



Exploring stakeholder perceptions of conservation outcomes from alternative income generating activities in Tanzanian villages adjacent to Eastern Arc Mountain forests

Sainsbury, K.; Burgess, Neil David; Sabuni, F.; Howe, C.; Puis, E.; Killenga, R.; Milner-Gulland, E. J.

Published in:
Biological Conservation

DOI:
[10.1016/j.biocon.2015.06.001](https://doi.org/10.1016/j.biocon.2015.06.001)

Publication date:
2015

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](#)

Citation for published version (APA):
Sainsbury, K., Burgess, N. D., Sabuni, F., Howe, C., Puis, E., Killenga, R., & Milner-Gulland, E. J. (2015). Exploring stakeholder perceptions of conservation outcomes from alternative income generating activities in Tanzanian villages adjacent to Eastern Arc Mountain forests. *Biological Conservation*, 191, 20-28.
<https://doi.org/10.1016/j.biocon.2015.06.001>



Exploring stakeholder perceptions of conservation outcomes from alternative income generating activities in Tanzanian villages adjacent to Eastern Arc Mountain forests



K. Sainsbury^{a,*}, N.D. Burgess^{b,c,f}, F. Sabuni^d, C. Howe^e, E. Puis^d, R. Killenga^d, E.J. Milner-Gulland^a

^a Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, UK

^b UNEP-WCMC, 219 Huntington Road, Cambridge CB3 0DL, UK

^c WWF US, 1250 24th Street, Washington, DC, USA

^d Eastern Arc Mountains Conservation Endowment Fund, P.O. Box 6053, Morogoro, Tanzania

^e Centre for Environment and Biodiversity Research, University College London, WC1E 6BT, UK

^f Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, Copenhagen, Denmark

ARTICLE INFO

Article history:

Received 27 January 2015

Received in revised form 21 May 2015

Accepted 1 June 2015

Available online 25 June 2015

Keywords:

Income-generating activities

Evaluation

Conservation

Tanzania

Alternative livelihoods

ABSTRACT

Critical evaluation of the impact of conservation actions is essential to meet the challenges posed by the biodiversity crisis. Conservationists need to understand which interventions work or fail, and how to improve them in order to invest limited funds wisely. Alternative income-generating activities (IGAs) are widely implemented within conservation and development projects, but their impact is rarely evaluated. The “ranked outcomes” evaluation methodology converts qualitative information on planned and realised outcomes into a score for comparison between projects. We test this methodology in two ways using a set of small scale IGAs implemented in communities adjacent to the Uzungwa Scarp proposed Nature Reserve in the Tanzanian Eastern Arc Mountains. The first approach used an independent evaluator and the second assessed project impacts from the perspective of target communities. Both evaluations rated Tree Planting as the most socially beneficial IGA, followed by Fish Farming. However, there was a high level of heterogeneity of perception between and within stakeholder groups (implementers and target communities), both in terms of which outcomes were most important and how well they had been achieved. Ranked outcomes emerged as a flexible framework that defines the terms of the evaluation for all stakeholders from the outset, even in cases when evaluation and clear goal-setting are omitted from original project design and planning. It can be modified for use as a component of rigorous impact assessment, to incorporate perspectives of all stakeholders, and provides important insights in data-poor situations and where baselines are not available.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Evaluation of conservation projects has become a focal issue for policy makers at the macro level, with the Convention on Biological Diversity (CBD) driving the agenda (Mascia et al., 2014). At a micro level, conservation practitioners have limited budgets and there is both a moral duty to spend money wisely and a practical need to do so cost effectively (James et al., 1999). Rigorous, evidence-based analysis is a pre-requisite to demonstrating that progress in conservation is being made (Sutherland et al., 2004) and also to validate that the strategies being deployed to achieve conservation goals are appropriate and do not have unintended

consequences for people living in the area (Ferraro and Pattanayak, 2006); indeed a natural extension of this is involving communities affected by an intervention in the process of the evaluation itself.

In spite of its importance, it is widely accepted that evaluation has been under-utilised in conservation (Stem et al., 2005; Mascia et al., 2014). In one of the few published analyses of the determinants of project success, a meta-analysis of 136 published evaluations concluded that project design is particularly important for the success of community-based conservation projects (Brooks et al., 2012). In the last decade, a growing number of organisations have published best practice frameworks to address this critical need for effective project design and evaluation. Examples include IUCN's Framework for evaluating Protected Area effectiveness (Hockings et al., 2006) and the GEF's Monitoring and Evaluation Policy (GEF, 2010). In addition, conservation

* Corresponding author.

E-mail address: ktsainsbury1@gmail.com (K. Sainsbury).

NGOs have published their own guidance, such as The Nature Conservancy's "Five – S Framework for site conservation" (TNC, 2000). Moreover, support tools are being developed by academic groups, for example the Cambridge Conservation Forum Conservation Evaluation Tool (Kapos et al., 2008) and the Ranked Outcomes approach (Howe and Milner-Gulland, 2012). Common features of these frameworks and tools include a focus on "outcomes" (the change resulting from an intervention) as well as "inputs" (what resources were expended), "activity" (how were they expended) and "outputs" (what was delivered; Cambridge Conservation Forum Measures of Success Project). The variety of frameworks available presents practitioners with a new challenge – which of the available approaches will best suit their particular project's need to return reliable and informative results, cost-effectively, as part of their ongoing programmes?

Despite the policy-level commitment to evaluation and the development of various evaluation tools, conservation organisations, governments and development agencies worldwide are still implementing numerous local-scale interventions without strong evidence for whether, where, or under what conditions these approaches are effective. Furthermore, local-scale evaluations are still not standard practice, and some types of intervention are implemented with only blind faith that they are working. In particular, there is a lack of evaluation of the effectiveness of alternative livelihoods or alternative income-generating activities (IGAs) as a conservation strategy (Wicander et al., 2014).

The logic of IGAs, which are very widely implemented in the developing world, often by local NGOs with limited capacity (Wicander et al., 2014), is that providing small scale local activities that focus on certain types of income generation activity, such as tree planting and small animal husbandry, will give local people the resources they need and hence reduce their need to go into protected areas to harvest resources. The lack of evidence for the effectiveness of alternative livelihoods was noted as a concern at the 2012 IUCN World Conservation Congress, where a resolution was passed that called for evidence to be gathered urgently on these kinds of interventions. In response, an evidence-gathering exercise from existing literature has been launched (Roe et al., 2014). However, Wicander et al. (2014) warn that post hoc meta-analyses are unlikely to succeed, given the poor evidence base which currently exists.

