

Going Beyond Mixed Methods to Mixed Approaches: A Systems Perspective for Asking the Right Questions

Jeneen R. Garcia and Aaron Zazueta

Abstract An impact evaluation's primary task is to determine which impacts were caused by an intervention, distinguishing them from those produced by other causes. However, in complex systems, interventions may contribute towards less apparent forms of impact (such as negative, unintended, indirect and secondary) that are no less significant, but which require a different way of asking questions. Rather than advocating for a particular evaluation approach or methodology, we propose that (1) beyond the intervention's theory of change (TOC), the complex system within which the intervention has taken place and intends to make an impact must also be examined, and (2) specific elements of an intervention's TOC and the larger system must be understood to have a systematic basis for determining what evaluation questions need to be asked. Only then can the appropriate evaluation approaches and methods be identified, and adapted as this understanding increases over the course of the evaluation.

1 Introduction

Impact, as defined by the OECD-DAC (OECD 2002), can come in many forms. It may be positive and negative, intended and unintended, direct and indirect, primary and secondary – effects that may or may not be attributable to an intervention, depending on what the evidence points to. An impact evaluation's primary task is to determine which impacts were caused by the intervention, distinguishing them from those produced by other causes. However, the evidence collected to determine this is itself constrained by the evaluation questions asked, i.e. what type of evidence to look for, when, where, and how, with an increasing trend to anchor these on an intervention's theory of change (TOC) (Vogel 2012; Stein and Valters 2012).

When evaluation questions are focused solely on the intervention logic – what the intervention intended to achieve, whether or not this was achieved, and whether this can be categorically attributed to the intervention or not – there is a danger of overlooking evidence that points to unintended, indirect and secondary impacts

associated with the intervention. The very nature of these types of impact suggests that they may not be categorically attributable to the intervention, but instead are largely the effects of the context interacting with the intervention. Interventions in complex systems, in particular, are prone to manifesting these less obvious types of impact. A complex system is characterised by the interconnectedness and interdependence of its elements and dimensions, feedback loops shaping how change occurs, and often unpredictable behaviour emerging from the interaction of its parts (Ramalingam and Jones 2008). Another way of defining a system's complexity is the degree of uncertainty and difficulty in achieving agreement on how problems are solved (Patton 2011). As such, any intervention involving human institutions or environmental interaction is likely to be complex, and have non-linear chains of causality caused by tipping points, spatial and temporal mismatches, and other surprises that do not follow the expected pathways to impact. While some schools of thought focus only on approaches that measure direct, attributable

impacts (e.g. randomised and quasi-experimental evaluations), knowledge of complex systems processes tells us that an intervention may also play a role in achieving impact in other ways.

The inability to establish direct causation and net results therefore does not necessarily diminish an intervention's transformative effects; instead, by investigating the ways in which interventions interact with the complex systems of which they are part, more attention to these less obvious types of impact may be a means to assess the role of the intervention as an insufficient but necessary ingredient in achieving the longer-term, large-scale impacts that a development organisation desires.

2 Asking the right questions

In this article, we offer a conceptual framework for developing the appropriate evaluation questions, taking into consideration the multiple factors and processes beyond an intervention's intended chain of causality. Rather than advocating for a particular evaluation approach or methodology, we propose that any inquiry begin with an in-depth understanding of what the evaluation needs to examine. Only then can appropriate evaluation approaches and methods of data collection and analysis be identified for each of the evaluation's components or phases. This conceptual framework proposes that (1) beyond the intervention's TOC, the complex system within which the intervention has taken place and intends to make an impact must also be examined, and (2) specific elements of an intervention's TOC and the larger system must be understood to have a sufficient basis for determining what evaluation questions need to be asked.

The understanding of these elements within the context of both the intervention's TOC and the complex system beyond it is expected to be iterative as more information is collected over the course of an evaluation. The result is that an evaluation's scope, units of analysis and methodologies may be adapted as needed to take into account new questions that arise, as well as to systematically identify those that may no longer be feasible or relevant. This allows the evaluation to integrate evolving knowledge in a timely manner, in order to remain consistent with the evaluation's agreed-upon purpose and objectives.

2.1 The theory of change and beyond

As with most theory-based evaluations, we begin by getting a good grasp of how the intervention intended (or intends, if *ex ante*) to achieve impact. 'Intervention' may be a strategy, programme, project, project component, or specific project activity. It is important to define at which of these scales the evaluation needs to be done, or if it is to be an evaluation of a combination of different types and scales of intervention. An intervention's TOC identifies its inputs and activities, expected outputs and impact, and key assumptions about how impact will be achieved and might not be achieved. Many texts discuss how to construct and analyse an intervention's TOC for use in an evaluation (e.g. see Mackenzie and Blamey 2005; Morra Imas and Rist 2009; Van den Berg and Todd 2011).

