“Leveling up” for change: Application of a multilevel system capacity change model to evaluate U.S. Department of Defense global health engagement activities

Eva Reed | James Harwood | Jessica L. A. Jackson | Tsega Gebreyesus

1 Uniformed Services University of the Health Sciences (USU), Center for Global Health Engagement (CGHE), Bethesda, MD, USA
2 Naval Medical Research Unit Three (NAMRU-3), Sigonella, Italy
3 Henry M. Jackson Foundation for the Advancement of Military Medicine (HJF), In support of the Uniformed Services University of the Health Sciences (USU), Center for Global Health Engagement (CGHE), Bethesda, MD, USA

Correspondence
Eva Reed, Uniformed Services University of the Health Sciences (USU), Center for Global Health Engagement (CGHE), 6720-B Rockledge Drive, Suite 350, Bethesda, MD, USA 20817.
Email: Eva.reed@usuhs.edu

Abstract
Evaluating within complex systems is challenging because of how complexity affects the identification and observation of outcomes. U.S. Department of Defense (U.S. DoD) capacity building global health engagements are often difficult to measure due to the conflation of levels of analysis and confounding variables, hindering the explanation of change effects. This article will illustrate two case examples where a boundary-driven systems framework was utilized to integrate systems thinking into U.S. DoD capacity building programs and associated evaluations. The findings from the first case led to developing a theory of change that was later tested and refined in the second case to establish the multilevel system (MLS) concept model. Based on these findings, the four distinct system boundaries and subcomponents of the MLS concept model were refined to include changes within the organizational system. The development of the MLS model allowed for the explicit framing of efforts, measurement and analysis, and the alignment of program activities and observed outcomes; while still allowing for the illumination of emergent change effects in a complex system.

INTRODUCTION

Despite the critical role that the U.S. Department of Defense (DoD) plays in building the capacity of international partners through global health engagement (GHE), the planning
and evaluation of these interventions often lack a comprehensive perspective of the intricacies of and interactions within complex systems (Diehl & Major, 2016). This article seeks to demonstrate the value of integrating systems thinking into the U.S. DoD GHE initiatives to delineate levels of analysis and measurement within and across systems. The article uses the example of a theory of change (TOC) developed in Fiji and tested and refined in the Solomon Islands to examine the application of a multilevel system (MLS) concept model into the planning, development, and evaluation of U.S. DoD GHEs. The illustration of the importance of boundaries seeks to decomplexify the change process by reducing the conflation of inferences across levels of analysis.

**Capacity building in practice**

“Assumptions are the Achilles heel of development programs …” (Archibald et al., 2016). U.S. DoD GHEs are characterized as any health-related activities conducted collaboratively between the DoD, and civilian and or military entities within partner nations (P.N.s). Capacity building GHEs are often besotted with measurement fallacies and errors due to incorrect assumptions about change (Paul et al., 2015; Ross, 2016; Sheperd, 2017). These change assumptions are commonly fraught with mismatched levels of measurement and analysis, resulting in incorrect reductionism or falling into the ecological fallacy trap (Lavrakas, 2008). DoD GHEs prescribe strategic (systems-level) change effects to technical level interventions, usually implemented at the individual or small group level (Diehl & Major, 2016).

Similarly, evaluations are expected to produce system-level results by making inferences from individual-level data. High-level return on investment evaluations that use aggregated individual-level data may show results at the group level, but this group result may not be reflective of the actual individual or sub-system-level change. For instance, if an evaluation aims to explore how teaching a technical class resulted in increased “system capacity,” the level of measurement is very technical. However, the level of analysis needed is rather strategic. To answer this type of evaluation question would necessitate analysis across multiple system hierarchies; thus, the problem cannot be explored only at the technical or individual component level. Crossing boundaries to measure across activity implementation and emergent change effect is necessary for complex change programs, meaning that boundaries must be explicit and the level of analysis clear for results to be significant and realistically understood.

Boundaries help to avoid these errors and fallacies by creating a better definition of the problem through the identification of focal phenomena of interest (MacDougall-Shackleton, 2011). Coherent problem framing better enables evaluation by focusing observation, clarifying levels of analysis and measurement, and identifying proximate and ultimate system change goals (MacDougall-Shackleton, 2011). The result is an enhanced ability to articulate dynamics across system levels (Eoyang & Berkas, 1998) because underlying relationships are more clearly observed and thus more easily measured (Kurtz & Snowden, 2003). A systems-driven approach using explicit boundary decisions breaks levels of change into levels of analysis using subsystems to organize the complexity.

