clark JW (1972) *Introduction to Systems Philosophy*. New York: Gordon and Breach.

Leeuw FL and Furubo JE (2008) Evaluation systems. What are they and why study them? *Evaluation* 14(2): 157–69.

Luhmann N (1995) *Social Systems* (trans. J Bednarz and D Baecker). Stanford, CA: Stanford University Press.

McCulloch WS (1965) *Embodiments of Mind*. Cambridge, MA: MIT Press.

Martin I and Carey JC (2018) Evaluation capacity within state-level school counseling programs: A cross-case analysis. *Professional School Counseling* 15(3). DOI: 10.1177/2156759X1201500302

Martinaitis Ž, Christenko A and Krauciuniene L (2019) Evaluation systems: How do they frame, generate and use evidence? *Evaluation* 25(1): 46–61.

Midgley G (2006) Systems thinking for evaluation. In: Williams B and Imam I (eds) *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes, CA: American Evaluation Association, 11–35.

Miller GA (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2): 81–97.

Miller JG (1978) *Living Systems*. New York: McGraw-Hill.

Mobus G and Kalton M (2015) Systems engineering. In: GE Mobus and MC Kalton (eds) *Principles of Systems Science*. New York: Springer, 289–96.

Morgan P (2005) *The Idea and Practice of Systems Thinking and Their Relevance for Capacity Development*. Maastricht: European Centre for Development Policy Management.

Newman MEJ (2005) Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46: 323–51.

Ortiz Aragón A and Macedo JCG (2010) A "systemic theories of change" approach for purposeful capacity development. *IDS Bulletin* 41(3): 87–99.

Pahl G, Beitz W, Feldhusen J, et al. (2011) *Engineering Design: A Systematic Approach*. Berlin: Springer.

Parsons T (1951) *Social Systems*. Glencoe: Free Press.

Pattee HH (1973) *Hierarchy Theory: The Challenge of Complex Systems*. New York: G. Braziller.

Pattee HH (1978) Biological systems theory: Descriptive and constructive complementarity. In: Klir GJ (ed.) *Applied General Systems Research*. New York: Plenum, 511–20.

Pérez Yarahuán G and Maldonado Trujillo C (eds) (2015) *Panorama de Los Sistemas Nacionales de Monitoreo y Evaluación en América Latina*, Primera edición. Coyuntura y ensayo. México, D.F: CIDE, Centro de Investigación y Docencia Económicas.

Preskill H and Boyle S (2008) A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation* 29(4): 443–59.

Prizzon A (2016) Where next for development effectiveness? Situating the debate in the country perspective. In: *2016 CAPE conference*, London, 18–19 October. Available at https://odi.org/en/events/2016-cape-conference-where-next-for-development-effectiveness/ (accessed 9 January 2022).

Raimondo E (2018) The power and dysfunctions of evaluation systems in international organizations. *Evaluation* 24(1): 26–41.

Reynolds M, Gates E, Hummelbrunner R, et al. (2016) Towards systemic evaluation. *Systems Research and Behavioral Science* 33(5): 662–73.

Richardson K (2004) Systems theory and complexity: Part 1. *Emergence: Complexity and Organization* 6: 75–9.

Rosenblueth A, Wiener N and Bigelow J (1943) Behavior. *Purpose and Teleology* 10(1): 18–24.

Serrano E (2020) The national evaluation subsystem of Costa Rica: Adaptations based on the SDGs. *Cuadernos del CLAEH* 39(112): 77–90.

Shannon CE (1948a) A mathematical theory of communication—Part 1. *Bell System Technical Journal* 27(3): 379–423.

Shannon CE (1948b) A mathematical theory of communication—Part 2. *Bell System Technical Journal* 27(4): 623–56.

Shannon CE and Weaver W (1949) *The Mathematical Theory of Communication*. Champaign, IL: University of Illinois Press.

Simon HA (1955) A behavioral model of rational choice. *The Quarterly Journal of Economics* 69(1): 99.