However, to uncover the less obvious types of impact that may have occurred, we need to look beyond the intervention to the larger complex system that it sought to influence, and which has influenced it. Many theory-based evaluation approaches (e.g. realist evaluation, results-based evaluation, contribution analysis) advocate the importance of understanding an intervention's context, such as political and social factors, either to distinguish an intervention's impacts from those of contextual variables (White 2009; Mayne 2011), or to identify which specific context triggers mechanisms that allow impact to be achieved (Pawson and Tilley 2004; Blamey and Mackenzie 2007). 'Context' refers to the background or setting against which a certain result can be fully understood; a 'system', on the other hand, is 'a set of interacting or interdependent components forming an integrated whole'.¹ The difference between the two, while subtle, is critical to how evaluative evidence is sought out and interpreted.

A complex system perspective emphasises the interconnectedness between an intervention and its context. Thus, while a 'contextual perspective' tries to assess the effects of contextual variables on the intervention, or to separate their effects from the intervention's, a complex system perspective assumes that both the intervention and its context are constantly evolving, and that the evolution of one is shaped by the other's. 'Context' is not treated as a static backdrop to the intervention, but rather as comprising many independent yet interconnected components,

each one also shaping the others. The intervention is thus no longer viewed as the main agent of change, but rather as one of these many components interacting to produce an often uncertain outcome. By framing what has traditionally been called ‘context’ in terms of the system that an intervention is part of, the evaluation has a clearer basis for identifying which contextual variables must be examined, and which are likely to be irrelevant. This applies both to discounting rival hypotheses on what caused the change (e.g. contribution analysis, quasi-experimental approaches), and for understanding how the intervention and other parts of the system interacted to produce the observed result (e.g. realist evaluation).

2.2 Elements that need to be understood

To determine the evaluation questions – i.e. what to assess, where and when – we would also need to define its scope and the units to be analysed. Scope refers to what is included within – and excluded from – the evaluation’s lens of observation, or the evaluand. The units of analysis pertain to discrete areas of impact that need to be assessed within that scope. A unit of analysis may be defined by a certain type of intervention or a particular geographical location or scale, for example. It is the context within which data are collected, analysed and interpreted.

How then do we determine the evaluation’s scope and units of analysis? The evaluation’s purpose and objectives, as well as time and budget constraints, are a starting point. These help narrow down the areas which an evaluation’s users are interested in, and have the resources to find out. The intervention’s design is another logical place to begin, as we cannot evaluate it against impacts it never aimed to achieve. But as argued previously, looking only at what the intervention intended to achieve increases the risk that less obvious types of impact will be missed. Drawing on complex systems theory, we highlight the following elements of a TOC and its larger system that need to be identified and understood to serve as a basis for defining the evaluation scope and units of analysis, and thus the specific evaluation questions. These are elements that have relevance insofar as their potential to influence an intervention’s achievement of impact, and at the same time an intervention’s potential to influence these elements towards achieving a desired impact.

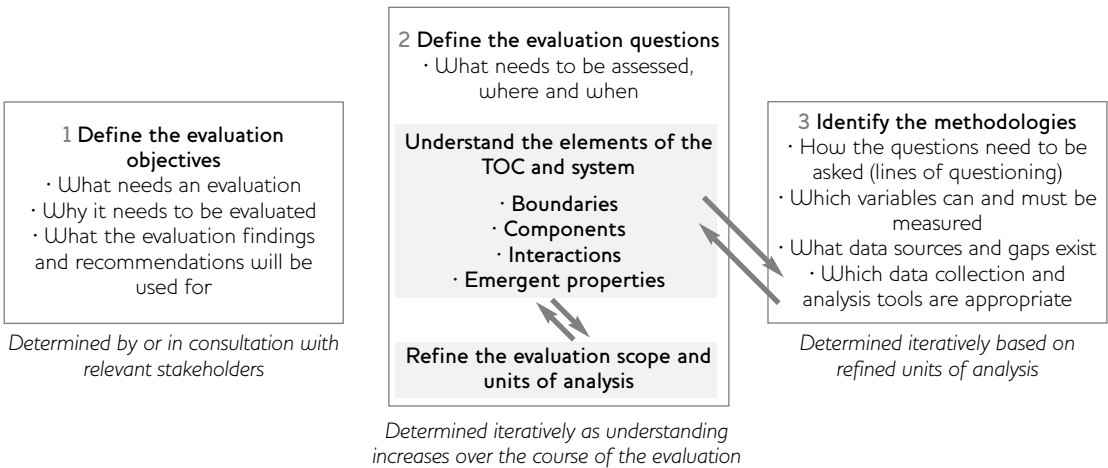
2.2.1 Boundaries

These refer to the geographical, temporal and domainial extent of the larger system that the intervention is part of. These are typically social, ecological, economic, political and administrative in nature, as well as whichever other domains have an influence on the intervention. Within each of these domain’s boundaries, it is also important to identify nested spatial and temporal scales that are likely to exist (Holling 2001; Cash *et al.* 2006). For example, what is the geographical coverage of an intervention that seeks to develop a national law to protect an ecosystem spanning several municipalities? What are the ecological boundaries of the ecosystem that the law seeks to protect? What are the political boundaries of the municipal governments that need to adopt the law? What are the social boundaries of the village-level communities that will be implementing the law? Are the boundaries covering the same geographical locations and scales, or are there mismatches (e.g. Folke *et al.* 1997)? By understanding how these different boundaries intersect (or do not intersect), we can draw the system boundaries that encompass where the intervention was implemented; where, when and how it intended to achieve impact; any factors that might influence the intervention’s achievement of impact; and any areas of the system that the intervention might influence.

