

The state of capacity development evaluation in biodiversity conservation and natural resource management

ELEANOR J. STERLING, AMANDA SIGOUIN, ERIN BETLEY, JENNIFER ZAVALETA CHEEK
JENNIFER N. SOLOMON, KIMBERLEY LANDRIGAN, ANA L. PORZECANSKI
NORA BYNUM, BAILEY CADENA, SAMANTHA H. CHENG, KAYLIN R. CLEMENTS
RYAN FINCHUM, MALLORY GERESY, ANDRÉS GOMEZ, MARTHA GROOM
THIRZA A. C. LOFFELD, DANIEL C. MILLER, DOMOINA RAKOTOBÉ, MADHU RAO
RYAN ROBERTS, XOCO ANNA SHINBROT, ERIN WILLIGAN and MEGAN S. JONES

ELEANOR J. STERLING* (Corresponding author, orcid.org/0000-0003-2692-8264, sterling@amnh.org), AMANDA SIGOUIN (orcid.org/0000-0002-1272-1898), ERIN BETLEY (orcid.org/0000-0002-0009-8939), KIMBERLEY LANDRIGAN (orcid.org/0000-0001-5555-6581), ANA L. PORZECANSKI (orcid.org/0000-0002-5157-4579), SAMANTHA H. CHENG (orcid.org/0000-0003-1799-6310), ANDRÉS GOMEZ (orcid.org/0000-0003-2462-0880) and ERIN WILLIGAN American Museum of Natural History, Center for Biodiversity and Conservation, New York, NY, USA

JENNIFER ZAVALETA CHEEK (orcid.org/0000-0002-5375-1516) Department of Natural Resources, South Dakota State University, Brookings, USA

JENNIFER N. SOLOMON† (orcid.org/0000-0001-9103-1600) and RYAN FINCHUM* (orcid.org/0000-0002-7735-3803) Center for Protected Area Management, Colorado State University, Fort Collins, USA

NORA BYNUM Amazon Center for Environmental Education and Research, Bahama, USA

BAILEY CADENA‡ (orcid.org/0000-0001-6465-0828) and MALLORY GERESY (orcid.org/0000-0002-2024-3654) Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, USA

KAYLIN R. CLEMENTS, RYAN ROBERTS (orcid.org/0000-0003-1593-3172), XOCO ANNA SHINBROT§ (orcid.org/0000-0002-4290-3916) and MEGAN S. JONES (orcid.org/0000-0002-4284-3650) Human Dimensions of Natural Resources Department, Warner College of Natural Resources, Colorado State University, Fort Collins, USA

MARTHA GROOM (orcid.org/0000-0002-4044-4613) University of Washington Bothell, Bothell, USA

THIRZA A.C. LOFFELD (orcid.org/0000-0002-1531-1069) Durrell Institute of Conservation and Ecology, University of Kent, Canterbury, UK

DANIEL C. MILLER¶ (orcid.org/0000-0001-6812-0314) Keough School of Global Affairs, University of Notre Dame, Notre Dame, USA

DOMOINA RAKOTOBÉ (orcid.org/0000-0003-1718-0645) Wildlife Conservation Society, Antananarivo, Madagascar

MADHU RAO** (orcid.org/0000-0002-5360-4073) IUCN World Commission on Protected Areas Capacity Development Thematic Group, Gland, Switzerland

*Also at: IUCN World Commission on Protected Areas Capacity Development Thematic Group, Gland, Switzerland

†Also at: Human Dimensions of Natural Resources Department, Warner College of Natural Resources, Colorado State University, Fort Collins, USA

‡Also at: Department of Food, Agricultural, and Biological Engineering, The Ohio State University, Columbus, USA

§Also at: Graduate Degree Program in Ecology, Colorado State University, Fort Collins, USA

¶Also at: University of Illinois at Urbana-Champaign, Champaign, USA

**Also at: Wildlife Conservation Society, Singapore, and National University of Singapore, Singapore

Received 6 November 2020. Revision requested 28 January 2021.

Accepted 12 April 2021.

Abstract Capacity development is critical to long-term conservation success, yet we lack a robust and rigorous understanding of how well its effects are being evaluated. A comprehensive summary of who is monitoring and evaluating capacity development interventions, what is being evaluated and how, would help in the development of evidence-based guidance to inform design and implementation decisions for future capacity development interventions and evaluations of their effectiveness. We built an evidence map by reviewing peer-reviewed and grey literature published since 2000, to identify case studies evaluating capacity development interventions in biodiversity conservation and natural resource management. We used inductive and deductive approaches to develop a coding strategy for studies that met our criteria, extracting data on the type of capacity development intervention, evaluation methods, data and analysis types, categories of outputs and outcomes assessed, and whether the study had a clear causal model and/or used a systems approach. We found that almost all studies assessed multiple outcome types: most frequent was change in knowledge, followed by behaviour, then attitude. Few studies evaluated conservation outcomes. Less than half included an explicit causal model linking interventions to expected outcomes. Half of the studies considered external factors that could influence the efficacy of the capacity development intervention, and few used an explicit systems approach. We used framework synthesis to situate our evidence map within the broader literature on capacity development evaluation. Our evidence map (including a visual heat map) highlights areas of low and high representation in investment in research on the evaluation of capacity development.

Keywords Biodiversity conservation, capacity development, conservation outcomes, evaluation, evidence map, literature review, systems approach, systematic map

Supplementary material for this article is available at doi.org/10.1017/S0030605321000570

Introduction

Capacity development can be defined as ‘the process through which the abilities of individuals, institutions, and societies to perform functions, solve problems, and set and achieve objectives in a sustainable manner are strengthened, adapted, and maintained over time’ (UNDP, 2010, p. 32). The term capacity development implies progress from a base of existing knowledge and skills, emphasizes that the process for building capacity can be dynamic and occur at multiple scales, including individual, organizational and systems levels (CADRI, 2011), and considers how to sustain the capacity that is built (Whittle et al., 2012). Capacity development for individuals includes strengthening technical and interpersonal skills, fostering knowledge acquisition and awareness, and effecting change in attitudes, behaviours and cultural norms. Capacity development also addresses organizational and system-level practices through use of organizational assessments, competence registers, and cross-sectoral and integrated planning approaches, among others (Kapos et al., 2008, 2009; Porzecanski, *in press*). It encompasses many different formats, such as formal training, peer learning, mentorship, community engagement, learning networks, communities of practice, and others. Capacity development can also encompass physical resources but we do not address this as it lies outside our scope.

Research identifies capacity development as crucial to achieve biodiversity conservation goals (e.g. Barnes et al., 2016; Gill et al., 2017; Geldmann et al., 2018; Coad et al., 2019). Although capacity development is considered important to many conservation programmes, there is limited evidence around its specific impacts. Rather, relevant studies often restrict their analyses to assessing broad measures of capacity. For example, Geldmann et al. (2018) and Gill et al. (2017) found that positive conservation outcomes in protected areas correlated with capacity. They measured capacity using general metrics such as number of staff and the presence or absence of a training plan. However, it is not clear which underlying factors contributed to these successes. Capacity comprises more than the sum of these commonly used quantitative measures. The impacts of capacity development intervention on knowledge, behaviour and attitudes are more difficult to measure and, therefore, not typically considered in such studies. Thus, even though there is evidence that capacity development benefits conservation, it is often not clear what interventions can most effectively and efficiently support and/or achieve conservation outcomes. A more thorough understanding of the outcomes from capacity development interventions (Ferraro & Pattanayak et al., 2006) could help address the twin challenges of limited resources allocated to conservation and an imperative to meet sustainability goals.