The impacts of most conservation-focussed IGA interventions are hard to evaluate because of their complex nature, small scale and case-specific outcomes. Perceptions of project success, particularly in terms of the social components, are inevitably subjective and dependent on the perspective of the person being asked. Post-hoc evaluation is generally based on academic publications, project reports or questionnaires aimed at project managers (e.g. Brooks et al., 2012; Wicander and Coad, 2014; Roe et al., 2014). However, managers' perspectives on what constitutes success, and on whether projects have fulfilled their goals, may well differ from the perspectives of the people targeted by the projects. These issues call for flexible evaluation frameworks which are inclusive of a range of stakeholders, including both the staff of the implementing organisation and the target communities. When interventions are implemented in developing countries, and particularly by local NGOs, there is also a need for low-tech, relatively simple but robust approaches that can be implemented without high level statistical skills and which can incorporate both quantitative and qualitative assessment of project outcomes. Frameworks that can use retrospectively gathered materials, including project reports, are also more likely to be adopted.

When considering which evaluation approaches can be used in a particular situation, a key question is why that evaluation is needed. Evaluations can be used to build an evidence-base to guide future conservation interventions (e.g. Brooks et al., 2012; Roe et al., 2014). They can also be aimed at donors or internal priority-setters, in which case there may be a need to calculate a return on

investment (Murdoch et al., 2007), or the quantitative effect size of the impact of the intervention on some metric of poverty or biodiversity loss (e.g. Clements and Milner-Gulland, 2015). These two needs are best met by rigorous, externally-valid evaluations which may be costly in both time and technical expertise. Alternatively, an organisation may require an evaluation of project outcomes to date, in order to guide learning and adaptive management (Jenks et al., 2010). It may be more important that this type of evaluation is internally valid (i.e. rings true to those involved in the intervention) than that it generates externally-valid results, as this makes it more likely to highlight areas in which changes could be implemented to improve project performance in the future.

Here, we explore the potential of a recently published evaluative approach, the Ranked Outcomes (RO) method (Howe and Milner-Gulland, 2012). This novel approach was selected for its apparent, although as yet untested, ability to provide a structured framework for guiding the adaptive management of conservation interventions in a low capacity setting. The approach enables the post hoc evaluation of the outcomes of individual projects within a portfolio with over-arching objectives. It translates qualitative statements about hoped-for, or achieved, outcomes at the portfolio level into quantitative scores reflecting the success of individual projects within the portfolio towards meeting these objectives. It may be particularly valuable when objectives are poorly defined, or the assessor wishes to include outcomes which were unanticipated when the projects were initiated. It is also potentially helpful for outcomes which cannot easily be expressed in quantitative terms or are not easily comparable with a single metric. The method was developed for the evaluation of qualitative statements about diverse outcomes achieved by projects funded within the portfolio of the UK Government's Darwin Initiative, contained in final reports by project leaders; Howe and Milner-Gulland (2012) demonstrated that the approach compared well to two less flexible approaches (Threat Reduction Assessment, Salafsky and Margoluis (1999); and scoring of quantitative outputs).

We explore the potential of the RO method using a portfolio of projects funded by a Tanzanian conservation funding organisation, the Eastern Arc Mountains Conservation Endowment Fund (EAMCEF; www.easternarc.or.tz) in the Kilolo district of Iringa region in the Southern highlands of Tanzania, adjacent to the Uzungwa Scarp proposed Nature Reserve (USpNR). In order to address the critical need for conservation evaluations to hear the perspectives of the people targeted by IGA-type projects, we modified and extended the framework to gather the views of local villagers as well as those of project implementers. We then used the approach to carry out a preliminary evaluation of EAMCEF's interventions in four villages and make initial recommendations to EAMCEF. We end with an assessment of the general applicability of the method to project evaluation within conservation and recommendations for improvement of the method in future applications.

2. Methods

2.1. Study site

Uzungwa Scarp proposed Nature Reserve (USpNR) is a central government-managed forest reserve that is in the process of being upgraded to the status of Nature Reserve. It is located within the Eastern Arc Mountains and is one of the most important sites for biodiversity in that globally recognised centre of endemism (Burgess et al., 2007; Rovero et al., 2014). The reserve is surrounded by eight villages (Tanzanian national census data, 2012). Monitoring between 1998 and 2008 identified that biodiversity depletion in USpNR is higher than neighbouring forests, particularly in respect of its endemic, and in some cases endangered, primate and duiker populations (Rovero

et al., 2010). During the same period, evidence of increased forest disturbance such as snare hunting and logging (Rovero et al., 2010) and plant collection for medicinal purposes (Ndangalasi et al., 2007) was observed. Urgent recommendations arising from research at the time included upgrading USpNR status to Nature Reserve, improving law enforcement and initiating livelihoods programmes to provide alternative protein sources to the local communities (Rovero et al., 2010).

The EAMCEF was established in 2001 as an independent non-governmental organisation aiming to support conservation efforts in the Eastern Arc Mountains (www.easternarc.or.tz). Funding and initial support was provided by the GEF through the World Bank and United Nations Development Programme (UNDP), World Bank IDA funds, and the Tanzanian Government. Between 2006 and 2010, EAMCEF distributed approximately \$1 million in the form of grants for projects to a wide range of institutions, from government departments to private entities, to support new and existing initiatives in priority locations. More than 28 small-scale community-based projects have been funded, involving tree planting, livestock management, fisheries development, beekeeping and the introduction of fuel efficient stoves. To date EAMCEF has primarily focused on ensuring project delivery takes place as contracted with grantees, and is yet to review project outcomes.

2.2. The EAMCEF projects

We implemented the RO evaluation for EAMCEF projects carried out in four villages in Kilolo District in the Eastern Arc Mountains: Idegenda, Ilutila, Masisiwe and Mbawi. They all lie adjacent to the USpNR boundary and are within approximately a 15 km radius, and in some cases less than one hour's walk, of one another. Due to their high poverty levels and the high biodiversity value in nearby forests, the villages have been the recipients of numerous projects in the last 20 years, most notably a large-scale tree planting project in the 1990s and more recently a range of livelihood enhancement projects. The majority of EAMCEF conservation projects in this area have been run by Kilolo District Council.