2.2.2 Components

These are the ‘actors’ or ‘agents of change’ in the system, and may be human or non-human (e.g. Anderies, Janssen and Ostrom 2004). These would include all key stakeholders at the minimum, but may also include those who may be indirectly affected by change (whether positive or negative, e.g. a new money-saving economic policy in one country may cause income loss in families producing artisanal goods on the other side of the world, which in turn may lead to cascading and feedback effects closer to the area of intervention). These may also include those that are not directly affected by change but have the power to effect change (e.g. mining operations on a mountain polluting water sources downstream). Understanding a system’s components would entail understanding the intrinsic characteristics of each component, such as their interests, capacities and resources, goals, and worldviews, in the case of human actors (Brugha and Varvasovszky 2000). Another important characteristic would be the scales at which these

Figure 1 The proposed conceptual framework and its place within the evaluation design and implementation process



How this conceptual framework fits into the evaluation design process

⚡ Indicates iterative reviews resulting from new information, thus leading to adaptation of evaluation process

Source Authors' own.

components operate (e.g. Berkes 2006). Similar to how an intervention's system boundaries are defined, relevant components are identified based on their potential to influence the intervention's intended impacts, the intervention's mechanisms for achieving impacts, and the intervention's potential to influence those components.

2.2.3 Interactions

These refer to any action by one component that has an effect on another. These may be either regular interactions between two or more components (e.g. Ostrom 2009), or significant events resulting from the interaction of various components (e.g. Gladwell 2000). These interactions may take place across different spatial and temporal scales, as well as across domains, which can create mismatches such as time lags and other phenomena obscuring an intervention's effects. One example is how pollution abatement efforts in the Baltic Sea, shown to be successful in other seas over a typical project time frame of five years, did not result in similar changes there, as nutrient cycling in this sea takes place between 4 and 23 years (McQuatters-Gollop *et al.* 2009; Mee 2010). This suggests that even if all pollutive activities were to stop, pollution levels in the Baltic Sea cannot be expected to decrease before this time frame. Understanding such interactions provides a more realistic picture of what impacts may and may not be achieved by an intervention, and where

and when these impacts can be measured, rather than depending only on what the intervention initially targeted based on previous experiences.

2.2.4 Emergent properties

These are characteristics that no individual component may possess, but all components together exhibit. These may likewise refer to processes triggered by a combination of all interactions that transform the entire system, which no specific set of interactions could have done (Duhigg 2012). In a sense, these are what may be called the intrinsic characteristics of the system itself. This could be the extent to which the system has high diversity in terms of subject-matter expertise, or the extent to which formal and informal decision-making authority is evenly distributed, which then affects how individual components interact at various scales (Considine 2005; Carlsson and Sandström 2008). It may also refer to overall functions of the system, if the system itself has been created for the achievement of a specific goal, or the degree to which there is unity or diversity of purpose across the system (e.g. Feeny *et al.* 1990).

An understanding of these four elements can be initially derived from an extensive review of scientific journals, case studies and previous evaluations, as well as consultation of experts on the subject matter. This can provide a good sense at the outset of how the TOC's identified

elements match those of the larger system's – and thus to what extent the intervention's TOC is technically valid, or if certain assumptions were either missed or wrongly made.

2.3 Adapting the evaluation to evolving knowledge

Understanding which boundaries, components, interactions and emergent properties are most relevant to how an intervention contributed (or did not contribute) to impact is necessarily an iterative process as more information comes to light. A well-researched TOC may already have incorporated causal chains that reflect a thorough understanding of the larger system when the intervention was designed. But given the dynamic nature of complex systems, certain characteristics and pathways might only be uncovered once the evaluator starts asking questions on the field. Consequently, the evaluation's scope, and especially its units of analysis, may evolve many times as understanding of these elements increases over the course of the evaluation.

While the evaluation scope specifies what is included and not included in the assessment, the units of analysis further break this down into which variables need to be measured (or which indicators, if the desired impact is not directly or immediately observable), and the appropriate mix of methodologies for measuring and analysing these variables (i.e. the when, where and how). A complex system perspective helps identify what data are feasible and available for collection and what constraints can be expected, and thus which variables are most appropriate for answering the evaluation questions. Furthermore, such understanding helps an evaluator pursue multiple lines of questioning that enable a more comprehensive assessment of how impact has or has not been achieved. Figure 1 illustrates how an increasing understanding of an intervention's TOC and the larger complex system that it is part of requires an iterative refinement of the evaluation scope and units of analysis, and consequently of the methodologies appropriate for assessing different types of impact.

3 What might have been missed

This section provides illustrations from *ex post* impact evaluations carried out by the Global Environment Facility Independent Evaluation Office (GEF IEO) since 2010.² Each one highlights the less apparent impacts that would have escaped our attention had we not made the

effort to understand the elements of the larger complex system that each intervention was interacting with, and adapted the evaluation scope and units of analysis as needed. As all GEF-supported interventions involve complex social-ecological systems, understanding the characteristics of complex systems in general was a crucial starting point in identifying aspects that needed further investigation.