A comprehensive summary of capacity development interventions and their evaluation can help build evidence-

based guidance to support conservation. Evaluating interventions is critical to determine what works and when; evaluations can focus on specific indicators (e.g. change in attitudes) and/or seek to uncover what would have happened in the absence of an intervention (i.e. a counterfactual; Ferraro & Pattanayak, 2006). Rigorous, measurable and impact-focused evaluation of capacity development investments helps practitioners make better design and implementation decisions and supports funding agencies to allocate resources more effectively (Aring & DePietro-Jurand, 2012; Loffeld et al., *in press*). Evaluators may employ different approaches to frame an intervention and design an evaluation. For instance, some researchers posit that evaluation requires advanced planning to set forth clear objectives and goals for outputs and outcomes, so that monitoring, evaluation and research are conducted to assess achievement (Baylis et al., 2016). The process for achieving goals need not be linear in nature; evaluations that apply approaches and tools from systems thinking and complexity may help to focus on different areas for action compared to a linear approach. These approaches to evaluation emphasize feedback and learning throughout the intervention, view capacity as an emergent process of a complex adaptive system, and adapt the evaluation to reflect this dynamic nature (Black et al., 2013; Knight et al., 2019; Patton, 2019).

Our goal is to provide the first comprehensive assessment of how capacity development in conservation and natural resource management has been evaluated. We seek to collate and synthesize existing knowledge to explore the following questions: What types of impacts/outcomes are being evaluated? Who are the subjects of evaluations? What types of methods are being used for evaluation? To what extent do current evaluations incorporate causal models and/or systems approaches?

We supplement this case study analysis with a review of documents that are focused on guidance and lessons learnt for effective capacity development evaluation. We assess these guidance documents to contextualize our findings on conservation and natural resource-specific synthesis within the broader field of capacity development evaluation.

Methods

We followed an a priori protocol adapted from the systematic map approach of Collaboration for Environmental Evidence (2018). We used two systematic review methods for our approach: framework synthesis and an evidence map. Framework synthesis is used to shape an initial conceptual framework for an issue, around which other aspects of a study can be organized and that can be amended as part of an iterative process (Brunton et al., 2020). An evidence map is a thematic collection of research articles. The map identifies and reports the distribution and occurrence of existing

evidence on a broad topic area, question, theme, issue or policy domain (McKinnon et al., 2016; Saran & White, 2018).

Framework synthesis

We used framework synthesis (Brunton et al., 2020) to develop broad themes about capacity development evaluation and situate our findings within the broader literature. We used expert solicitation to familiarize ourselves with key topics, ideas and publications regarding evaluation of capacity development in the fields of conservation, development, and planning, monitoring and evaluation. We contacted the World Commission on Protected Areas Capacity Development Evaluation Working Group, which comprises more than 20 global experts in capacity development and evaluation, through e-mail and directly at workshops. This expert solicitation ultimately led to a core literature dataset that provided lessons learnt, guidance, best practices and other recommendations. We used the resulting dataset for two purposes. Firstly, we developed a preliminary framework consisting of various implementation approaches for a range of capacity development interventions that led to various outcomes, including conservation. These interventions are evaluated in different ways and contexts (Fig. 1). We used this framework to develop our codebook system, described below and in Supplementary Material 1. Secondly, we undertook a qualitative inductive thematic review (Thomas & Harden, 2008) of the dataset to identify patterns and distil key themes on guidance for effective capacity development evaluation, to situate the findings from our evidence map within this broader context in the discussion.

Evidence map search strategy

We determined search terms based on our framework synthesis and a scoping process that included expert review and focused on peer-reviewed and grey literature published since 2000. Starting in 2015, we carried out a pilot search and scoping exercise to hone the search strategy for relevant articles. Iterative searches were conducted subsequently and reviewed by outside experts, with the final set of searches performed in June 2019. All searches were restricted to English language works. We searched for articles from a literature database (Web of Science, Clarivate, Philadelphia, USA), a search engine (Google Scholar, Google, Mountain View, USA), and several grey literature/specific organization portals (OpenGrey, Collections at United Nations University, U.S. Environmental Protection Agency, My Environmental Education Evaluation Resource Assistant, and United States Agency for International Development). We supplemented our core database search strategy by contacting experts to solicit relevant case study literature (specifically experts from the World Commission on Protected Areas Capacity

Development Evaluation Working Group as detailed above). For further details on the search strategy, see Supplementary Material 1.

Evidence map inclusion/exclusion process and critical appraisal

Abstract/title inclusion/exclusion analysis Three team members (EJS, AS, EB) used Colandr (Cheng et al., 2018) for screening titles and abstracts for relevance. Colandr removes duplicates, allows review by multiple screeners, and applies machine learning and natural language processing algorithms to sort articles according to relevance, which can reduce the time taken to screen compared with traditional methods (Cheng et al., 2018). At least two team members (EB, AS or EJS) assessed each article; in instances of disagreement, the involved team members discussed the specific study and a consensus decision was reached in consultation with a third-party reviewer when necessary. We erred on the side of including a paper if it seemed to meet our inclusion criteria: (1) actions/interventions related to capacity development, (2) explicit evaluation of capacity development outcomes or impacts, (3) in the fields of conservation and/or natural resource management, and (4) type of capacity development recipient (included recipients were conservation and natural resource management professionals, educators, Indigenous peoples and local community stewards, pre-professionals (e.g. interns, young people involved in training programmes), community scientists, and conservation and Indigenous peoples and local community organizations. We use the inclusive term community scientist because citizenship, or the perception that an individual may or may not be a citizen, is not germane and can impact inclusion efforts. In terms of recipient type, we chose to exclude evaluation of youth environmental education programmes (e.g. for ages 5–14 in classroom or informal settings) as there is an extensive body of literature on the evaluation of environmental education (see systematic reviews by Stern et al., 2014; Ardoin et al., 2015; Thomas et al., 2018; Monroe et al., 2019).

Full-text inclusion/exclusion and critical appraisal of case studies The full text of articles was further assessed by one reviewer (EJS, AS, EB or MSJ) for relevance and articles were excluded from the dataset if they met any of the following additional criteria: (1) insufficient detail on evaluation, (2) focus only on youth environmental education programmes, or (3) the only evaluation was comparing outcomes of volunteers with professionals. In instances where an intervention was evaluated in more than one publication, we chose case studies that had the most detail related to our inquiry and those most recently published. For literature deemed relevant at the full-text level, we performed a critical appraisal to