Since 2006, 17 projects have been funded by EAMCEF in the study villages. Each project typically lasts 2 years. We evaluated IGA projects which were at least 90% complete by the time of our evaluation in June 2012. We grouped related projects into "programmes" to assist evaluation. Six programme groups were identified; beekeeping ("Bees"), dairy goat husbandry ("Goats"), fish farming ("Fish"), fuel efficient stoves ("Fuel"), rabbit farming ("Rabbits") and tree planting ("Trees"). Every village hosted five of the programme groups; *Bees*, *Fuel*, *Goats* and *Trees* appeared in all villages, *Fish* and *Rabbits* were implemented in two villages each. A summary of the programmes and their implementation is in Table 1.

We carried out a two-part evaluation of the six programmes carried out in the four villages using an adaptation of Ranked Outcomes framework based on the principles set out in Howe and Milner-Gulland (2012). Our adapted method is more general than the original approach, allowing us to evaluate the potential of the method in a range of circumstances and from a range of perspectives.

2.3. The ranked outcomes method

An idealised project design and implementation process might start with a theory of change, which then guides project activities, each of which is associated with a set of outcome measures against which project success can be evaluated. Increasingly, projects are expected to set and report against Specific, Measureable, Achievable, Realistic, Timebound (SMART) targets (e.g. Darwin Initiative, 2014). However, in the real world, interventions adapt their objectives to changing circumstances, unexpected outcomes occur, monitoring and evaluation are not built into project design, and if targets exist, they are often not SMART. In these circumstances, post hoc reconstruction of the outcomes

that an implementer would like to achieve, and their importance to their mission, is required.

The RO method which we developed for this study, based on Howe and Milner-Gulland's (2012) original approach, followed a five-step procedure:

- (1) *Identify* and agree with stakeholders a list of potential intervention outcomes. These could be based on statements within portfolio documentation or project reports, or on stakeholder consultation. The outcomes can be positive or negative, and may take the form of statements such as "Improved legal protection for priority conservation areas, e.g. gazetting new reserves, expanding existing ones or upgrades to legal status" (see [Supplementary Material](#) for a full list of outcomes identified for the EAMCEF case study). Care should be taken at this stage to keep the outcomes list broad and inclusive, so as to minimise the potential for selection bias.
- (2) *Sort* outcomes into groups, such that comparisons are made between similar types of outcome, enabling meaningful ranking to take place (e.g. in this case study, we grouped outcomes into Education & Awareness; Research & Infrastructure; Species & Habitat status, Legacy, Negative impacts). Groups may be decided adaptively after listing the outcomes, or may be pre-determined.
- (3) *Rank* outcomes within each group, according to the perceived relative importance to overall conservation success of achieving the outcome compared to the other outcomes (e.g. in the "Species & Habitat" group, "Infractions, e.g. illegal logging or bushmeat hunting, are reduced" may be ranked higher than "Creation of appropriate ex situ conservation strategies"). The ranking can be carried out separately by individuals and then a combined rank derived, or by a focus group, and can be carried out independently by different stakeholder groups or in consultation. If negative outcomes are included in the list, these can be ranked in order of detrimental effect.
- (4) *Score* each individual project according to whether it has or has not achieved each outcome; this can be done as a yes/no or a degree of achievement, and can be based on a single assessment (e.g. through individuals reading project reports) or based on opinions expressed in focus groups or surveys. The scoring is carried out separately from, and by different people to, those involved in selecting and prioritizing the outcomes, in order to maintain independence.
- (5) *Calculate* the RO score for each project by multiplying the rank of each outcome by its achievement, summing over outcomes for each outcome group and normalising (dividing the sum by the maximum achievable score). Outcome groups can then be combined, either directly or with a weighting, to get an overall RO score for a given project within the portfolio, which can be compared with other projects. If different stakeholders have ranked and scored outcomes separately, their RO scores can be compared.

The RO method therefore disassociates the assessment of outcome importance from the ranking of projects, and bases scores on stakeholder perspectives rather than predetermined weightings. The approach of asking people to agree and rank outcomes, then assess projects against these outcomes, also provides a framework for discussion and learning about the factors affecting project success.

2.4. Implementing ranked outcomes for EAMCEF

In order to investigate the different perspectives of programme success held by villagers compared to an external assessor working with

Table 1

Summary of programmes evaluated, including the number of projects per programme, number of participants and amount spent (\$000s).

	Bees	Fish	Fuel	Goats	Rabbits	Trees
Year started ^a	2007	2008	2008	2008	2006	2007
Study villages ^b	Idegenda Ilutila Masisiwe Mbawi	Idegenda Ilutila	Idegenda Ilutila Masisiwe Mbawi	Idegenda Ilutila Masisiwe Mbawi	Masisiwe Mbawi	Idegenda Ilutila Masisiwe Mbawi
Number of projects in this programme	3	2	1	1	1	3
Amount spent (\$000s) ^c	8.74	11.31	7.81	8.77	1.92	27.77
Amount spent per village (\$000s)	1.46	5.66	1.30	1.46	0.96	4.63

^a Typically the year after the proposal was submitted and accepted by EAMCEF.^b Although only four of the projects are included in the scope of this study.^c Some of the *Bees* and *Trees* projects were submitted under joint proposals. Costs have been split between the projects using the itemised budgets. Based on expenditure on project themes at point of evaluation, using exchange rate of: 1 TZS = USD0.000622665. Accessed on xe.com on 27th November 2013.

EAMCEF staff, we carried out two RO evaluations of the same IGA projects. The evaluations were tailored to the interests and capacity of each stakeholder group (i.e. villagers versus external assessor) and so were not directly comparable, but followed the same basic structure (Table 2). The aim was to make the results as comparable as possible while respecting the perspectives and capacities of different stakeholder groups. The outcomes for both the IE and VE were identified by the researcher using EAMCEF's strategy as the basis for selection, and validated with EAMCEF stakeholders.

The scoring step of the independent evaluation (IE) was carried out by KS, who qualitatively assessed the six programme groups for their performance against the ranked outcomes. The evaluation was independent in the sense that it was carried out by a researcher with no affiliation to the project or vested interest in a given result. The assessment was based on information gathered from a combination of reading project reports, direct onsite observations, and semi-structured interviews with village members participating in the project and with project leaders. Project reports were a mixture of progress updates made by the project manager, and summaries from EAMCEF head office's site visits made at regular intervals throughout the project. Direct onsite observations were made by KS to validate and supplement information provided in the written reports. This involved viewing the projects in situ to cross-reference their status with the information provided by the written reports and carrying out semi-structured interviews (see [Supplementary Information](#)). These aimed to probe areas of ambiguity in outcome achievement identified during examination of the reports. The final project scores were based on the combined findings from all three activities across all four villages; where there were discrepancies in findings or unsubstantiated statements in reports, information collected by direct observation and in interviews took precedence.