**Boundary framing in the context of complexity evaluation**

Systems thinking is a means to analyze complex problems (Hummelbrunner, 2011; Jackson, 2001). It centers around three main concepts: Perspectives, inter-relationships, and boundaries (Williams & Imam, 2007). Perspective is how the system is seen –as part of the
bigger picture and across varying points of view (Fujito, 2010; Williams & Imam, 2007). Inter-relationships are internal system connections, the interconnected components that interact in a complex and dynamic way, understood in terms of system structures and functions (Fujito, 2010; Williams & Imam, 2007). Boundaries are explicit choices made to compartmentalize larger systems into smaller subsystems that drive the analytic and conceptual components of systems analysis (Midgley, 2003). Identifying boundaries starts with explicit decisions about the system level at which the intervention operates and what broader systems will be considered outside of this system, knowing that connections are passing through this arbitrary and imaginary border. Boundary choices allow for judgment about what is in and what is out of the frame of analysis (Midgley, 2003) and are evidenced through patterns (Cabrera, et al., 2006), often using critical system heuristics to acknowledge the influence of perspective in boundary selection (Fujito, 2010; Hummelbrunner, 2011; Ulrich, 2005; Williams & Imam, 2007).

Boundary framing is a systems thinking tool to improve understanding of the system (Fujito, 2010; Midgley, 2003). Boundaries provide an opportunity for multiple levels of problem identification and analysis (Hummelbrunner, 2011; Williams & Imam, 2007) by showing how parts connect within and outside of the system (Trochim et al., 2006). The use of explicit boundaries breaks down complex, messy problems into smaller, more manageable ones while also orienting intervention activities and efforts at appropriate system levels. Boundaries are not discrete lines but conceptual margins. They frame the system to look across system components while providing the ability to also work within specific aspects of the system (Bosomworth et al., 2017). This framing allows for independent methodology and evaluation at each level through compartmentalization (Hawkes et al., 2016).

Beyond a conceptual frame of reference, boundaries also serve as analytic categories to isolate the fractions or levels within a system, without losing the whole system perspective (Hummelbrunner, 2011). Without boundary choices, systems are too abstract to evaluate (Pawson, 2013). Thus, boundaries can reduce, and at times eliminate, confusion about levels of analysis (MacDougall-Shackleton, 2011). However, as boundaries are imposed on the system design, an important aspect of boundary identification is awareness of the influence of perspective and the role of power dynamics in applying boundary choices and how these decisions ultimately effect analysis and understanding of systems change (Ulrich, 2005).

Capacity building and complexity

For interventions into complex systems, the change environment consists of a system of interconnected actors and activities (Trochim et al., 2006). As such, the evaluation of complex interventions requires multiple methods and levels of inquiry, each involving different strategies to evaluate within the levels of change under development (Fujito, 2010; Peter & Swilling, 2014). This difficulty is often rooted in programs having to translate change activities implemented at the individual or technical level (the micro-level) into outcomes at the institutional or community levels (macro-level). Small programs or individual activities enable larger system effects, resulting in a faulty linear or misaligned level of analysis (Reimann, 2012) and often suffer from ecological fallacy or reductionism. The result is the inability to link the observed changes with intervention activities. This conflation across different levels makes it difficult to distinguish between problems inherent to the program itself and those intrinsic to the program evaluation (Eoyang & Berkas, 1998; Reimann, 2012).
As a solution, capacity building interventions for complex problems require multiple strategies of change (Sim et al., 2010), and evaluations must valuate this interconnected web of interventions and effects (Moore et al., 2019). To this end, capacity building evaluations must require specification of a change purpose, identify a bound evaluand, allow for complexity in methodologies, offer a scalable approach, and incorporate contextual nuance (Moore et al., 2019). Thus, a multilevel evaluation strategy is needed to assess multiple impacts, relationships, and interactions (Midgley, 2006). Further, because complex problems are not bound by a single discipline, to observe emergent change, a whole systems approach that is focused enough to achieve harmony between levels of analysis and levels of measurement is necessary (Jackson, 2001; Klein et al., 1994).