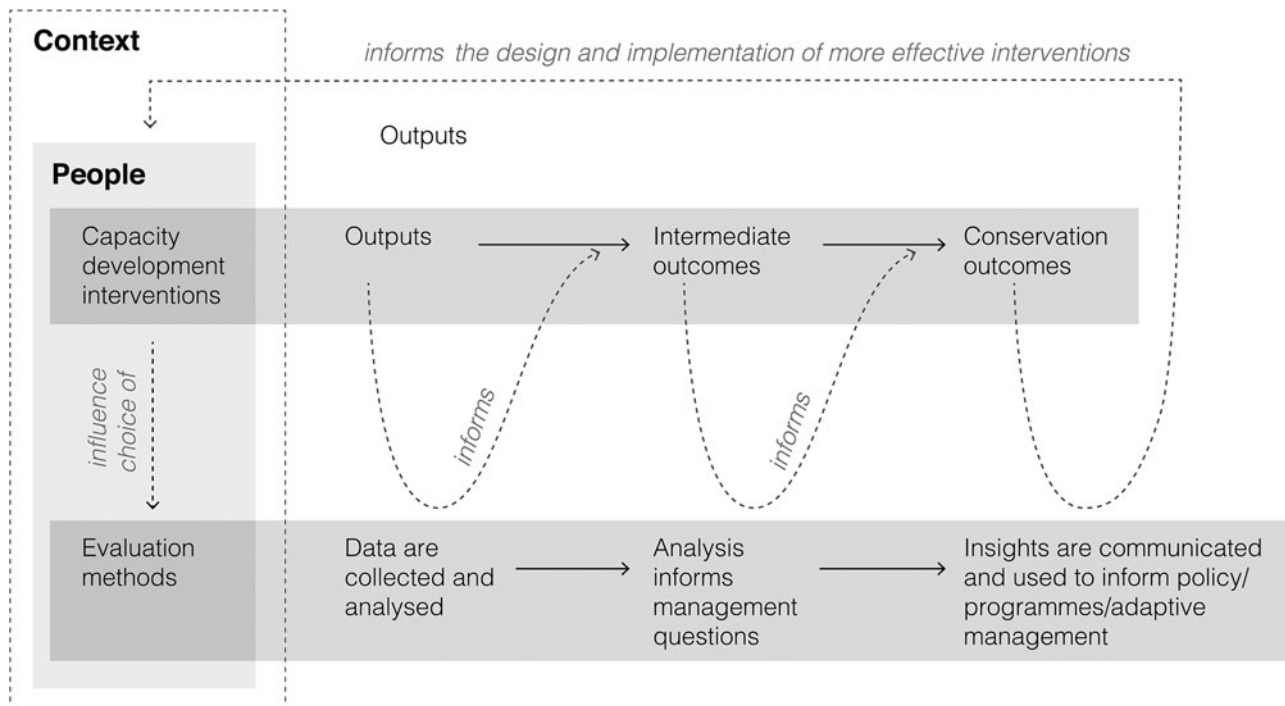


FIG. 1 Causal model for evaluation of capacity development interventions. Choice of capacity development intervention and evaluation methods is influenced by people, who implement and are targeted by the intervention and evaluation, and by the broader context (we note that the influence of people and broader context extend throughout the causal model; for our purposes we are emphasizing how they affect intervention type and evaluation method related to decisions in particular). As shown, the capacity development intervention leads to outputs, intermediate outcomes, and ultimately to conservation outcomes. The evaluation of outputs and outcomes along the intervention results chain in turn informs ongoing adaptive management as well as the design of more effective interventions in formative and summative ways (modified from CAML, 2020).

ensure our study included literature appropriate for our research aims by assessing the following four principles of quality: conceptual framing, validity of study design, quality of data sources, and quality of analysis (see Supplementary Material 1 for additional detail). Reviewers assessed each principle of quality to determine inclusion in the coding analysis.

Evidence map data collection

For included studies, we developed an a priori codebook system to extract data for 21 key fields, covering general information on each study (e.g. geography, sector), the capacity development intervention, evaluation methods, types of data and analyses, and output and outcome categories assessed. Coding fields included a combination of a priori categories and free text. We used the framework synthesis to develop a pilot codebook and refined it through an iterative process with the multidisciplinary coding team. A subset of the authors coded the included articles (Supplementary Material 1).

Evidence map analysis

We calculated descriptive statistics for the final case studies to assess the overall landscape of capacity development

evaluation in conservation and natural resource management. We created a relational database to sort and cross reference the data for the coded variables across the case studies. We used this database to generate a visual heat map of the distribution and frequency of all variables from the 21 coded fields in comparison with every other coded variable.

Results

We identified 1,206 potentially relevant studies, of which 85 case studies were included in the final evidence map (Fig. 2, Supplementary Material 1 & 2). Key portions of the visual heat map of evidence are presented in Supplementary Material 3 and Supplementary Figs 1–8. For the framework synthesis thematic review, our dataset consisted of 12 articles (Supplementary Material 1). The themes from this dataset have been integrated into our discussion.

Summary of included case studies

Of the 85 included studies, 42% assessed programme interventions at one specific locality and 42% examined interventions that took place at multiple localities within a country.

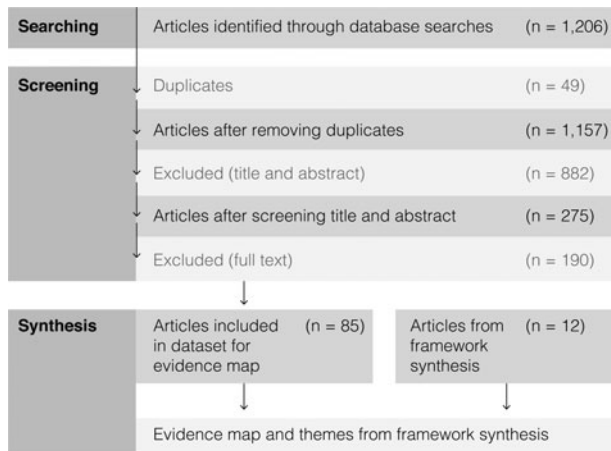


FIG. 2 Search and inclusion process for the evidence map and framework synthesis. For details on the criteria used for inclusion, exclusion, and critical appraisal see Methods and Supplementary Material 1.

Only a few assessed interventions on multi-national (10%) or global (6%) scales. The geographical distribution of the case studies was predominantly in North America (31%), Asia (25%) and Africa (20%), with fewer studies in Oceania, Europe and South America (Fig. 3). All interventions involved conservation and/or natural resource management sectors, and many were multi-sectoral, including education (19%), economic development (18%), agriculture (16%) and governance/policy (13%).

We found that 42% of the included studies described an explicit causal model. A causal model could be inferred in most of the remaining cases (46%); in 10 instances (12%) it was difficult to identify a causal model. We did not assess if the explicit or inferred models accurately reflected the causal thinking employed in the study.

Capacity development interventions overview

The capacity development interventions in the evidence map were wide-ranging and some incorporated multiple approaches. Interventions most often included training, informal education, and learning networks and communities of practice; mentoring and formal adult education were less common (Fig. 4a). Training included initiatives to train the trainers and on-the-job-training, including geospatial equipment, species monitoring and interpersonal skills. Community scientists were also trained in how to monitor coral reefs, mammals, invasive species and water quality. Informal education included conservation corps opportunities for high school and university students, landowner and farmer workshops on management techniques, community workshops about alternative livelihoods, and work experience such as internships. Mentoring primarily involved peer mentoring of new participants by programme alumni.