3.1 One plus one does not equal two

The 'Reversing Environmental Degradation Trends in the South China Sea and Gulf of Thailand' project (GEF ID 885, also known as the 'UNEP SCS' project)³ had as one of its objectives 'to create an environment at the regional level, in which collaboration and partnership in addressing environmental problems of the South China Sea, between all stakeholders, and at all levels is fostered and encouraged'. Thus, while the project's ultimate intended impact was to improve environmental trends in the South China Sea⁴ and Gulf of Thailand (collectively referred to here as SCS), its immediate target within the project time frame was regional collaboration among the seven GEF-eligible countries bordering the SCS, primarily through the adoption of a strategic action programme (SAP). The SAP was to outline regional activities to be undertaken by each country within the SCS, with the corresponding costs and timeline.

The SCS is known to be a particularly difficult area to work in. Maritime territorial disputes have put almost every country in the region at odds with each other even before the United Nations Convention on the Law of the Sea set out to delineate national exclusive economic zones in 1982. For a project to aim at fostering cooperation among these countries was an ambitious yet important step in achieving the kind of global environmental benefits that GEF has specifically been created to fund.

A terminal evaluation focused only on the project's TOC found that the UNEP SCS project was in fact successful in achieving its ambitious goal. Right before the project closed in January 2009, the seven countries, including China and Malaysia which had initially been very reluctant, signed a non-binding SAP that was acceptable to all. This meant that some issues, such as transboundary fisheries and certain coral reef areas, were not covered by the SAP; nevertheless,

it was considered an achievement to have all countries come to some agreement over one of the most disputed bodies of water in the world.

Project activities at the regional level were beneficial, according to participants interviewed (UNEP 2009; GEF IEO 2013). Scientific conferences and the Regional Scientific and Technical Committee formed by the GEF project convened the top scientists from each country to share ecological data and discuss common ecosystem management concerns. Similarly, Mayors' Roundtable meetings allowed local government chiefs across the region to share experiences and improve coordination in transboundary areas, such as the seagrass beds between Kampot Province in Cambodia and Kien Giang Province in Vietnam. An agreement to jointly manage the seagrass beds was eventually signed between the two provinces. Driving all these activities was a very determined project director, a typical 'champion', who did everything necessary to ensure that these agreements would be in place by project end.

By all appearances, GEF supported the right activities and the right persons in the right way. The project's focus on regional cooperation at the scale of the SCS rather than just on the delivery of interventions at more manageable national and local scales, as many other projects do, reflects the latest science on coastal and marine management, which has been moving away from the management of resources, sectors and single habitats towards ecosystem-based management (Vo and Pernetta 2010).

In the impact evaluation of GEF support to the SCS, however, further understanding of the system within which regional collaboration was aimed to take place showed that the SCS large marine ecosystem was not the appropriate unit of analysis for assessing this specific impact.

Delving deeper into the history of coastal and marine initiatives in the SCS,⁵ several inherent system characteristics soon became evident. First, countries historically did not allocate funds for regional environmental action plans; implementation was typically left to donors for funding through individual projects. As such, even though everyone had signed, no funds were forthcoming from the countries to implement the SAP. Second, none of the proposed multilateral

environmental agreements in the region had ever been adopted. Individually, the countries had no problems committing to global conventions, but as a region, they shied away from signing any legally binding environmental agreements. However, previous regional coastal-related initiatives had begun to bring countries into dialogue not by rallying them around a critical body of water or ecosystem as a common cause, but by appealing to their shared regional identity as sovereign East Asian nations. Assessing the impact of a direct environmental intervention (e.g. wastewater treatment plants, regulations on fish catch) requires paying attention to the SCS's oceanographic boundaries to measure the extent of the intervention's effects. But to assess the impact of a regional intervention with a political nature, the evaluation had to take on not the SCS but the larger East Asian geopolitical region as the unit of analysis.

Two main regional marine initiatives that geographically encompassed the SCS had existed prior to the UNEP SCS project, and continued to exist during its implementation: the Coordinating Body on the Seas of East Asia (COBSEA), and the Partnership for Environmental Management of the Seas of East Asia (PEMSEA), which was a series of projects also funded by GEF. Both initiatives included the seven countries bordering the SCS as well as other countries in the region. Realising that the system had been shaped by several similar stakeholder interventions long before the UNEP SCS project began, the evaluation had to assess the value of GEF's regional contributions using as a baseline the progress that had already been made and that continued to take place.