Evaluations of capacity building interventions must address the levels of complexity within systems to focus analysis and measurement (Hirsch et al., 2012). For example, levels within systems vary due to changes in the density of interconnections and relationships, thus becoming more complex with greater abstraction at higher levels but less complex at lower system levels (Fujito, 2010). Evaluations of interventions used to address complex problems cannot simply aggregate findings across project outcomes and offer a systemic change summary. Further, capacity is dependent on context such that each problem and intervention will act and react differently across different systems and influences. As a result, complex system problems lack a single solution, and often the goals or end states are just as difficult to entangle. Instead of a single or goal-oriented approach, a complexity aware evaluation strategy provides better methods to observe a multitude of effects within and across levels where clusters of system outcomes are inherent (Paz-Ybarnegaray & Douthwaite, 2017).

Within the context of the U.S. DoD GHE enterprise, the use of a multilevel approach to designing and evaluating interventions is paramount to teasing out activities and effects within and across the boundaries of each system. The purpose of the MLS model was to develop a systems-oriented tool able to provide heuristic support to address planning and evaluation concerns related to levels of analysis in capacity building interventions to address complex problems. The objective was to understand how observations within a complex system can connect to contributing activities to make explicit and credible value judgments. The specific aim was to critically examine system boundaries and provide factors for each boundary set to offer categories of change capacity within each level.

The following case examples provide details about the development of the MLS capacity change model to better understand the components of capacity building, the levels at which capacity building activities occur within DoD GHE, and the enablers necessary for change to occur within interconnected systems. The first case discusses the development of a capacity-building TOC, and the second illustrates the testing and refinement of this MLS model in a new environment.

**Multilevel system capacity change concept model**

**Case illustration 1: Conceptual model emergence and testing in Fiji**

A program evaluation of the integrated vector control initiative between the U.S. DoD and the Republic of Fiji’s Ministry of Health and National Vector Control Unit occurred between 2016 and 2017. The goals and objectives of this capacity building program were to assist Fiji in becoming a regional center of excellence in vector control and management for the Oceania region. Capacity change activities were implemented across individual/technical, organizational, and institutional levels. At the technical/individual level, workforce
development training aimed to increase knowledge, skills, and abilities. At the organizational level, a Train the Trainer (ToT) program intended to improve organizational workforce training capabilities. Additionally, the drafting of process and practice guidance (standard operating procedures) sought to increase the effectiveness and efficiency of services and to develop organizational leadership capacity further. At the institutional level, the program reviewed business plans to address funding priorities and policies, and foster cross-organizational partnerships (Ministry of Health (MoH), Ministry of Agriculture, Ministry of Defense, and the Ministry of Education).

Process evaluation was used to monitor for intervention performance and adjust activities based on the adaptive capacity of the organization (i.e., the Vector Control Unit and broader MoH). Formative monitoring points supported the summative evaluation conducted at program completion. This summative evaluation employed a mixed-methods approach to investigate outcomes compared to changes in performance using contribution analysis and process tracing with Bayesian updating. Process tracing used a tripartite conceptual framework that compared a TOC with the original hypothesis about change (the initial assessment) and the testable existence of that reality (evidence) (Stein & Valters, 2012; Wimbush et al., 2012). Evidence was collected from a series of interviews ($N = 11$) conducted after the conclusion of the program. Interviews with staff ($n = 7$) validated findings in program assessment scores, while leadership interviews ($n = 4$) provided triangulation for findings from staff interviews. Change activity, mechanisms, and effect patterns emerged around four distinct system boundaries: Individual, Organizational, Institutional, and Environmental/Community levels.

Using a post hoc boundary reflection from the Fiji evaluation and TOC, we wanted to explore these boundary patterns in a model that could serve as an overarching framework for implementation and evaluation but was not specific to one intervention. The model needed to be adaptive and able to span all types of capacity building initiatives. Boundary critique followed a critical systems approach to approximate our understanding of the system as it pertained to the construct of capacity building. Boundaries were examined regarding judgements about practical use from the U.S. program evaluation perspective based on how implemented activities would align with a capacity building framework. A literature review of capacity building factors provided a reference system to frame known conditions (those that emerged from the evaluation) into broader categories. Mutual understanding of these boundary choices was not sought from the partner nation due to time constraints, but is a necessary step in critical boundary framing and a suggested activity with subsequent implementation of the model. Despite certain limitations, an examination of system boundaries clarified where activities could orient towards specific change goals, and were ultimately used to delineate observation of effects both proximately and emergently. The result was a conceptual approach to how system capacity change might work (see Figure 1) given a set of standard system properties and clusters of outcomes (see Figure 2) from the perspective of a program planner or evaluator.