Using the *Beekeeping* project as an example, an outcome clearly awarded is as follows:

- Outcome: "Infrastructure: The capital resources required for the project (e.g. seedlings, hardware) are provided".
- Project report: outlines that 140 beehives have been distributed to the four villages.
- Interviews: verify that modern beehives were provided to the villages.
- Observations: show the location of the beehives and their current state.

An example of an outcome not awarded for *Rabbits* is:

- Outcome: "Infrastructure: Livelihoods are established".
- Project report: states that the rabbits are providing additional income.
- Interview with Project Coordinator: states that there are 500 rabbits in the village.
- Observations and participant interviews: locate just 12 rabbits remaining from the project across the two villages.

The 'Villager Evaluation' (VE) aimed to evaluate any disconnect between the perceptions of the project implementers and project recipients, by repeating the evaluation process with village members. In this evaluation, three focus groups ranked a simplified list of 25 outcomes, chosen to reflect specific outcomes that were of relevance to local people. Focus groups aimed to reflect the perspective of a broad group of villagers. Participants were volunteers who came forward after the process was announced at a village meeting. This is likely to have resulted in somewhat biased viewpoints, but was deemed necessary according to cultural norms. Scoring was then carried out using interviews with 132 individuals within the four villages. Interviewees were asked firstly

Table 2Summary of the application of the RO method for EAMCEF. H&MG = Howe and Milner-Gulland (2012). See [Supplementary Information](#) for full list of grouped and ranked outcomes.

	Independent Evaluation (IE)	Villager Evaluation (VE)
Step 1 – Identification	60 project outcomes proposed by the researcher and agreed with EAMCEF staff. Outcomes based on original lists in H&M-G, adjusted using EAMCEF strategy documentation created by a wide range of stakeholders when EAMCEF was established	25 outcomes most relevant to villagers chosen from original 60. EAMCEF's lowest-priority outcomes and those unlikely to relate to the villagers' experience were removed
Step 2 - Sorting	Outcomes grouped into 10 categories based on H&M-G, including a separate "negatives" category	Outcomes grouped into 5 of the 10 original categories (e.g. 'Research and Planning' was removed)
Step 3 – Ranking	EAMCEF staff ($n = 3$) completed individual prioritisation exercises. Results compared using Fleiss' (1981) Kappa Statistic and Landis & Koch's (1977) guide on interpreting the kappa statistic in terms of strength of agreement between participants	3 focus groups (FGs) carried out per village to rank outcomes. Rankings from the FGs were averaged to create an overall rank per outcome
Step 4 – Scoring	Independent evaluator reviewed project documentation ($n = 55$), interviewed implementing staff and local people ($n = 52$) & directly observed the project in all 4 villages. A binary yes/no scoring system used to indicate where outcomes were met	Surveys were carried out with 132 participants across the study villages. Outcomes were presented as statements for participants to agree or disagree with. The proportion of "yes" scores was used as the score for each outcome
Step 5 – Calculation	The rank position of outcomes with a "yes" were added together for each programme to create category scores and overall programme results	The rank position of outcomes were multiplied by the proportion of "yes" results and then totalled for each programme

whether they were aware of each project which had been implemented in their village, and if so, whether or not each outcome had been met for each of the projects of which they were aware. They were also asked whether they had personally participated in the projects. Survey participants were selected by stratifying villages by subdivision. House-to-house walk-arounds were carried out in each subdivision, with every second person approached to be interviewed until the requisite sample size had been reached. Focus group participants were excluded from the surveys.

3. Results

3.1. Outcome ranking

3.1.1. Independent evaluation

The EAMCEF Secretariat staff had a “Fair” agreement on the priority they gave to different outcomes of their investments (Kappa = 0.221; Landis and Koch, 1977). None of the outcome groups showed particularly strong agreement, ranging from 0.01 (poor) for the Negatives category to 0.42 (moderate) for Species & Habitat. It appeared therefore that EAMCEF staff had somewhat differing views on the most important outcomes of their funding. Feedback from the respondents was that it was in some cases “difficult to prioritise as all of the outcomes are important” but that it was a “useful exercise to think again about what we are trying to achieve and why we are here”. There was also recognition that many of the projects are small scale and not likely to meet all of the outcomes on their own as they form part of a broader strategy. The median of the priority scores from the EAMCEF staff interviews was used as the outcome ranking for the next stage of the evaluation.

3.1.2. Villager evaluation

The three focus groups (FGs) gave quite different rankings of the outcomes, with an overall Kappa statistic of zero (poor agreement), and the largest value being a slight negative agreement for Species & Habitat (suggesting opposing categorisations between FGs). This was in spite of each group articulating logical, if diverse, rationales for their decisions. For example, one of the few points of agreement was within Education, with all the groups agreeing that “projects providing more environmental education in schools and clubs” was the most important outcome. The group said this was because “to conserve the environment, you need the young people to take part”. In several cases, the outcomes that one might expect to be highest priority, for example “the project leads to improvements in the number of naturally occurring plants and animals in our environment” (Species & Habitat) were considered lowest priority overall. When asked, this was explained by some groups as being lower priority as “this [outcome] can’t happen without other changes happening first”.

This low level of agreement between FGs is perhaps less surprising than it at first appears, as one might expect that the villagers are less likely to be united by shared conservation goals in the same way as people who work for a conservation organisation. This lack of agreement presents a challenge for the method, because if participants do not agree on what is important, the validity of using the median ranking to represent such diverse views is questionable. For the purpose of this study, which was primarily a methodological exercise, the decision was made to continue with the median.

3.2. Outcome scoring

3.2.1. Independent evaluation

At the end of the project documentation review, 52% of outcomes required further clarification through interviews and field observations. Only those outcomes which were directly relevant to the specific projects being assessed were taken forward to the evaluation (82%). On a number of occasions the project documentation differed from the

information collected during the observations and interviews (see for example the *Rabbits* outcome described in the Methods section).