The assessment revealed that GEF support had indeed increased linkages across different scales and among different stakeholder groups by funding various communication activities. It had also allowed other donors to link their lower-scale interventions to higher-scale ones through the regional frameworks provided by the GEF projects. Upon closer examination, however, it became obvious that the GEF-funded PEMSEA projects had also successfully accomplished very similar things to what the UNEP SCS project had: bringing together the region's top scientists, enabling local government chiefs across East Asia to exchange experiences, and having 12 countries as well as 20 non-government stakeholders agree

Table 1 UNEP SCS: how understanding system elements reveals the appropriate unit of analysis, requiring an expanded scope when assessing the baseline

| Elements to be understood to define questions | TOC-only perspective | Complex system perspective |
|---|---|---|
| Boundaries | Degraded ecosystem covers South China Sea and Gulf of Thailand (oceanographic). | Discussions and decision-making on maritime issues within South China Sea take place among countries in East Asian region (geopolitical). |
| Components | National governments, local governments, and scientists undertake different streams of project activities towards reducing ecosystem degradation. | Several bilateral donors and regional organisations have implemented regional interventions since the late 1970s; GEF itself has another regional-scale intervention covering East Asia doing similar activities as this project. |
| Interactions | Steering committee meetings comprising participating countries, scientific conferences involving region's top scientists, and roundtable meetings among local government chiefs encourage trust-building and exchange of ideas on ecosystem management. | Long-standing maritime disputes continue between almost every country in the region; national governments have traditionally adopted regional environmental action plans but depend on donors to implement activities. |
| Emergent properties | Science-based discussions will convince countries to agree to fund and collaboratively implement SAP across the ecosystem. | Countries reluctant to enter into legally binding regional environmental agreements due to need to assert sovereignty. |
| Line of questioning | Did the disputing countries agree and adopt the SAP developed through the project? | What contributions did the project make in relation to what other regional initiatives have accomplished? What was unlikely to be achieved considering the nature of country relations in the region? |
| Findings on impact of GEF intervention | Interactions built trust and led to participating countries adopting modified SAP; environmental agreement forged between two provinces in two countries. | Similar regional activities of two GEF projects and existing organisations have increased communication, but diluted country resources and caused fatigue due to lack of coordination. |

Source Authors' own.

to adopt a Sustainable Development Strategy for the Seas of East Asia. PEMSEA was also steered by a very prominent project director who had led many of the region's previous coastal initiatives.

The sum of two successful GEF-supported projects, unfortunately, did not equal to greater impact in this instance. Instead, negative effects such as fatigue among government officials who had to attend two sets of regional activities organised by the two projects made them less willing to invest more resources over time, as these tended to interfere with their regular work. Furthermore, some countries indicated their

reluctance at having to create – and fund – new parallel structures when they had already begun to put money into COBSEA, which was formed ten years before GEF came into existence. Having two very strong, visionary project directors, while critical to the extraordinary success of both pioneering projects, became an obstacle to coordination between the projects. Each started with robust, complementary schools of thought that GEF has supported, but many times, the unresolved personal differences became choke points in implementing GEF's overall strategy in the region. Understanding that outcomes in complex systems are highly

dependent on initial conditions, the evaluation concluded that support for ‘champions’, when not properly managed, can result in big losses in momentum and synergy, which in this case has lasted more than a decade. Had the evaluation not modified its scope of questioning beyond the TOC’s specified boundaries, it would not have brought to light the emergent negative impacts of two otherwise individually stellar projects.

3.2 *Many roads lead to impact*

In contrast, the case of the ‘Mexico High-efficiency Lighting’ project (GEF ID 574, or ILUMEX) illustrates how an evaluation focused on an intervention’s proposed causality can result in alternative causal pathways being missed, and consequently to contributions being underestimated.⁶ ILUMEX was implemented by the World Bank from February 1995 to December 1997.⁷ It had as a main objective to demonstrate the technological and financial feasibility of reducing greenhouse gas emissions through the widespread installation of energy-efficient lighting. This objective was to be accomplished by building the capacity of the Federal Commission on Electricity (CFE), Mexico’s electric utility, for demand-side management.

The project also intended to develop a model to expand efficient lighting in the context of developing countries. Several barriers had to be overcome. On the supply side, a product that functioned under conditions of fluctuating electrical current had to be developed, and a production and distribution chain for this product had to be established. On the demand side, barriers such as the high upfront cost of compact fluorescent lamps (CFLs) compared to the traditional incandescent bulbs had to be mitigated. The project worked with large bulb manufacturers and CFE to develop extensive product marketing and promotion, strict purchasing specifications, laboratory testing, and public awareness campaigns. The project also introduced credit and subsidy instruments to provide incentives to consumers to purchase the new bulbs. When a follow-up independent evaluation took place in 1999, two years after project closing, 2.6 million CFLs had been sold – over 900,000 greater than the number that the project had targeted (OED 2002).

CFE found the lessons and management tools developed by the project useful, including such

things as development of standards, ways to engage the private sector, and testing of incentive mechanisms. While acknowledging these accomplishments, the follow-up evaluation downgraded the project’s rating from satisfactory to marginally satisfactory because the project had failed to demonstrate the financial feasibility of reducing greenhouse gas emissions. The independent evaluation also considered the sustainability of the project’s impact to be uncertain, because while the project did provide a replicable model of demand-side management, project design and implementation did not address how the model would be operationalised. Contributions to institutional capacities were rated as modest, as the project added little to CFE’s institutional arrangements.