Simon HA (1956) Rational choice and the structure of the environment. *Psychological Review* 63(2): 129–38.

Simon HA (1974) How big is a chunk? By combining data from several experiments, a basic human memory unit can be identified and measured. *Science* 183(4124): 482–8.

Skyttner L (2005) *General Systems Theory: Problems, Perspectives, Practice*. Singapore; Hackensack, NJ: World Scientific.

Smuts JC (1927) *Holism and Evolution*, 2nd edn. London: MacMillan. Available at: https://ia802907. us.archive.org/10/items/holismandevoluti032439mbp/holismandevoluti032439mbp.pdf (accessed 29 January 2020).

Stockmann R (2018) *Der Evaluation Capacity Development Ansatz Des Centrums Für Evaluation (Ceval)*. 21, Arbeitspapier. Saarbrücken: Centrum für Evaluation (Ceval).

Tarafdar P and Bose I (2019) Systems theoretic process analysis of information security: The case of Aadhaar. *Journal of Organizational Computing and Electronic Commerce* 29(3): 209–22.

Tseng SH (2011) An assessment scale of ECB: Perspectives from the faculty and staff of technical colleges in Taiwan. *Evaluation and Program Planning* 34(2): 154–59.

Von Bertalanffy KL (1949) General system theory. *Biologie Generalis* 19: 114–29.

Von Bertalanffy KL (1950) An outline of general system theory. *The British Journal for the Philosophy of Science* 1: 134–65.

Von Bertalanffy KL (1967) General theory of systems: Application to psychology. *Information (International Social Science Council)* 6(6): 125–36.

Von Bertalanffy KL (1968) *General System Theory: Foundations, Development, Applications*, Rev. ed. New York: Braziller.

Von Bertalanffy KL (1972) The history and status of general systems theory. In: Klir GJ (ed.) *Trends in General Systems Theory*. New York: Wiley, 21–41.

Waddington CH (1957) *The Strategy of Genes: A Discussion of Some Aspects of Theoretical Biology*. London: Allen & Unwin.

Waddington CH (1968) Towards a theoretical biology. *Nature* 218(5141): 525–7.

Wan PY (2011) Emergence à la systems theory: Epistemological totalausschluss or ontological novelty? *Philosophy of the Social Sciences* 41(2): 178–210.

Whitney K, Bradley JM and Baugh DE (2015) Systems theory as a foundation for governance of complex systems. *Int J System of Systems Engineering* 6: 15–32.

Wiener N (1948) *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press.

Wilby J (1994) A critique of hierarchy theory. *Systems Practice* 7(6): 653–70.

Williams B and Imam I (eds) (2006) *Systems Concepts in Evaluation: An Expert Anthology*. Point Reyes, CA: American Evaluation Association.

Williams B and Van't Hof S (2016) *Wicked Solutions: A Systems Approach to Complex Problems*, 2nd ed. Wellington, New Zealand: Bob Williams.

Sarah D. Klier is the Technical Director of the ECD project Focelac+ at the German Evaluation Institute for Development Cooperation, DEval. She is an expert in Evaluation Capacity Development.

Raphael J. Nawrotzki is an Evaluator at the German Institute for Development Evaluation (DEval) with a specialization in the use of quantitative methods, including geodata, in evaluations.

Nataly Salas-Rodríguez, MA, Social Psychologist and Evaluator, works at DEval as a specialist in Evaluation Capacity Development.

Sven Harten, Deputy Director at DEval. He holds a PhD in Political Science from LSE and has more than 20 years of experience in the field of development evaluation.

Charles B. Keating is a Professor of Engineering Management and Systems Engineering at Old Dominion University in Norfolk, Virginia, USA. His expertise includes Systems Theory/Engineering and Complex System Governance.

Polinpapilinho F. Katina is an Assistant Professor in the Department of Informatics and Engineering Systems at the University of South Carolina Upstate (Spartanburg, South Carolina, USA).