Case illustration 2: MLS capacity building model application in the Solomon Islands

The analytic framework for this evaluation case example was developed based on the findings from the capacity building program evaluation in Fiji. Capacity building-specific boundaries were then applied as “levels of analysis” to measure the contribution of complex change activities to observed change effects in a subsequent program evaluation in the Solomon Islands. This case illustration focused on efforts between the U.S. DoD, the
Solomon Islands Ministry of Health, and Medical Services National Vector Borne Disease Control Unit. The summative evaluation of this engagement aimed to test the validity of the MLS model developed in Fiji through replication in the Solomon Islands.

Data collection methods included a document review \( (N = 70), \) semi-structured interviews with course participants and key informants \( (N = 9), \) and semi-structured observations \( (N = 15). \) The research methodology used for the Solomon Islands evaluation followed an embedded QUAL-quan mixed method design (Creswell & Clark, 2007, p. 67). An embedded design allowed interpretation of qualitative evaluation findings to be verified against quantitative change evidence to allow for the triangulation of insights about the effectiveness of GHE on building capacity at the individual and organizational levels.

The emergent MLS model posited that there were four components of the capacity needed to obtain full sustainable capacity: (a) individual (e.g., an increase in knowledge of vector management and control), (b) organizational (e.g., buy-in from leadership, adequate resources, advocacy for change, and empowerment), (c) institutional (e.g., policy and partnerships), and (d) community or environmental (e.g., network and functional capacity). By evaluating the capacity-building engagement in the Solomon Islands, granular aspects relevant to capacity change emerged. While there were many salient findings across each of the levels after the Fiji case, this second case illustration will focus on how using the MLS model in the Solomon Islands led to a revised version of the MLS concept model (see Figure 1).

First, the concept of communication within the organizational system expanded to include both internal and external communication pathways. Based on the findings in the Solomon Islands, barriers to effective collaboration were related to both intra- and inter-organizational communication systems. Change in capacity at this level depended in part on ensuring cooperation between divisions and directing appropriate communication to external entities. As such, the capacity change model evolved to highlight the importance of internal and external communication in strengthening processes and practices. These boundary framing choices were mutually agreed upon with the United States and partner nation program staff.
Second, service capacity, as a function of organizational capabilities, expanded to encompass the concept of excellence in service delivery, leveraging an additional qualitative component emerging from the perspective of the partner nation. In the first case, service capacity and organizational performance had a more direct or linear relationship. This assumption was based on findings that increased or more efficient products, services, or programs lead to increased organizational service capacity. The second case example showed service capacity was increased by the number of products or services delivered, but also by the quality of those services. Quality of service was influenced by how well aligned the services were with external professional standards or credentialing bodies. Greater alignment ensured safety, increased the effectiveness of efforts, and demonstrated commitment to regional malaria-reduction goals and objectives. In this sense, service capacity

---

**Figure 2** Multilevel system (MLS) capacity change indicators and outputs across levels of analysis
was not just the ability to deliver services, but also the ability to deliver activities in line with external standards to affect broader environmental changes.

Finally, the Solomon Islands case posed the question of where and how to situate community-level health and public outreach efforts within the current capacity change model. At the country level, community health outreach and engagement activities were necessary drivers for creating population-level changes in health risk behaviors related to vector-borne diseases. Stakeholders within the program viewed community empowerment and education as a natural outcome of the training and necessary for ensuring sustainability. However, within the current model, the boundaries of the environmental system exclude population-level shifts in health outcomes, behaviors, or attitudes as an indicator of system resilience. Understanding the interactions between community outreach, organizational relationships, and associated outcomes would require further inquiry and identification of representative units of analysis at the community-level.

**Revised multilevel system capacity change concept model**

The MLS model set out to group individual, interpersonal, and collective processes that both comprise and influence change by setting preconditions for change to occur (Henry & Mark, 2003). Since capacity building is more adaptive than merely increasing the technical ability of an individual or groups of individuals, this model intended to provide an adaptive approach to whole systems change: *technical, operational, and strategic*.

At very high levels in the system, strategic effects must target multiple outcomes from improved subsystem capabilities across several institutions, organizations, groups of individuals, and other interacting variables. The process of achieving these more strategic outcomes requires improvements across systems. Results cannot be solely understood as singular or even small aggregates of engagements, but viewed instead as systemic change across multiple projects and programs with similar outcome goals.