Based on the evaluation of the outcomes achieved in each category, weighted by each outcome’s rank and summed over all categories, the IE suggested that *Trees* was the best initiative (Table 3). This was primarily due to the outcomes achieved in Education & Awareness; the study found that the greatest impact of *Trees* was improved understanding by the community of the ecological benefits associated with tree planting (one respondent volunteered “everyone, even the smallest child in this village, understands the value of planting trees now”), and minimal deductions in the Negatives grouping. *Fish* was second best, with the highest marks in the Legacy category and second lowest deductions in the Negatives category. The worst performing initiative was *Goats*. This was due to its low scores in Education & Awareness and Research & Planning, as a result of poor training and the provision of native rather than dairy goats, along with high deductions in the Negatives category. Overall, projects split into two rough groups, with a top four that attained more than 50% and a bottom three that scored less than 30% of the theoretical maximum number of points.

3.2.2. Villager evaluation

In the questionnaire survey of villagers, respondents were on average aware of 3.9/5 projects that had run in their village, and only one respondent had not heard of any of the projects. This was higher than might be expected given that some of the projects only involved 10 people per village. 70% of respondents had heard of EAMCEF. *Bees* had the highest and *Trees* the lowest awareness of the projects delivered in all four villages. The overall participation rate in any project was relatively high given the small scale of some of the projects, at 29% of respondents. This could be due to the sampling, which was unavoidably limited by availability of participants (many villagers were at their farms or working away from home). The *Trees* project had the highest participation rate at 20% of respondents, and the *Bees* project had the highest awareness, at 92% of respondents (see Supplementary Materials for more detail).

According to the villagers, the highest scoring project was *Trees*, with a score of 45.1 out of a possible maximum of 60 (Table 4). It achieved an average “yes” response of 81% across all of the positive outcomes. The second best project, *Fish* (22.0), scored less than half the number of points compared to *Trees*. The worst performing project was *Fuel*, with an overall score of 6.9. This was due to the poor adoption in most of the villages (except Masisiwe), and so the project was viewed negatively throughout.

There was substantial variation between villages in their perceptions of project outcomes, for some of the projects. *Bees*, *Goats* and *Fuel* varied substantially between villages in their perceived success, while *Trees* was viewed consistently positively by all villages. *Fish* and *Rabbits*

Table 3

Independent Evaluation results, ordering projects by best to worst achievement. Each category can score a maximum of 55 points, making a theoretical maximum score of 275 (with Negatives scoring minus points). Numbers in brackets show the number of outcomes achieved per category. See Supplementary Information for the full list of outcomes and ranks.

Outcome	Trees	Fish	Bees	Fuel	Rabbits	Goats
Education & Awareness	40 (7)	27 (4)	34 (5)	27 (4)	18 (3)	7 (1)
Research & Planning	13 (2)	23 (3)	10 (2)	24 (3)	13 (2)	6 (1)
Infrastructure	54 (9)	54 (9)	46 (7)	28 (5)	42 (6)	42 (6)
Species & Habitat	27 (4)	18 (3)	27 (4)	15 (2)	10 (2)	18 (2)
Legacy	34 (4)	42 (6)	37 (5)	22 (3)	0 (0)	21 (3)
<i>Subtotal positive outcomes</i>	168	164	154	116	83	94
Negatives	−4 (1)	−9 (2)	−18 (3)	−39 (6)	−20 (4)	−34 (6)
Overall	164	155	136	77	63	60
Score% of maximum	60%	56%	50%	28%	23%	22%
IE ranking	1	2	3	4	5	6

The italics denote a subtotal and the bold highlights the ranking position.

Table 4

Villager Evaluation project scores subdivided by outcome category. Project scores are calculated by multiplying the outcome rank by the proportion of respondent agreement per outcome and totalling the results. Maximum total score is 60. Maximum score per outcome category is 15. See [Supplementary Information](#) for the full list of outcomes and ranks.

	Bees	Goat	Fish	Fuel	Rabbit	Trees
Legacy	7.1	6.0	7.5	2.6	5.2	11.6
Species & Habitat	7.2	6.0	7.0	3.9	5.8	12.6
Infrastructure	7.8	6.6	8.7	3.6	6.2	12.0
Education	5.3	4.6	5.1	3.3	4.0	12.0
<i>Subtotal positives outcome</i>	<i>27.4</i>	<i>23.2</i>	<i>28.3</i>	<i>13.4</i>	<i>21.2</i>	<i>48.2</i>
Negative	-6.2	-6.3	-6.3	-6.5	-5.5	-3.1
Overall total	21.2	17.0	22.0	6.9	15.6	45.1
VE ranking	3	4	2	6	5	1

The italics denote a subtotal and the bold highlights the ranking position.

were implemented in only two villages and so comparisons were less robust. Masisiwe was consistently the most positive village about all the projects, while Ilutila perceived the projects worst, even giving two projects (*Goats & Fuel*) negative overall scores (Fig. 1).

Participants in the projects were more positive about the projects' outcomes than non-participants. The level of additional positivity among participants varied by project and village, reflecting the level of success with project implementation in the different villages. Masisiwe participants were most positive overall about the projects, Ilutila participants were least positive. Fuel was the only project to be scored more negatively by the project participants than non-participants and this happened in two villages: Mbawi and Ilutila.

4. Discussion

4.1. General assessment of the RO approach

Despite the awareness within academic and policy circles of the critical importance of evaluating alternative livelihood initiatives, gathering evidence on the impact of these conservation interventions continues to pose a challenge. Our aim was to extend the RO method and trial it on a set of income-generating activity (IGA) projects under the conditions

encountered by conservationists in the messy real world, where there is limited time and money for evaluations, little or no quantitative information, poor documentation and poor initial – or shifting – articulation of priorities and programme outcomes. We found the method stood up to this test, producing relative project scores that enabled us to rank the projects in order of achievement with some consistency in results between the Independent and Villager Evaluations; highlighting gaps in knowledge (for example a lack of data pertaining to biodiversity); and identifying a set of lessons learned that will be beneficial for EAMCEF, for future IGA projects in general, and for future application of the RO method (see [supplementary materials Table S6](#) for our recommendations for future application of RO to IGA projects). Our experience suggests that RO could provide a useful framework for evaluating IGAs, contributing to improving understanding of the role and effectiveness of IGA approaches in conservation. In order to consider the merits of the RO approach, we look critically at each step in the evaluation process in turn.

4.1.1. Outcome listing

Progress cannot be measured against ambiguous or unmeasurable objectives; hitherto this has been a common criticism of conservation evaluations (Clarke, 1996). Investing time in agreeing and appropriately wording a project's outcomes is fundamental to a meaningful evaluation. In the event that objectives are unclear or evaluation has not been planned as part of the original project definition, RO provides a practical solution for a lack of clarity by (re)agreeing strategic priorities at the outset as part of a transparent and inclusive process.