The evaluation focused on assessing the extent to which the project achieved its specific objectives and to which it had clarity on how those objectives would be reached, but the evaluation did not give sufficient attention to the documented results of higher-than-expected CFL demand, and to how this had taken place. While CFE stated in its final self-evaluation that it would stop financing the trust funds established under the project to subsidise CFLs, it also indicated that it would integrate the lessons derived from ILUMEX to the Trust Fund for Energy Conservation (FIDE). This, while not representing an institutional change nor the sustainability of the project’s structures, did indicate CFE’s intention to mainstream project activities and lessons into an existing institution. Mainstreaming had in fact already begun to take place when the evaluation was conducted in 1999. Subsequently, key actors from the ILUMEX team moved to FIDE. But instead of shifting attention towards the unexpected positive developments that had taken place and investigating how the project might have contributed to it, the evaluation remained focused on the project’s lack of attention to specific sustainability pathways during project design and implementation. Indeed, after the project ended, FIDE continued and expanded the project activities, and by January 2007 had sold about 14 million additional CFLs (FIDE 2007).

Because of the bulk sales generated by ILUMEX and FIDE, manufacturers and distributors felt confident enough in the market’s sustainability to invest in distribution channels and marketing. According to a market study done in 2011, CFL

Table 2 ILUMEX: how a systems perspective reveals alternative causal pathways, requiring a shift in the line of questioning towards what happened rather than what should have happened

| Elements to be understood to define questions | TOC-only perspective | Complex system perspective |
|---|--|--|
| Boundaries | Energy-efficient lighting market in Mexico, including consumers and production and distribution chains (economic, national). | Market for energy-efficient lighting in Latin America, as well as for other sectors covered by Mexico's electric utility [in hindsight] (economic, regional). |
| Components | Mexico's electric utility and CFL manufacturers implementing project activities to influence consumers. | Mexico and other Latin American countries on steep economic growth curve, resulting in power supply shortages. |
| Interactions | Product development with large bulb manufacturers to supply bulbs suited to developing countries; incentives and marketing campaigns through electric utility for consumers to purchase CFLs. | Project consultants later hired to design and implement efficient lighting programmes within Mexico and other Latin American countries that also suffered from blackouts [in hindsight]. |
| Emergent properties | Influencing both manufacturers and the electric utility will result in increased demand for CFLs and therefore decreased GHG emissions. | The large influence of FIDE over different electric markets combined with the increase in CFL demand and improvements in technology triggered an increase in CFL supply, lower prices and consequently even higher demand. |
| Line of questioning | Were enough CFLs sold to reduce GHG emissions? Were financial and institutional arrangements set up to sustain or replicate the model developed by the project? | What has led to the unexpectedly high demand for CFLs? How did project outcomes contribute and might continue to contribute to this? |
| Findings on impact of GEF intervention | Target for number of CFLs sold exceeded, and project lessons adopted by electric utility, but no new institutional arrangements or financial commitments in place to allow replication as planned. | Higher-than-expected demand for CFLs and commitment to adopt lessons in existing Trust Fund through integration of project team indicated impact being achieved through alternative pathways. |

Source Authors' own.

sales in Mexico increased from 2.7 million units in 2000 (three years after ILUMEX) to an estimated 52.8 million by the end of 2010, increasing the percentage of CFLs to total light bulb sales from about 1 per cent to 20 per cent in ten years. Benefits to consumers took place further downstream when competition and technological innovations reduced CFL prices from US\$14 in 2000 to US\$3 in 2010.⁸

The project's transformative effects have spilled over to other areas as well. ILUMEX has been a model for more recent programmes in Mexico, particularly Luz Sustentable, which has delivered close to 44 million CFLs (SENER 2012). In December 2010, a mandatory energy efficiency

standard set by the national government prohibited the sale of low-efficacy lamps (such as the standard incandescent lamp) throughout the whole country by the end of 2013 (SENER 2010). Drawing on the economic analysis methodology developed by ILUMEX, FIDE introduced close to 180,000 high-efficiency motors in Mexico from 1998 to 2002, thus also transforming the market of electric motors in Mexico.

Cuba's Electricity Savings Programme initiated in 1998 also had ILUMEX as a reference. For this programme, Cuban government officials visited Mexico to learn from FIDE and CFE about the Mexican experience (OLADE 2012; Ramirez 2012). At the end of 2008, more than

9 million CFLs had substituted an equal number of incandescent lamps in Cuba, and the country was in the process of prohibiting the importation of incandescent lamps (CEPAL 2009).⁹ Following Cuba's example and advice, Venezuela began in November 2006 what was called Mission Energy Revolution that included a massive lamp replacement programme. With Venezuela's financial support and Cuba's technical advice, the model was also replicated in the Caribbean. Similarly, local utilities across Central and South America adapted and started using ILUMEX's utility service centre sales model as countries faced fast-growing power demand that was leading to blackouts (Marbek Consultants 2006; de Buen and Carpio 2010). In all of these cases, the key players from ILUMEX (manufacturers and lighting consultants) were part of the design and implementation teams (Ramirez 2012; pers. comm. Villaseñor 2012).¹⁰