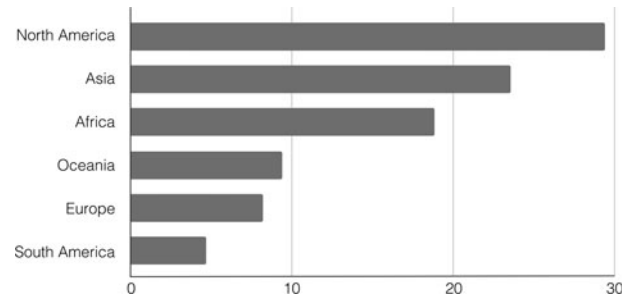


FIG. 3 Geographical distribution of the case studies (n = 85), excluding five case studies that involved multiple continents.

Formal adult education included establishing graduate and certificate programmes in the fields of conservation and natural resource management. Additionally, 8% of cases included capacity methods labelled as ‘other’, such as partnerships and workshops with peer learning and networking.

Of the cases that focused on specific global regions, the use of training as a capacity development method was highest in Africa (100% of cases used training), followed by Asia (80%) and North America (72%). Capacity development interventions were most often implemented by individuals working in non-profit organizations, government and higher education (Fig. 4b). Interventions were also implemented by more than one organization (32%).

Capacity development interventions mostly targeted local community members and employees working within governments or non-profit organizations, followed by multi-stakeholder working groups, teachers, pre-professionals, facilitators and interest groups (Fig. 4c). Community members included farmers, fishing clubs, and Indigenous-led groups. External actors regularly initiated the capacity development interventions (58%), meaning the need for capacity building was determined by an outside party, rather than initiated internally from the group that received the intervention (28%), although who initiated the study was difficult to determine in 14% of cases. Viewing the relationship between types of capacity development approaches and capacity development targets in terms of the approaches used, mentoring was not used to build capacity in teachers, facilitators/trainers or interest groups, and it was used most often in cases targeting pre-professionals (Fig. 5).

Evaluation overview

In parallel with our analysis of the different types of capacity and intended beneficiaries, we also analysed the methods used in evaluations. In our evidence map only a small proportion (16%) included a control group. In 68% of the case studies, the organization that ran the capacity development intervention also conducted the evaluation of the intervention; in 32% of cases the organization tasked with evaluation differed from the implementing group. Evaluations

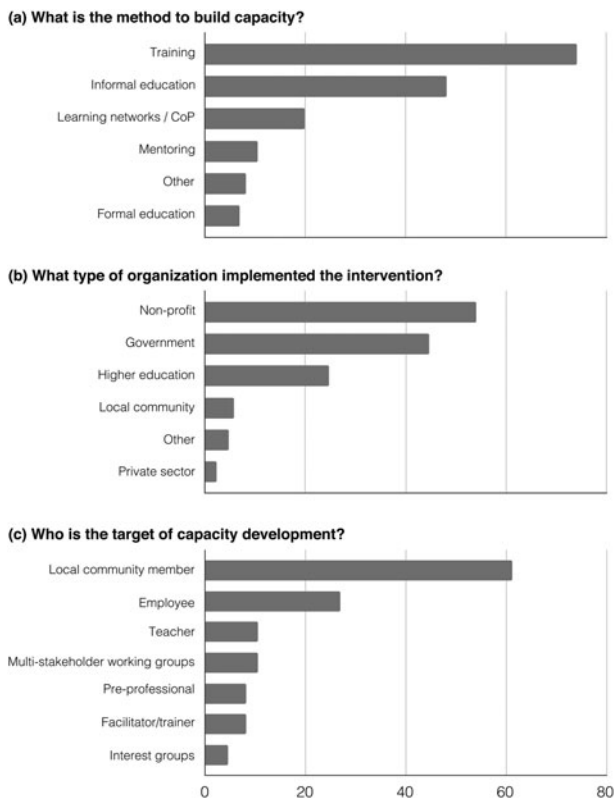


FIG. 4 Overview of capacity development intervention factors as a per cent of cases overall ($n = 85$). A case study could be in more than one category (so the sum of all bars can be $> 100\%$). (a) What is the method to build capacity? (b) What type of organization implemented the intervention? (c) Who is the target of capacity development? CoP, communities of practice.

were most commonly led by individuals working in higher education, followed by employees in non-profit organizations and government (Fig. 6a).

Case studies used a range of evaluation tools, with most employing more than one (80%), most commonly interviews and one-off questionnaires or tests (Fig. 6b). Only 22% of case studies used pre- and post-tests to evaluate capacity development; fewer used physical measurements, graphical tools such as concept mapping, or diaries/scrap-books. Some case studies employed other types of evaluation tools such as those from systems modelling. Approximately 65% of the case studies evaluated interventions with interactive oral methods (interviews and/or focus groups) whereas c. 20% used solely written methods (questionnaires and pre- and post-tests). By region, Africa had the highest use of oral methods (75%) and their use was lowest in Europe (47%).

In most case studies, the evaluation tools were used to evaluate the individual targets of the capacity development intervention directly, who most often were community members and employees (Fig. 6c). In 14% of case studies, the tools were used on a different group from the target of

the intervention. For example, managers were interviewed to understand how the capacity of an employee had developed.

We found that in $> 60\%$ of all cases the evaluation took place either immediately after the intervention or within 1 year of the intervention (Fig. 6d), with few evaluations (6%) taking place ≥ 10 years afterwards. These figures ($n = 53$) exclude cases where there was insufficient data to determine the timing gap and also evaluations that assessed multiple cohorts, leading to varied times post intervention. We separately assessed the articles with multiple cohorts and available timing data ($n = 12$) taking into account the longest time frame for each; these evaluations were evenly spread over 2–4 years, 5–9 years, and ≥ 10 years post intervention. Regarding frequency, most cases had one evaluation point (74%), with fewer having two (11%), or three or more evaluations (8%); the frequency was unclear in 7% of articles. All evaluations encompassed summative (at the end of intervention) components, and only a small portion were evaluated during the intervention (18%).

Analysis approach

Approximately half (51%) of the studies used a combination of quantitative and qualitative methods to evaluate capacity development interventions, 26% analysed outcomes with only quantitative methods and 23% using only qualitative methods. Quantitative analyses generally included descriptive and inferential statistics, and qualitative analyses included methods or approaches such as grounded theory, thematic coding, and most significant change. Data were validated through cross-verification, or using more than one data collection method, in 54% of the studies.

Outputs and outcomes

In our sample, evaluation studies assessed six categories of outputs and outcomes (Fig. 7). A majority of studies assessed three categories: learning, knowledge or awareness outcomes; application of new behaviour or skills; and attitude change. A smaller portion assessed participant satisfaction regarding the intervention. More studies assessed medium- and long-term, non-conservation related outcomes, such as poverty alleviation, than conservation related outcomes (41 vs 15%).