Analysis within and across system levels was a critical aspect of the model to understand proximate and ultimate effects and cross boundary interactions. At the greater system level, analysis of strategic outcomes is used to evaluate results of multiple capabilities (e.g., improved relationships, reduced vulnerabilities) through the understanding of proximate causes (i.e., increased capabilities and sustainability of capacity within and across systems), and in terms of ultimate outcomes (i.e., improved system resiliency). Institutional-level outcomes focus on the sustainability of interconnected organizational and grouped capacities (Luederitz et al., 2017) though maximizing the integrative effects of policy and regulation (Jones et al., 2017) and cross-sector partnerships and communication (Selsky & Parker, 2005), and evidenced by changes in network capacities (Monaghan et al., 2017). In comparison, organizational level outcomes were considered in terms of unit or organizational level capabilities, such as group practices and processes, guidelines to affect change in group behaviors, to include leadership, resources, and internal communications. Lastly, individual level outcomes, being the most proximate measures, focused on key individual change effects, including individual practice capacity and resource utilization, and internal motivations and beliefs.

**Lessons learned**

Final revisions to the MLS model after the last case example resulted in a purpose-driven approach to evaluation by framing efforts, measurement and analysis, targeting intended
beneficiaries and stakeholders, focusing change opportunities, and intentional data and information collection. The final model (see Figure 1) can be used as an adaptive tool to work within various capacity change contexts by providing a semi-structured method to articulate change intentions and effects (see Figure 2). Use of the model increased program evaluability by allowing multiple methods of inquiry and also increased utilization of evaluation findings by targeting tiered results to the appropriate stakeholder groups.

Adaptability

This model provides a semi-structured process with options for methodological adaptability (Parsons et al., 2016) and analysis across dimensions and degrees of abstraction. Application of the model is shaped by locally relevant strategies, orienting whole and subsystems to unique and varying contexts. Each level uses actionable and specific inputs, which are unique to outputs and outcomes sought at that level. Effects can range from more simplistic to more complex to allow for a multiplicity of options to understanding capacity building change. This model is also useful as an assessment tool to identify current states of capacity to focus intervention activities within the level at which change is sought and target specific stakeholder groups to focal systems areas.

Evaluability

Use of this boundaries-driven model increased the evaluability of engagements through intentional framing of inquiry using focused evaluation questions, methods, and data. Boundaries highlighted clear direction on “when, what, and how” based on system simplification (Rogers, 2008), and to make appropriate inferences from the findings more easily. This framing allowed for assessment, monitoring, and evaluation at regular intervals using a phased approach to evaluation, engaging across multiple scales of capacity to interpret change (Peter & Swilling, 2014). In a phased approach, immediate results using proximate outcomes satisfy funders quickly. At the same time, phased evaluation can capture delayed effects that may not occur until well after the program has ended.

Utilization and learning

Boundary framing focused relevant outputs and outcomes to the most appropriate stakeholders resulting in increased consumption of findings due to targeted stakeholder action and review. Isolating what is at stake clarified which stakeholders were most relevant for each level of change. As well, the interdisciplinary orientation to levels of capacity building forced collaboration among stakeholders that may or may not routinely collaborate, creating a greater awareness of systemic issues and concerns. This cross-sector potential for collaboration about system level change goals enabled multiple organizations with connected practices to work together to develop sustainable institutional policies and mandates. Another effect of cross-sector collaboration was developing relationships and partnerships, which emerged as a strengthened network and subsequently increased the capacity for broader system change.

This article discussed how a system-based, boundary-driven capacity change concept is interpreted as an evaluation framework to detail how technical change-oriented activities at the individual user level interact with organizational, institutional, and community-level
processes across system boundary “levels” to more effectively understand complex system change (Boriani et al., 2017). The purpose of the MLS model is to make explicit the relationships between capacity building activities and change effects across system boundary levels. The aim is to reduce change complexity by organizing activity objectives within defined system boundaries, or levels, relevant to program outcomes. Specifically, this MLS model intends to make the evaluation of complex capacity change programs more effective through explicit framing and alignment of program activities and observed outcomes, while still allowing for illumination of emergent change effects.