That the initial success of ILUMEX has created cascading effects 16 years down the line in areas beyond the original project scope is an example of how impacts in a complex system may be greatly delayed, and manifest in unexpected places. Neither the project design nor the follow-up evaluation, which had expected market change to take place after only three years of implementation, seemed to take this into account. While such far-reaching influences could not have been predicted, early successes such as the higher-than-expected demand for CFLs and the mainstreaming of project lessons within an existing institution did indicate the existence of alternative pathways to which the project also contributed. Had the follow-up evaluation understood that impact in complex systems can occur through multiple causal pathways, and followed up with a line of questioning grounded on what had actually occurred rather than what the TOC specified should be occurring, it might have recognised that the conditions for achieving ILUMEX's long-term transformational goal were even then already in place.

3.3 Multiple lines of questioning

The example of ILUMEX shows how understanding the system and exploring a different line of questioning can reveal indications of impact being achieved in ways that the original TOC might not have foreseen. Another way to uncover alternative pathways to

impact is by asking about any contributions the intervention may have made within the larger system rather than focusing only on where it was intended to have an effect. This entails directly asking what difference an intervention has made, how different it is from other stakeholder interventions, and what might have happened had the intervention not taken place.

While potentially biasing key informants towards looking for positive effects to attribute to an intervention, these questions can nevertheless shed light on unexpected contributions whose impact can then be verified through other information sources and methodologies. To further mitigate bias, this line of questioning must only be taken once a good understanding of more direct causal pathways has been established. The aim is to discover other ways by which an intervention contributes to impact, allowing an assessment of an intervention's added value to a system where several other institutions and drivers may be dominant. A good knowledge of what activities the intervention implemented, and what previous evaluations have reported about similar interventions are essential in uncovering less obvious links between the intervention and the observed results.

This line of questioning has resulted in remarkably similar findings across geographically diverse evaluations on the impact of GEF support. One example is how GEF funding has enabled local government officials, civil society organisations and private land owners to collaborate in different protected areas in Indonesia, Namibia and Uganda, as well as facilitated dialogue among high-level ministers at the regional level in the Black Sea, South China Sea and other major transboundary bodies of water.¹¹ While meetings and discussions seem to have no clear impact on the environment as ultimately intended by these GEF-supported interventions, stakeholders interviewed credit them for cultivating the necessary conditions of trust and communication that precede any concerted action that leads to environmental impact on the ground in polycentric settings such as these (Ostrom 2008; Brondizio, Ostrom and Young 2009). Although it is impossible to predict any outcomes with certainty in complex systems, an understanding of which conditions would be necessary and/or sufficient for achieving the desired impact within

Table 3 Types of less apparent impacts uncovered by using different lines of questioning

| Line of questioning | Example in this article | Types of less apparent impacts uncovered in the example |
|--|---|---|
| Baseline-focused e.g. What similar interventions have existed within the system? How do the intervention's contributions interact with these? How has the system typically responded to this type of intervention? | Having two successful but uncoordinated regional interventions in the South China Sea. | Unintended and secondary (negative). |
| Results-focused e.g. What is the current trend or state of the system? What factors and events led to this state? What contributions did the intervention make in shaping these factors and events? | Successful but unexpected outcomes leading to long-term, large-scale effects beyond intended geographical area and market. | Unintended, indirect and secondary (positive). |
| Intervention-focused e.g. What contributions did the intervention make? What difference did these contributions make in the system? Are these contributions typically known to increase likelihood of impact in such systems? | Increasing communication in polycentric systems such as multi-stakeholder management of local protected areas and regional management of large marine ecosystems. | Intended and indirect (positive). |

Source Authors' own.

the specific system allows us to assess the extent to which an intervention has increased the likelihood of impact occurring.

Using multiple lines of questioning allows us to uncover different types of impact that might not be immediately apparent if we approach an evaluation with only the TOC as a lens. Table 3 highlights the different lines of questioning that were used in the examples presented, and how they led us to assessing the extent of impacts (or contributions towards achieving impact) that either were not identified in the TOC, or not immediately observable during data collection. Among other things, it shows how asking similar questions, but with a different focus each time, allows us to estimate the maximum extent of impact that might be expected, as a function of both the specific complex system's responses and the intervention's contributions. Combining these and other lines of questioning develops a multi-faceted understanding of the system, and of the different ways an intervention may have contributed towards achieving impact in that system – in effect, estimating the various counterfactuals and discounting rival hypotheses by asking the same question in different ways.

4 Conclusion

Theory-based evaluation evolved as an approach that specifically counters evaluations driven purely by methodologies, or purely by the need to measure impact, without regard to the how or why of it (Chen 2012). This approach is based on 'a model, theory, or philosophy which indicates the causal relationships supposedly operating in the program' (Fitz-Gibbon and Morris 1996). However, many evaluators apply this approach by deriving evaluation questions primarily from an intervention's TOC, without considering the effects produced by an intervention's interactions with other elements that can also influence the achievement of impact. Understanding the characteristics and effects of these elements is particularly critical in complex systems due to the non-linear pathways confounding an intervention's contributions.