Conservation related outcomes included empirical evidence, observations, and/or perceptions regarding the desired state of biodiversity and/or threats to biodiversity. Fourteen studies assessed these conservation outcomes criteria, of which 79% targeted local community members. The majority of the 14 case studies (72%) that assessed conservation outcomes were implemented by a single organization, and the majority (79%) of evaluations were predominantly led by individuals working in higher education, sometimes working with other organizations

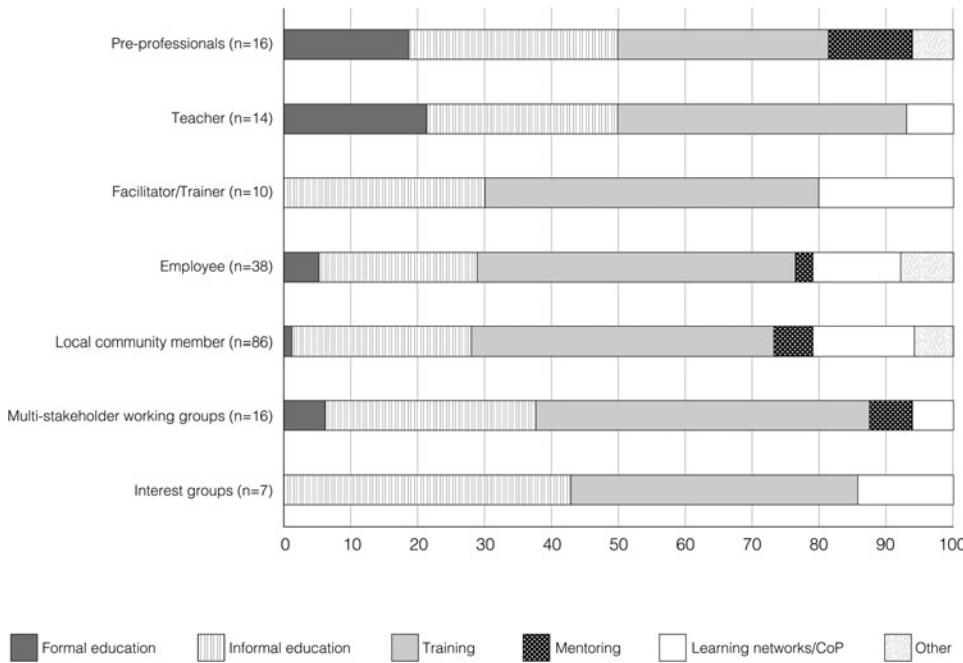


FIG. 5 Relationship between types of capacity development and capacity development targets; n is the number of times a capacity development approach was coded with a specific target. The figure maps which capacity development approaches were used to build capacity in each targeted group. CoP, communities of practice.

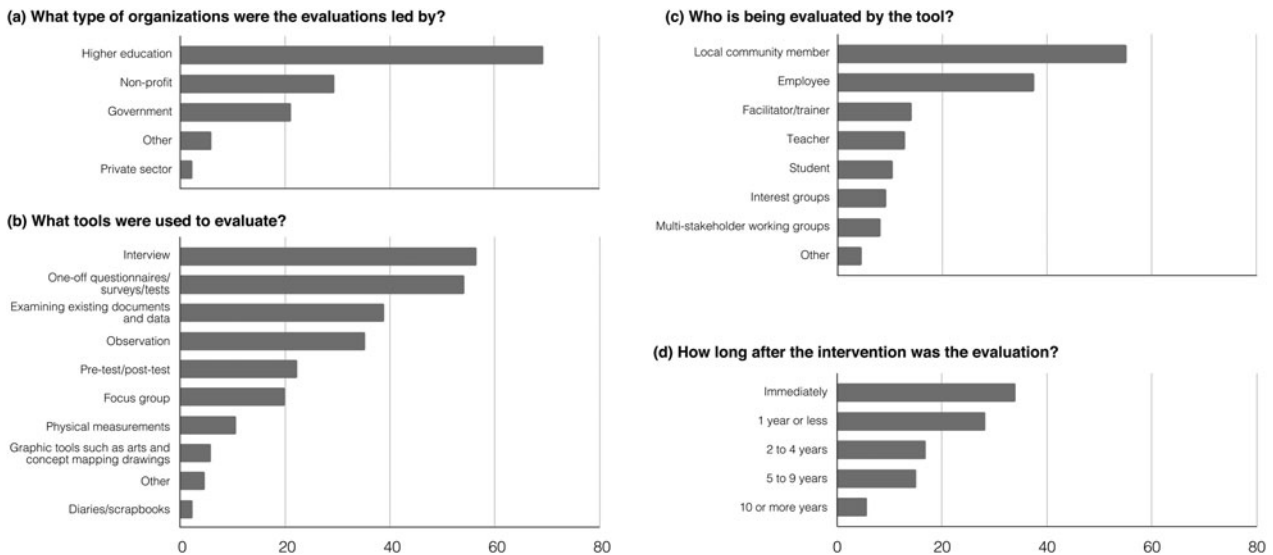


FIG. 6 Overview of capacity development evaluation factors as a per cent of cases overall (n = 85, unless specified otherwise). (a) What type of organizations were the evaluations led by? (b) What tools were used to evaluate? (c) Who is being evaluated by the tool (in 14% of case studies, the tools were used on a different group from the target of the intervention)? (d) How long after the intervention was the evaluation (n = 53; 15 cohort studies and 17 studies without timing information were excluded)?

in government or the private sector. Over 75% of the conservation outcomes cases described evaluation efforts in Africa and Asia. Among the 16 case studies in Africa, 31% assessed conservation outcomes, followed by 20% of the 20 studies in Asia, and of the 25 case studies in North America, only 8% assessed conservation outcomes. Most of the 14 cases (79%) used training alone or in combination with other methods to develop capacity. Half of the conservation outcomes involved direct measurement of conservation data, and others

assessed perceptions of the state of biodiversity (21%), changes in threats to biodiversity (14%), and changes in biodiversity data quality or conservation practices (14%). Of the 10 case studies assessing conservation outcomes that provided enough detail to assess the timing of the evaluation, 60% evaluated ≥ 5 years after intervention. In comparison, across all 65 cases that provided timing information, only 29% had ≥ 5 year time frames. For case studies assessing conservation outcomes, the presence of a clear or explicit

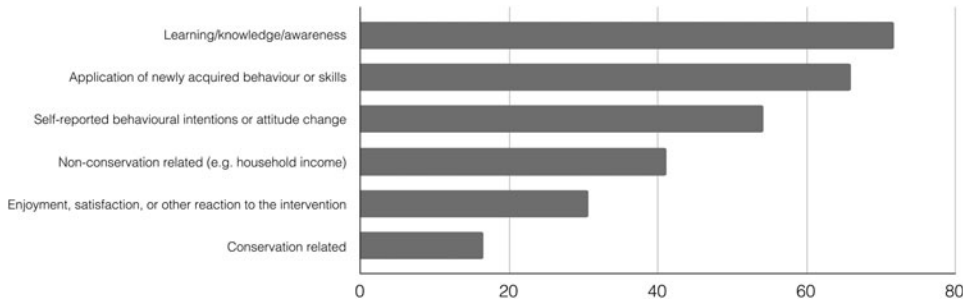


FIG. 7 Categories of outputs/outcomes assessed among evaluations of capacity development interventions shown as a per cent of cases overall ($n = 85$). Case studies are included multiple times if they assessed multiple categories.

causal model was slightly higher compared with the overall cases (50 vs 42%).

Systems approaches

We identified an explicit systems approach in only 6% of the case studies in our evidence map. For example, Elbakidze et al. (2015) used systems thinking methods to evaluate strategic spatial planning initiatives in municipalities in Sweden, including the use of causal loop diagrams to assess stakeholder participation. In their analysis of an integrated conservation and development project in Guyana, Mistry et al. (2010) used a system viability approach for monitoring and evaluation of impact.