REFERENCES
Archibald, T., Sharrock, G., Buckley, J., & Cook, N. (2016). Assumptions, conjectures, and other miracles: The application of evaluative thinking to theory of change models in community development. Evaluation and Program Planning, 59, 119–127.
Boriani, E., Esposito, R., Frazzoli, C., Fantke, P., Hald, T., & Rüegg, S. R. (2017). Framework to define structure and boundaries of complex health intervention systems: The ALERT project. Frontiers in Public Health, 5, 182.
Bozomworth, K., Owen, C., & Curnin, S. (2017). Addressing challenges for future strategic-level emergency management: Reframing, networking, and capacity-building. Disasters, 41(2), 306–323.
Cabrera, D. A. (2006). Boundary critique: A minimal concept theory of systems thinking. In Proceedings of the 50th Annual Meeting of the ISSS-2006, Sonoma, CA.
Creswell, J. W., & Clark, V. P. (2007). Designing and conducting mixed methods research. Sage.
Diehl, G., & Major, S. (2016). Finding a healthy match: A discussion on applying the appropriate level of analysis for global health engagement activities. Military Medicine, 181(3), 189–190. https://doi.org/10.7205/milmed-d-15-00403
Eoyang, G. H., & Berkas, T. (1998). Evaluation in a complex adaptive system. Managing complexity in organizations. Quorum Books.
Fujito, N. (2010). Beyond logframe: Using systems concepts in evaluation. Issues and prospects of evaluation for international development—Series IV. Foundation for Advanced Studies on International Development.
Hawkes, S., Aulakh, B. K., Jadeja, N., Jimenez, M., Buse, K., Anwar, I., … Whitworth, J. (2016). Strengthening capacity to apply health research evidence in policy making: Experience from four countries. Health Policy and Planning, 31(2), 161–170.
Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation’s influence on attitudes and actions. American Journal of Evaluation, 24(3), 293–314.
Hirsch, G., Homer, J., Milstein, B., Scherrer, L., Ingersoll, C., Landy, L., … & Fisher, E. (2012, July). Rethink health dynamics: Understanding and influencing local health system change. In 30th International Conference of the System Dynamics Society (Vol. 80). St. Gallen, Switzerland.
Hummelbrunner, R. (2011). Systems thinking and evaluation. Evaluation, 17(4), 395–403.
Jackson, M. C. (2001). Critical systems thinking and practice. European Journal of Operational Research, 128(2), 233–244.
Jones, C. M., Clavier, C., & Potvin, L. (2017). Adapting public policy theory for public health research: A framework to understand the development of national policies on global health. Social Science & Medicine, 177, 69–77.
Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. Academy of Management review, 19(2), 195–229.
Kurtz, C. F., & Snowden, D. J. (2003). The new dynamics of strategy: Sense-making in a complex and complicated world. IBM Systems Journal, 42(3), 462–483.
Lavrakas, P. J. (2008). Encyclopedia of survey research methods. Sage.
Luederitz, C., Schäpke, N., Wiek, A., Lang, D. J., Bergmann, M., Bos, J. J., … Westley, F. R. (2017). Learning through evaluation—A tentative evaluative scheme for sustainability transition experiments. Journal of Cleaner Production, 169, 61–76.
MacDougall-Shackleton, S. A. (2011). The levels of analysis revisited. Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1574), 2076–2085.
Midgley, G. (2003). Science as systemic intervention: Some implications of systems thinking and complexity for the philosophy of science. Systemic Practice and Action Research, 16(2), 77–97.
Midgley, G. (2006). Systemic intervention for public health. American Journal of Public Health, 96(3), 466–472. https://doi.org/10.2105/AJPH.2005.067660
Monaghan, S., Lavelle, J., & Gunnigle, P. (2017). Mapping networks: Exploring the utility of social network analysis in management research and practice. Journal of Business Research, 76, 136–144.
Moore, M., Parsons, B., & Jessup, P. (2019). How evaluators can use a complex systems lens to get “untrapped” from limiting beliefs and assumptions. American Journal of Evaluation, 40(1), 75–91.
Parsons, M., Glavac, S., Hastings, P., Marshall, G., McGregor, J., McNeill, J., … Stayner, R. (2016). Top-down assessment of disaster resilience: A conceptual framework using coping and adaptive capacities. *International Journal of Disaster Risk Reduction, 19*, 1–11.

Paul, C., Moroney, J. D., Grill, B., Clarke, C. P., Saum-Manning, L., Peterson, H., & Gordon, B. (2015). *What works best when building partner capacity in challenging contexts*. Rand Corporation.