The selection and articulation of appropriate outcomes is fundamental to the success of the method. As highlighted in the management science literature, care should be taken to ensure that the selected outcomes for RO evaluation are "mutually exclusive, but collectively exhaustive" (Rasiel, 1999). This means ensuring that the outcomes selected are broad enough in scope to cover all priorities (and no more), but not so numerous that they are indistinct from one another from a benefits accounting perspective. Too many outcomes can complicate the results and dilute the lessons learned. If the outcomes are too similar, the prioritisation is less meaningful because participants struggle to differentiate between them, and evaluators risk double-counting benefits

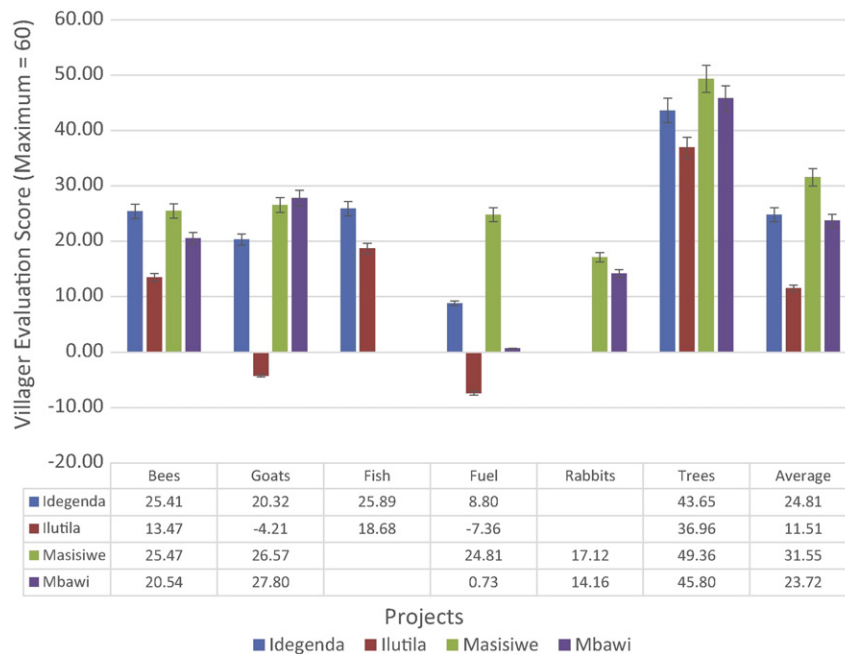


Fig. 1. Villager Evaluation project scores by average project group results in each village. Project scores are calculated by multiplying the outcome prioritisation by the proportion of respondent agreement per outcome and totalling the results. Maximum project score is 60. Results for Bees (KW = 11.23, df = 3, p = 0.011), Goats (KW = 50.64, df = 3, p < 0.0001) and Fuel (KW = 46.62, df = 3, p < 0.0001) were significantly different by village. Trees has less differentiation, which is borderline significant (KW = 7.21, df = 3, p = 0.066).

under multiple outcomes. Consideration should be given to whether it is desirable to measure progress at an organisational or project level. If the former, the outcomes should be used as a target for all key organisational priorities; the benefit of this approach being that the organisation can identify at a strategic level whether it is progressing towards its outcomes. If the latter, care should be taken to remain focussed on the most important goals of the project, particularly if designing outcomes after the start of a project. Including outcome measures which are of secondary importance may divert attention from the primary project outcomes. There are circumstances where it may be right to introduce new objectives to a project, but in general terms it is unfair to expect a project to be measured against new or secondary objectives where there may be little or no data to support the evaluation.

Although the potential for RO to be applied after a project has been implemented is a clear advantage of the approach, there is an obvious risk of choosing outcomes that are known to be achievable, thereby positively biasing the results. It is also important to ensure an appropriate balance of outcomes, particularly with livelihoods projects where outcomes may conflict with, or not necessarily clearly relate to, biodiversity conservation goals. With the aim of minimising bias in the outcome selection, the outcomes in this case study were drawn from EAMCEF's founding strategic objectives, rather than being newly created for the purpose of the evaluation. The challenges with this approach were (a) the high number of outcomes ($n = 60$), which made the evaluation results complex and limited the transferability of the outcome list to the villager evaluation; (b) the breadth of outcomes, which not all the projects could be expected to meet and which led to overall low scoring by the projects (although arguably if the original aim was to assess progress against EAMCEF's strategic objectives, the evaluation worked to highlight gaps in their focus); and (c) the differences in the time period for realising the strategic objectives versus the lifespans and scale of the projects, which again pre-disposed the projects to perform poorly against the measures chosen.

4.1.2. Outcome grouping

The grouping of outcomes was intended to ensure comparisons were made like-with-like, and to reduce the burden of ranking a large number of outcomes where judgements were likely to be difficult to make. It was relatively effective for the Independent Evaluation (IE), although the Negatives category may have better been amalgamated within the relevant categories rather than acting as a grab-bag for a range of different potential unintended consequences.

However, villagers tended to answer consistently in all their responses for a given project, rather than differentiating between a project's achievement of particular outcomes or outcome groups. It may be that there were too many outcomes for the villagers to distinguish clearly between and so their responses were based on an overall perception of the projects. Better results may potentially have been obtained by limiting villager evaluations to a much shorter list of outcomes than 25, ungrouped. This may enable villagers better to rank and score projects against a more focussed list.

4.1.3. Outcome ranking

The purpose of the outcome ranking process was to provide a weighting based on the relative importance of the outcomes as judged by the participants in the process. In this case study, EAMCEF staff were selected to prioritise the IE and the villagers the Villager Evaluation (VE). The most interesting thing about this step was the lack of agreement about the rankings, both by staff and villagers. The lack of agreement between FGs in the VE may be unsurprising due to the potential range of opinions represented across the village. In the IE, it may be less expected, given that individuals within an organisation might be expected to share a common view of what was most important in their conservation programme. In future, a Delphi method may be worthy of investigation, in which individuals rank independently,

the ranks are revealed to the group, followed by a period of discussion and reflection followed by a re-ranking. This has proven successful in improving the accuracy and information content of expert judgements in a range of contexts (Martin et al., 2012).