Although few cases in our evidence map took an explicit systems approach, we note that systems approaches to evaluation can consist of several implicit elements, including the level of assessment (e.g. individual, organizational, system) and the degree to which an evaluation considered intervening variables (i.e. those outside the study's control) that can impact the end result of capacity development interventions. Of all 85 case studies, we found that the majority of the case studies assessed at the individual level (85%). Organizational (40%) and systems levels (14%) were less well-represented. Two or more levels were assessed in 34% of the case studies, and of this subset of case studies ($n = 29$), 14% assessed all three levels, 59% assessed individual and organizational levels, 17% assessed individual and systems level, and 10% assessed organizational and systems levels. Of all 85 case studies, almost half (47%) assessed how outcomes were mediated by intervening variables that can affect the end result of capacity development interventions. Of these, 36% assessed individual-level intervening variables (e.g. motivation to learn, level of expertise) and 26% considered contextual factors (e.g. work environment, socio-economic status).

Discussion

From our review, we identified that the literature on capacity development evaluation in conservation and natural resource management spans: (1) case studies of single

evaluated interventions, (2) aggregate studies that consider only general categories of capacity development (e.g. presence or absence of a training plan), and (3) general recommendations and lessons learnt in monitoring, planning and evaluation of capacity development. Here we provide the first overview documenting the state of the available evidence of capacity development evaluation in conservation. Our purpose was to identify the evidence and existing knowledge gaps to inform future research on and implementation of capacity development interventions.

In our framework synthesis, we found that much of the evaluation literature on capacity development has focused on formal, results-based management and project framework approaches such as use of standardized templates and tools (Watson, 2006; Baser & Morgan, 2008; Carleton-Hug & Hug, 2010; Ling & Roberts, 2012). There is consensus that evaluation emphasis must shift from inputs (e.g. funding) and outputs (e.g. number of workshops) to outcomes resulting from the interventions (Ferraro & Pattanayak, 2006). Outcomes can better reflect the actual conservation impacts of an intervention because they link the intervention to changes in biodiversity or people and place.

Our evidence map highlighted that much of the literature does not provide sufficient description on how capacity interventions lead to conservation outcomes, including the factors that modulate the effectiveness of capacity development interventions. Most studies assessed learning and behavioural change outcomes from capacity development interventions. Of the few studies that assessed conservation outcomes directly, the majority were related to interventions undertaken by academics and targeted at building the capacity of local community members, a similar pattern to the cases overall. This may be because local community members are frequently the stakeholders most likely to have a strong influence over biodiversity (Garnett et al., 2018), in comparison with employees, pre-professionals and trainers, and because research on community-based natural resource management has increased over the past 4 decades (Milupi et al., 2017).

The small number of studies evaluating the conservation outcomes of capacity development investments may relate to how difficult it is to make a direct connection between a single intervention and conservation outcomes. Challenges

associated with evaluating these connections may stem from temporal scale problems; there is often a lag between capacity development interventions and when the status of conservation targets can be expected to change. Monitoring and evaluation resources rarely continue for sufficient time to assess these outcomes in the long-term (Verissimo et al., 2018). Our evidence map shows that about half of the evaluations took place within a year of the intervention, and relatively few had a time horizon of ≥ 5 years. This gap may be related to project funding timelines that often exclude post-project follow-up, and is particularly problematic given that the goal of many conservation efforts is to achieve enduring change. Funding constraints can hinder proper design, implementation and assessment of capacity development evaluation (Carleton-Hug & Hug, 2010). Going forward, evaluators and donors should work together to align the knowledge and resources necessary for high quality evaluation. Such evaluations should provide accountability to internal and external audiences, as well as important insights regarding how to improve future interventions. These efforts could be based on a recognition that rigorous evaluation can meet the needs of both donors and practitioners by contributing to more effective allocation of funding and greater achievement of conservation goals (Aring & DePietro-Jurand, 2012; Loffeld et al., *in press*). This can be supported by approaching evaluation through a systems lens that emphasizes the dynamic feedback loops between learning and impact over the life of a project or programme (Black et al., 2013; Knight et al., 2019; Patton, 2019).

In our framework synthesis, we identified arguments made by some authors that effective evaluation relies on (1) clarifying achievable programme objectives, (2) identifying a causal model and testable hypotheses, and (3) defining the purpose of evaluation along with performance indicators and data collection protocols (Baylis et al., 2016). Evaluators should develop credible causal models (i.e. models that map all pathways between actions and outcomes, including intermediate outputs, and describe the assumptions and mechanisms between steps; Cheng et al., 2020), to help identify relevant outputs and outcomes, and support effective evaluation. We note the need to be flexible and consider advice to prioritize feedback, learning and adaptation even when starting with clear, predetermined objectives. Such thinking draws from systems evaluation approaches that view capacity as an emergent process (Lusthaus et al., 1999; Watson, 2006).

Our framework synthesis indicated that to fully understand the impacts of capacity development efforts, conservationists may need to apply a systems approach to evaluation that explores the dynamic relationship between investment in one set of capacity development interventions and effects at other levels. As an example, organizational readiness for change may affect results of individual capacity development.

Our framework synthesis also highlighted the importance of considering readiness for effective evaluation, related to the unique socio-political context of a capacity development programme (Lusthaus et al., 1999; Carleton-Hug & Hug, 2010; Pearson, 2011). In our evidence map, there is a significant focus on evaluations of individual-level capacity development, with less attention at the organizational or systems levels. Although not every intervention needs to result in systems-level outcomes, and not every evaluation needs to assess systems-level outcomes, it is important to acknowledge that systems-level forces shape every capacity development initiative. For instance, political and social systems of power influence whether and how investments in individual capacity have an effect in organizations or on conservation outcomes (Porzecanski et al., *in press*).

Our framework synthesis highlighted the importance of considering intervening variables when evaluating capacity development (Ferraro & Pattanayak, 2006). Fewer than half of the studies in our evidence map assessed how intervening variables affected the outcomes of capacity development interventions. This may be because of limited resources and time, as well as limited knowledge about these approaches and their importance in evaluation. Our framework synthesis found that evaluating capacity development requires a comprehensive analytical framework that accounts for the individual, organizational and systems levels in assessing the complex process of learning, adaptation, and attitudinal change. Nevertheless, it can be challenging to identify appropriate measurement tools and indicators to measure these complexities (Mizrahi, 2004).

However, the majority of the studies in our database did not include an explicit causal model linking interventions and expected outcomes. Lack of causal models can lower effectiveness, limit replicability, and hinder the development of the evidence base about effectiveness of interventions and best practices (Watson, 2006; Cheng et al., 2020). Employing systems approaches to capacity development could improve effectiveness through their emphasis on careful consideration of linkages and causal relationships in understanding where to invest scarce conservation resources.

Our framework synthesis identified the importance of using a diversity of research approaches, including a mixed-methods design, conducting different types of evaluation for different user groups and needs, and making use of participatory methodologies (Baser & Morgan, 2008; Carleton-Hug & Hug, 2010; Horton, 2011). We found some support for this in our evidence map, where most studies used more than one evaluation method, and some seemed to tailor methods to the capacity development recipients. A few studies using both written and oral evaluation methods cited lower comprehension from written methods than oral methods (Scholte et al., 2005; Amin & Yok, 2015),

showing that the method of evaluation can affect the results and thus should be considered carefully.