Pawson, R. (2013). *The science of evaluation: A realist manifesto*. SAGE Publications Ltd. https://www.doi.org/10.4135/9781473913820

Paz-Ybarnegaray, R., & Douthwaite, B. (2017). Outcome evidencing: A method for enabling and evaluating program intervention in complex systems. *American Journal of Evaluation, 38*(2), 275–293.

Peter, C., & Swilling, M. (2014). Linking complexity and sustainability theories: Implications for modeling sustainability transitions. *Sustainability, 6*(3), 1594–1622.

Reimann, C. (2012). *Evaluability assessments in peacebuilding programming* (Working Papers on Program Review & Evaluation: #3). CDA Collaborative Learning Projects.

Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation, 14*(1), 29–48.

Ross, T. W. (2016). Enhancing security cooperation effectiveness: A model for capability package planning. *Joint Force Quarterly, 80*(1), 25–34.

Selsky, J. W., & Parker, B. (2005). Cross-sector partnerships to address social issues: Challenges to theory and practice. *Journal of Management, 31*(6), 849–873.

Sheperd, T. (2017, July 25). *Strategic ends not tactical means: Better ways to build partner capacity*. War Room-U.S. Army War College-Creative Thinking about National Security and Defense. https://warroom.armywarcollege.edu/articles/strategic-ends-not-tactical-means-better-ways-build-partner-capacity/

Sim, L. J., Parker, L., & Kumanyika, S. K. (Eds.). (2010). *Bridging the evidence gap in obesity prevention: A framework to inform decision making*. National Academies Press.

Stein, D., & Valters, C. (2012). Understanding Theory of Change in International Development, JSRP Paper 1. Justice and Security Research Programme, London: LSE. https://gsdrc.org/document-library/understanding-theory-of-change-in-international-development/

Trochim, W. M., Cabrera, D. A., Milstein, B., Gallagher, R. S., & Leisichow, S. J. (2006). Practical challenges of systems thinking and modeling in public health. *American Journal of Public Health, 96*(3), 538–546.

Ulrich, W. (2005). A mini-primer of critical systems heuristics. Rev. version of “Critical systems heuristics,” in H. G. Daellenbach, & R. L. Flood (eds.), *The Informed Student Guide to Management Science*, London: Thomson Learning, 2002, p. 72f. Expanded version in Werner Ulrich’s Home Page, http://wulrich.com/csh.html, first published 17 October 2005, last updated 30 April 2018.

Williams, B. & Imam, I. (Eds.). (2007). *Systems concepts in evaluation: An expert anthology*. EdgePress.

Wimbush, E., Montague, S., & Mulherin, T. (2012). Applications of contribution analysis to outcome planning and impact evaluation. *Evaluation, 18*(3), 310–329.

**AUTHOR BIOGRAPHIES**

**Eva Reed**, DrPH, is a military researcher and global health engagement planner and evaluator serving as the Director of Assessment, Monitoring, and Evaluation (AME) and an assistant professor with the Department of Preventive Medicine & Biostatistics at the Uniformed Services University for the Health Sciences, Center for Global Health Engagement, located in Bethesda, Maryland.

**James Harwood**, PhD, is an entomologist and military global health engagement planner serving as a research scientist with the Naval Medical Research Unit No. 3, located in Sigonella, Italy.

**Jessica L. A. Jackson**, MA, MPH, is an Assessment, Monitoring, and Evaluation (AME) Specialist with the Henry M. Jackson Foundation for the Advancement of Military Medicine and an Instructor in AME with the Department of Preventive Medicine & Biostatistics at the Uniformed Services University for the Health Sciences, located in Bethesda, Maryland.
Tsega Gebreyesus, PhD, MPH, is a formerly senior global health researcher with the Henry M. Jackson Foundation for the Advancement of Military Medicine and an assistant professor with the Department of Preventive Medicine & Biostatistics at the Uniformed Services University for the Health Sciences, located in Bethesda, Maryland.

How to cite this article: Reed, E., Harwood, J., Jackson, J. L.A., & Gebreyesus, T. (2021). “Leveling up” for change: Application of a multilevel system capacity change model to evaluate U.S. Department of Defense global health engagement activities. In E. F. Gates, M. Walton, & P. Vidueira (Eds.), Systems and Complexity-Informed Evaluation: Insights from Practice. New Directions for Evaluation, 2021, 39–50. https://doi.org/10.1002/ev.20458