4.1.4. Outcome scoring

As is likely to be the case in many situations, the written reports available at EAMCEF for the IE were short, focussed on outputs, and descriptive rather than based on hard data. In our case study, we supplemented these documents with information gathered from conversations with on-the-ground implementers and project participants. Without this additional verification of the materials contained in the reports, the IE would have remained substantially uninformed and potentially compromised. However, focussing the verification on the relatively straightforward question of whether particular outcomes had or had not been fulfilled meant that information-gathering could remain limited and focussed. Methods which require nuanced assessments (rather than yes/no), interpretation and multiple proof points may create a reliance on individual evaluator judgement that could lead to significant variation in the results between different independent evaluators. On the other hand, a simple binary assessment such as in the RO method also leaves room for bias in interpretation, either through uninformed scoring due to absence of evidence, or overreliance on snap judgements. It also meant that an outcome that had only just been achieved received the same value as one that had been overachieved. The RO approach does not require rankers to consider the counterfactual case and does not attempt to assign causative mechanisms to outcome fulfilment; despite this, the discussions promoted by the method gave insights which could be used to inform future, more mechanistically-based evaluations once data became available to support them.

One approach for future testing could be to introduce a scale of achievement per outcome with a clearly defined scorecard of what it means to achieve each level. This is similar to Goal Attainment Scaling, a tool that originated in clinical fields (Marson et al., 2009). In Goal Attainment Scaling, outcomes are measured against a five-point scale, for example, -2 to $+2$ where -2 is a deterioration in status, 0 is neutral and $+2$ shows a considerable improvement. This would remove the need for outcomes that describe negative or progressive states and allow progression to be demonstrated cleanly through repeat evaluations, which lends itself to adaptive management.

4.1.5. Overall evaluation of the RO method

Any evaluation depends on the perspective of the evaluator (Scriven, 2011). One strength of the RO method is the separation of the outcome ranking from the assessment of outcome achievement, both because they are explicitly separate stages of the process, and because different people can contribute at each stage. The method also enables an outsider to evaluate progress independently, and enables a range of perspectives to be taken into account using a relatively comparable framework. Incorporating Return On Investment (ROI) measures is viewed by some as the logical next step in the evolution of conservation planning (Murdoch et al., 2007). One of the benefits of RO is that it could provide a numerator for an ROI or cost-benefit analysis, with the project budget as the denominator.

4.2. Evaluation of the EAMCEF programme

On one hand, it could be considered that the six programme groups scored relatively poorly in the evaluation, with the highest score being 60% of the theoretical maximum (Table 3). On the other hand, the outcomes were selected based on EAMCEF's overall strategy and the small size of the programmes means they were unlikely to be broad enough in scope to meet all the outcomes. In addition, the programmes are still less than five years old and it may take some time to accrue benefits. For example, *Trees* will not be ready to harvest for another decade,

while *Goats* required the livestock to mature before breeding. Anecdotes at interview suggested that a subset of participants were enjoying socio-economic benefits resulting from the projects; typically these were proactive early adopters, who others were slowly beginning to copy.

Bearing in mind that the focus of the study was trialling the RO approach rather than carrying out a full assessment of EAMCEF's IGA projects, useful lessons were still learned. One lesson that came out was the difficulty in attributing outcome achievement to a particular programme when a plethora of initiatives was being carried out in a single area. This particularly affected the VE, where villagers were not clearly able to distinguish between the long-term benefits of different programmes. For example, both *Trees* and *Fuel* sought to reduce dependence on fuelwood collection. It could be argued that reduced visits to USpNR for this purpose could be due to *Trees* providing plenty of offcuts close to the village, or *Fuel* reducing the frequency of collection. In three of the four villages, *Fuel* was perceived to have performed poorly, so any benefit may be more likely to be due to *Trees*. However, this was not the first tree planting project in the area. In the 1990s, a similar, larger tree planting project was implemented. The education benefits and perceived success of the *Trees* projects could have been due to it being a repetition of an initiative that was already well known and understood by the communities.

Any biodiversity conservation outcomes that occurred at the portfolio level could not be evaluated due to a lack of monitoring of threats such as hunting or deforestation. An understanding of the relationship between IGAs and biodiversity would require further research, and baseline monitoring to have taken place. This highlights a challenge of post hoc evaluation – any analysis can only be as good as the available data. Even if monitoring had taken place, it seems unlikely that any change in threats could be robustly attributed to the activities of EAMCEF, due to the small scale of the programme and its limited scope. There was no conditionality built into any of the IGA projects, so the programme did not provide an incentive for villagers to reduce any environmentally-damaging activities.

That the IE evaluation and VE judged the overall performance of the projects similarly suggests that villager perceptions are broadly in line with the independent evaluation. However one key message from the RO evaluation was the importance of recognising and addressing heterogeneity of perceptions; the RO framework lends itself to capturing this variety. Different villages had very different perceptions of the success of projects; individual projects varied in their performance between villages, and individual villages varied in their overall perception of portfolio success. Similarly there was a lack of agreement between EAMCEF staff and between FGs about the ranking of outcomes. These results suggest that perceptions are important, and that this heterogeneity should be sought out and addressed in adaptive management. This will enable managers to target improvements in programmes to those groups who are not seeing the benefits. Knowing the discrepancy between programme staff and villagers' views of project outcomes (both in terms of their importance and their achievement) can promote dialogue about the causes and consequences of these differences.

5. Conclusion

In the future, donor pressure and changing norms about best practice in conservation may mean that conservation implementers start to design projects with evaluation in mind, and have a clear understanding of their project's outcomes and a strong set of appropriate quantitative and qualitative indicators from the start. In the meantime, there is an urgent need for feasible, easy-to-implement methods that still provide a robust evaluation and can feed into adaptive management. This is particularly important for conservation interventions such as IGAs, where relationships between the intervention and outcomes can be complex and understanding of factors affecting success continues to be limited. The RO method fulfils this need, enabling qualitative

statements in documents and stakeholder interviews to be used to produce a quantitative score, weighted according to perceived priorities rather than arbitrarily developed weights. Even in a challenging environment for evaluation, RO can provide a framework for a rapid impact assessment which enables a range of perspectives to be included, and which can inform the design of future more in-depth evaluations.

Acknowledgements

This study was supported by the Tropical Agriculture Association Fund. We thank the EAMCEF and its staff, Kilolo District Council and the people of Idegenda, Ilutila, Masiwiwe and Mbawi for their support and assistance in completing fieldwork. This paper is an output of Imperial College's Grand Challenges in Ecosystems and the Environment initiative. We acknowledge the support to EJM/G of ESRC/DFID project RS/J018155/1.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.biocon.2015.06.001>.