In terms of capacity development interventions, we noted that across all our evidence map cases, the predominant recipients were community members and, although most studies used a diversity of techniques, the most frequent type of intervention was training. Other forms of capacity development, such as peer learning or learning networks that might mirror existing forms of exchange within a community, were less common. This may result, in part, from the fact that capacity development interventions in this study were generally not initiated internally (i.e. by those within the group that received the intervention) and this may limit community leadership, ownership and decision-making power (Dobson & Lawrence, 2018). Evaluations in our evidence map were undertaken mostly by academic organizations, perhaps because this group benefits professionally from publishing research literature. We note that the types of cases in our evidence map might have been affected by our literature search having focused exclusively on English language sources.

Our framework synthesis highlighted the need to shift the emphasis of evaluation from accountability to external stakeholders such as donors, towards accountability to internal stakeholders and to evaluation that informs learning and programme improvement (Baser & Morgan, 2008; Simister & Smith, 2010; Horton, 2011). Although programme accountability is important to inform future funding decisions, programmes can only adapt, grow and improve when an evaluation focuses on understanding what worked, what did not, and why. In addition to why the evaluation is being done, it is important to consider who is performing an evaluation. In the majority of our evidence map cases, the groups leading an intervention also undertook evaluation of their own work; only in one-third of the cases was an evaluation done by an independent group. Self-evaluation is important, particularly formative evaluation that helps with adaptive management. However, independent evaluators can bring fresh eyes to an initiative and help identify gaps as well as successes that could be overlooked by those closest to a project. The fact that few cases had regular, periodic evaluation suggests there is a need for more evaluation to support course corrections for ongoing interventions. Evaluation of longer-term interventions could also uncover important cyclical or nonlinear patterns in changes to intervention participants or the social-ecological systems within which they are nested.

The need for more robust knowledge-sharing among evaluators was a theme identified in our framework synthesis and further supported by our evidence map. Over 20% of the articles were excluded (at full-text assessment) because of a lack of evaluation details. Even among the case studies included, we found insufficient information to code certain fields. Although this lack of information may be a result of word limitations in peer-reviewed

journals, it hampers the ability to build the evidence base on capacity development evaluation. Evaluators in conservation should share more detailed information on methodologies and the results of evaluations, to inform the evaluation field and practitioners (Carleton-Hug & Hug, 2010; Horton, 2011). Evaluators can overcome word limitations by providing detailed supplementary information, or publishing this information in public data repositories or in grey literature. Furthermore, the challenges of linking capacity development investment to achieving conservation outcomes point to the possibility that we will need to be able to track progress towards conservation outcomes at scale, which would involve collaboration to establish indicators that are meaningful across multiple capacity development initiatives.

In conclusion, our evidence map and framework synthesis clarifies the need for better evidence about the effectiveness of capacity development, its linkages to conservation objectives, and the contextual factors that affect the implementation of capacity interventions. Some of this evidence will need to come from evaluations and assessments, and we suggest that future efforts consider developing credible causal models (Cheng et al., 2020). We also suggest that more evaluations employ systems approaches that draw from the theories, methods, tools and approaches of systems thinking and complexity, and utilization-focused, developmental, and principles-focused evaluation (e.g. Knight et al., 2019; Patton, 2019). We recognize that operationalizing capacity development evaluation in conservation is a long-term, collaborative endeavour, but without it, we risk continuing to overlook or misunderstand the impacts of conservation interventions.

Future research should examine the evidence summarized in this evidence map to assess the impacts of capacity development, looking at a range of outputs and outcomes and seeking lessons learnt for practitioners. To develop decision-making tools that give capacity development implementers guidance on which evaluation approach to use under what circumstances, we need to understand better what outcomes emerge from different types of capacity development interventions on different targeted groups, as assessed by different evaluation methods. This information would facilitate the design of more effective capacity development interventions and more useful evaluations.

We view the evidence presented here as setting a baseline for the ongoing and long-term efforts to improve capacity development evaluation in conservation and natural resource management. It is a starting point for future review and evidence synthesis of when, how and why capacity development interventions and their evaluation are most impactful. In the meantime, we hope this evidence map provides critical information to help evaluators and practitioners in their work to develop the capacity needed to protect and manage social-ecological systems globally.

Acknowledgements We thank Sarah Trabue and Kathryn Powlen for help coding, compiling and sorting relevant data; Audrey Ek for performing initial scoping searches and participating in early coding that helped develop the codebook; Nikolas Merten for assisting with coding of conservation outcomes; and Nadav Gazit for graphic design expertise and assistance with figures. We gratefully acknowledge support from the John D. and Catherine T. MacArthur Foundation for this research (JZC; DCM).

Author contributions Conceptualization and design: EB, NB, JZC, RF, MG, MSJ, KL, TACL, DCM, ALP, XAS, AS, JNS, EJS; data collection, analysis and interpretation, writing and revision: all authors.

Conflicts of interest None.

Ethical standards This research abided by the *Oryx* guidelines on ethical standards.