References

- Brooks, J.S., Waylen, K.A., Mulder Borgerhoff, M., 2012. How national context, project design, and local community characteristics influence success in community-based conservation projects. In: Proceedings from the National Academy of Sciences of the United States of America, vol. 109 no. 52.
- Burgess, N.D., Butynski, T.M., Cordeiro, N.J., Daggart, N., Fjelds , J., Howell, K., Kilahama, F., Loader, S.P., Lovett, J.C., Mbilinyi, B., Menegon, M., Moyer, D., Nashanda, E., Perkin, A., Stanley, W., Stuart, S., 2007. The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biol. Conserv.* 134, 209–231.
- Cambridge Conservation Forum Measures of Success Project (2013–14). <<http://www.cambridgeconservationforum.org.uk/initiative/harmonising-measures-conservation-success>>.
- Clarke, T., 1996. Learning as a strategy for improving endangered species conservation. *Endangered Species Update*, vol 13: 1&2. Online source: <<http://hdl.handle.net/2027.42/39326>>.
- Clements, T., Milner-Gulland, E.J., 2015. The impacts of payments for environmental services and protected areas on local livelihoods and forest conservation in northern Cambodia. *Conserv. Biol.* 29, 78–87.
- Darwin Initiative, 2014. Round 21 of Funding – How to apply for funding: guidance note for applicants to the Darwin Initiative. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/317932/R21_Main_Projects_Guidance.pdf>.
- Ferraro, P.J., Pattanayak, S.K., 2006. Money for Nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biol.* 4 (4), e105.
- GEF, 2010. The GEF Monitoring and Evaluation Policy. Washington: Global Environment Facility.
- Hockings, M., Stolton, S., Leverington, F., Dudley, N., Courrau, J., 2006. Evaluating effectiveness: a framework for assessing management effectiveness of protected areas. 2nd ed. IUCN, Gland, Switzerland and Cambridge, UK.
- Howe, C., Milner-Gulland, E., 2012. Evaluating indices of conservation success: a comparative analysis of outcome- and output-based methodologies. *Anim. Conserv.* 15, 217–226.
- James, A., Gaston, K., Balmford, A., 1999. Balancing the Earth's accounts. *Nature* 401, 323–324.
- Jenks, B., Vaughan, P.W., Butler, P.J., 2010. The evolution of rare pride: using evaluation to drive adaptive management in a biodiversity conservation organization. *Eval. Prog. Plann.* 33 (2), 186–190.
- Kapos, V., Balmford, A., Aveling, R., Bubb, P., Carey, P., Entwistle, A., et al., 2008. Calibrating conservation: new tools for measuring success. *Conserv. Lett.* 1, 155–164.
- Landis, J.R., Koch, G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Marson, S., Wei, G., Wasserman, D., 2009. A reliability analysis of goal attainment scaling (GAS) weights. *Am. J. Eval.* 30, 203.
- Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M., Mengersen, K., 2012. Eliciting expert knowledge in conservation science. *Conserv. Biol.* 26, 29–38.
- Mascia, M.B., Pailler, S., Thieme, M.L., Rowe, A., Bottrill, M.C., Danielsen, F., Geldmann, J., Naidoo, R., Pullin, A.S., Burgess, N.D., 2014. Commonalities and complementarities among approaches to conservation monitoring and evaluation. *Biol. Conserv.* 169, 258–267.
- Murdoch, W., Polasky, S., Wilson, K., Possingham, H., Kareiva, P., Shaw, R., 2007. Maximizing return on investment in conservation. *Biol. Conserv.* 139, 375–388.
- Ndangalasi, H.J., Bitariho, R., Dovie, D.B., 2007. Harvesting of non-timber forest products and implications for conservation in two montane forests of East Africa. *Biol. Conserv.* 134, 242–250.
- Rasiel, E., 1999. *The McKinsey Way*. McGraw-Hill.

- Roe, D., Day, M., Booker, F., Zhou, W., Allebone-Webb, S., Kümpel, N., Hill, N.A.O., Wright, J., Rust, N., Sunderland, T.C.H., Redford, K., Petrokofsky, G., 2014. Are alternative livelihood projects effective at reducing local threats to specified elements of biodiversity and/or improving or maintaining the conservation status of those elements? A systematic review protocol. *Environ. Evidence* <http://www.environmentalevidencejournal.org/content/3/1/6>.
- Rovero, F., Mtui, A., Kitegile, A., Nielsen, M., Jones, T., 2010. Uzungwa Scarp Forest Reserve in Crisis. An urgent call to protect one of Tanzania's most important forests. Dar es Salaam.
- Rovero, F., Menegon, M., Fjeldsa, J., Collett, L., Doggart, N., Leonard, C., Norton, G., Owen, N., Perkin, A., Spitale, D., Ahrends, A., Burgess, N.D., 2014. Targeted vertebrate surveys enhance the faunal importance and improve explanatory models within the Eastern Arc Mountains of Kenya and Tanzania. *Divers. Distrib.* <http://dx.doi.org/10.1111/ddi.12246>.
- Salafsky, N., Margoluis, R., 1999. Threat reduction assessment: a practical and cost-effective approach to evaluating conservation and development projects. *Conserv. Biol.* 13 (4), 830–841.
- Scriven, M., 2011. The Evaluation Advisor: A New Role for Evaluators? <http://michaelscriven.info/images/EVALUATION_ADVISOR.6.3.11.pdf>.
- Stem, C., Margoluis, R., Salafsky, N., et al., 2005. Monitoring and evaluation in conservation: a review of trends and approaches. *Conserv. Biol.* 19, 295–309.
- Sutherland, M., Pullin, A., Dolman, P., Knight, T., 2004. The need for evidence-based conservation. *Trends Ecol. Evol.* 19, 305–308.
- The Nature Conservancy, 2000. The Five-S Framework for Site Conservation. Arlington, VA, USA: The Nature Conservancy.
- Wicander, S., Coad, L., 2014. Learning our Lessons: A Review of Alternative Livelihood Projects in Central Africa. Technical Report. University of Oxford, UK and IUCN CARPE, Cameroon.
- Wicander, S., Coad, L., Angu Angu, K., Walters, G., 2014. Moving Beyond Assumptions: Can alternative livelihoods projects reduce hunting pressure? Technical brief. <<http://www.eci.ox.ac.uk/publications/downloads/coad14-brief-long.pdf>>.