References

- AMIN, V.L. & YOK, M.C.K. (2015) Thematic interpretation approach in environmental adult education. *Procedia-Social and Behavioral Sciences*, 167, 261–266.
- ARDOIN, N.M., BIEDENWEG, K. & O'CONNOR, K. (2015) Evaluation in residential environmental education: an applied literature review of intermediary outcomes. *Applied Environmental Education & Communication*, 14, 43–56.
- ARING, M. & DEPIETRO-JURAND, R. (2012) *Promising Youth Development Strategies*. USAID, Washington, DC, USA.
- BARNES, M.D., CRAIGIE, I.D., HARRISON, L.B., GELDMANN, J., COLLEN, B., WHITMEE, S. et al. (2016) Wildlife population trends in protected areas predicted by national socio-economic metrics and body size. *Nature Communications*, 7, 12747.
- BASER, H. & MORGAN, P. (2008) *Capacity, Change and Performance Study Report*. ECDPM Discussion Paper 59B. European Centre for Development Policy Management, Maastricht, The Netherlands.
- BAYLIS, K., HONEY-ROSÉS, J., BÖRNER, J., CORBERA, E., EZZINE-DE-BLAS, D., FERRARO, P.J. et al. (2016) Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, 9, 58–64.
- BLACK, S.A., GROOMBRIDGE, J.J. & JONES, C.G. (2013) Using better management thinking to improve conservation effectiveness. *International Scholarly Research Notices*, 2013, 784701.
- BRUNTON, G., OLIVER, S. & THOMAS, J. (2020) Innovations in framework synthesis as a systematic review method. *Research Synthesis Methods*, 11, 316–330.
- CADRI (CAPACITY FOR DISASTER REDUCTION INITIATIVE) (2011) *Basics of Capacity Development for Disaster Risk Reduction*. International Environment House II, Geneva, Switzerland.
- CAML (CONSERVATION ACTIONS AND MEASURES LIBRARY) (2020) *High Level Standard Results Chain on Evaluation, Effectiveness Measures & Learning*. miradishare.org/ux/project/cmp-conservationaction-2016-00015?nav1=toc&nav2=diagrams [accessed 30 June 2021].
- CARLETON-HUG, A. & HUG, J.W. (2010) Challenges and opportunities for evaluating environmental education programs. *Evaluation and Program Planning*, 33, 159–164.
- CHENG, S., AUGUSTIN, C., BETHEL, A., GILL, D., ANZAROOT, S., BRUN, J. et al. (2018) Using machine learning to advance synthesis and use of conservation and environmental evidence. *Conservation Biology*, 32, 762–764.
- CHENG, S.H., MCKINNON, M.C., MASUDA, Y.J., GARSIDE, R., JONES, K.W., MILLER, D.C. et al. (2020) Strengthen causal models for better conservation outcomes for human well-being. *PLOS ONE*, 15, e0230495.
- COAD, L., WATSON, J.E., GELDMANN, J., BURGESS, N.D., LEVERINGTON, F., HOCKINGS, M. et al. (2019) Widespread shortfalls in protected area resourcing undermine efforts to conserve biodiversity. *Frontiers in Ecology and the Environment*, 17, 259–264.
- COLLABORATION FOR ENVIRONMENTAL EVIDENCE (2018) *Guidelines and Standards for Evidence Synthesis in Environmental Management*. Version 5.0. Collaboration for Environmental Evidence, Bangor, UK.
- DOBSON, C. & LAWRENCE, S. (2018) *Our Voices, Our Environment. The State of Funding for Women's Environmental Action*. Global Greengrants Fund and Prospera International Network of Women's Funds, London, UK.
- ELBAKIDZE, M., DAWSON, L., ANDERSSON, K., AXELSSON, R., ANGELSTAM, P., STJERNQUIST, I. et al. (2015) Is spatial planning a collaborative learning process? A case study from a rural-urban gradient in Sweden. *Land Use Policy*, 48, 270–285.
- FERRARO, P.J. & PATTANAYAK, S.K. (2006) Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLOS Biology*, 4, e105.
- GARNETT, S.T., BURGESS, N.D., FA, J.E., FERNÁNDEZ-LLAMAZARES, Á., MOLNÁR, Z., ROBINSON, C.J. et al. (2018) A spatial overview of the global importance of Indigenous lands for conservation. *Nature Sustainability*, 1, 369.
- GELDMANN, J., COAD, L., BARNES, M.D., CRAIGIE, I.D., WOODLEY, S., BALMFORD, A. et al. (2018) A global analysis of management capacity and ecological outcomes in terrestrial protected areas. *Conservation Letters*, 11, e12434.
- GILL, D.A., MASCIA, M.B., AHMADIA, G.N., GLEW, L., LESTER, S.E., BARNES, M. et al. (2017) Capacity shortfalls hinder the performance of marine protected areas globally. *Nature*, 543, 665–669.
- HORTON, D. (2011) Evaluating capacity development. *Capacity*, 43, 5–9.
- KAPOS, V., BALMFORD, A., AVELING, R., BUBB, P., CAREY, P., ENTWISTLE, A. et al. (2008) Calibrating conservation: new tools for measuring success. *Conservation Letters*, 1, 155–164.
- KAPOS, V., BALMFORD, A., AVELING, R., BUBB, P., CAREY, P., ENTWISTLE, A. et al. (2009) Outcomes, not implementation, predict conservation success. *Oryx*, 43, 336–342.
- KNIGHT, A.T., COOK, C.N., REDFORD, K.H., BIGGS, D., ROMERO, C., ORTEGA-ARGUETA, A. et al. (2019) Improving conservation practice with principles and tools from systems thinking and evaluation. *Sustainability Science*, 14, 1531–1548.
- LING, C.M. & ROBERTS, D. (2012) *Guide to Evaluating Capacity Development Results: A Collection of Guidance Notes to Help Development Practitioners and Evaluators Assess Capacity Development Efforts*. World Bank Institute, Washington, DC, USA.
- LOFFELD, T.A.C., HUMLE, T., CHEYNE, S.M. & BLACK, S.A. (in press) Professional development in conservation: an effectiveness framework. *Oryx*, in press.
- LUSTHAUS, C., ADRIEN, M.-H. & PERSTINGER, M. (1999) Capacity development: definitions, issues and implications for planning, monitoring and evaluation. *Universalia Occasional Paper*, 35, 1–21.
- MCKINNON, M.C., CHENG, S.H., DUPRE, S., EDMOND, J., GARSIDE, R., GLEW, L. et al. (2016) What are the effects of nature conservation on human well-being? A systematic map of empirical evidence from developing countries. *Environmental Evidence*, 5, 1–25.
- MILUPI, I., SOMERS, M.J. & FERGUSON, J.W.H. (2017) A review of community-based natural resource management. *Applied Ecology and Environmental Research*, 15, 1121–1143.

- MISTRY, J., BERARDI, A., SIMPSON, M., DAVIS, O. & HAYNES, L. (2010) Using a systems viability approach to evaluate integrated conservation and development projects: assessing the impact of the North Rupununi Adaptive Management Process, Guyana. *Geographical Journal*, 176, 241–252.
- MIZRAHI, Y. (2004) *Capacity Enhancement Indicators: Review of the Literature*. World Bank Institute, Washington, DC, USA.
- MONROE, M.C., PLATE, R.R., OXARART, A., BOWERS, A. & CHAVES, W.A. (2019) Identifying effective climate change education strategies: a systematic review of the research. *Environmental Education Research*, 25, 791–812.
- PATTON, M.Q. (2019) *Blue Marble Evaluation: Premises and Principles*. Guildford Press, New York, USA.
- PEARSON, J. (2011) *Training and Beyond: Seeking Better Practices for Capacity Development*. Organisation for Economic Co-operation and Development, Paris, France.
- PORZECANSKI, A., STERLING, E., COPSEY, J., APPLETON, M., BARBORAK, J., BRUYERE, B. et al. (in press) A systems framework for planning and evaluating capacity development in conservation: recommendations for practitioners. *Oryx*, in press.
- SARAN, A. & WHITE, H. (2018) Evidence and gap maps: a comparison of different approaches. *Campbell Systematic Reviews*, 14, 1–38.
- SCHOLTE, P., DE GROOT, W.T., MAYNA, Z. & TALLA, M. (2005) Protected area managers' perceptions of community conservation training in West and Central Africa. *Environmental Conservation*, 32, 349–355.
- SIMISTER, N. & SMITH, R. (2010) *Monitoring and Evaluating Capacity Building: Is it really that difficult?* Praxis Paper 23. International NGO Training and Research Centre, Oxford, UK.
- STERN, M.J., POWELL, R.B. & HILL, D. (2014) Environmental education program evaluation in the new millennium: what do we measure and what have we learned? *Environmental Education Research*, 20, 581–611.
- THOMAS, J. & HARDEN, A. (2008) Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 45.
- THOMAS, R.E., TEEL, T., BRUYERE, B. & LAURENCE, S. (2018) Metrics and outcomes of conservation education: a quarter century of lessons learned. *Environmental Education Research*, 25, 172–192.
- UNDP (UNITED NATIONS DEVELOPMENT PROGRAMME) (2010) *Measuring Capacity*. UNDP, New York, USA.
- VERÍSSIMO, D., BIANCHESSI, A., ARRIVILLAGA, A., CADIZ, F.C., MANCAO, R. & GREEN, K. (2018) Does it work for biodiversity? Experiences and challenges in the evaluation of social marketing campaigns. *Social Marketing Quarterly*, 24, 18–34.
- WATSON, D. (2006) *Monitoring and Evaluation of Capacity and Capacity Development*. European Centre for Development Policy Management, Maastricht, The Netherlands.
- WHITTLE, S., COLGAN, A. & RAFFERTY, M. (2012) *Capacity Building: What the Literature Tells Us*. The Centre for Effective Services, Dublin, Ireland.