TOCs simplify the complex process of achieving impact by outlining a path towards a desired goal. While a TOC-focused approach is useful for targeting project or programme activities that are likely to yield the desired outcomes, its narrow use in impact evaluation should be approached with caution. Impact at scale

requires transformational processes that may be composed of multiple pathways going in different directions. As the intervention's TOC is likely to direct the attention of the evaluator to the expected causal links, it can also act as a blinder to chains of causality that had not been contemplated in the intervention design, and are rarely apparent by project completion. Excessive focus on the intervention and its effects also increases the risk that the effects of other factors will be downplayed in the assessment.

In addition to understanding how the intervention has been designed to achieve impact, the evaluator must take on a complex system perspective and have a solid, well-researched understanding of how the larger system influences and is influenced by the intervention. The intervention then is viewed as only one of the many agents of change contributing towards a certain outcome, whether intended or unintended. This allows the evaluator to uncover how the intervention contributes to the various types and degrees of

impact in a specific setting, and not just through direct attribution. This is accomplished through the identification and understanding of the intervention's and the larger system's boundaries, components, interactions and emergent properties, which typically include different domains, and go across spatial and temporal scales. The examples show that such understanding helps determine where, when, and which impacts may be measured.

Furthermore, understanding these elements with a complex system perspective helps identify indicators of progress towards impact, in cases where impact takes place over a timescale greater than the intervention's, or where it cannot be measured directly. Finally, by using the deepening understanding of both the intervention and the system to refine the evaluation questions, evaluators are able to systematically adapt approaches and methods to emerging information while still meeting the evaluation's agreed-upon objectives – a necessary skill for successfully carrying out evaluations that themselves take place within complex systems.

Notes

- 1 See www.merriam-webster.com.
- 2 The GEF is a financial mechanism created in 1992 to help developing countries fulfil their obligations to global environmental conventions. It is a partnership composed of 183 countries, development banks, UN agencies, civil society organisations and the private sector. More information can be found at www.thegef.org/gef/whatisgef.
- 3 The UNEP SCS project cost US\$32.5 million, of which US\$16.75 million was a GEF grant. It was implemented by UNEP from 2002 to 2009.
- 4 In China, this sea is called the 'South Sea', 南海 Nánhai, and in Vietnam, it is known as the 'East Sea', Biên Đông. In September 2012, the president of the Philippines issued Administrative Order No 29 renaming '...maritime areas on the western side of the Philippine archipelago as the West Philippine Sea...', to cover the areas within the Philippines' Exclusive Economic Zone. 'South China Sea' is the international name for this area as recognised by the International Hydrographic Organization, and is therefore the name used throughout this evaluation.
- 5 This was done through the analysis of project documents and meeting minutes of previous initiatives, regional reports, databases of laws and donor funding in the region, and relevant websites, as well as through interviews of persons with historical knowledge. The findings in this example are derived from the Impact Evaluation of GEF Support to the South China Sea and Adjacent Areas (GEF IEO 2013).
- 6 The findings in this example are derived from a case study developed for the Impact Evaluation of Climate Change Mitigation: GEF Support to Market Change in China, India, Mexico and Russia (GEF IEO 2014) by ENTE, led by Odón de Buen.
- 7 The total cost of the project was US\$25.9 million, of which US\$10.7 million was financed by the GEF, US\$3.1 million by the Kingdom of Norway, and US\$12.1 million by the CFE.
- 8 See www.globaldata.com/static/120312_twt_Compact_Fluorescent_Lamp.html.
- 9 The population of Cuba is close to 11 million people.
- 10 J. Villaseñor, interview regarding ILUMEX, 2012.
- 11 These are based on findings from the Impact Evaluation of GEF Support to the South China Sea and Adjacent Areas (GEF IEO 2013) and the Impact Evaluation of GEF Support to Protected Areas and Protected Area Systems, which is currently in progress.

References

- Anderies, J.M.; Janssen, M.A. and Ostrom, E. (2004) 'A Framework to Analyze the Robustness of Social-ecological Systems from an Institutional Perspective', *Ecology and Society* 9.1: 18
- Berkes, F. (2006) 'From Community-based Resource Management to Complex Systems: The Scale Issue and Marine Commons', *Ecology and Society* 11.1: 45
- Blamey, A. and Mackenzie, M. (2007) 'Theories of Change and Realistic Evaluation: Peas in a Pod or Apples and Oranges?', *Evaluation* 13.4: 439–55
- Brondizio, E.; Ostrom E. and Young, O.R. (2009) 'Connectivity and the Governance of Multilevel Socioecological Systems: The Role of Social Capital', *Annual Review of Environmental Resources* 34: 253–78
- Brugha, R. and Varvasovszky, Z. (2000) 'Stakeholder Analysis: A Review', *Health Policy and Planning* 15.3: 239–46
- Carlsson, L.G. and Sandström, A.C. (2008) 'Network Governance of the Commons', *International Journal of the Commons* 2.1: 33–54
- Cash, D.W.; Adger, W.N.; Berkes, F.; Garden, P.; Lebel, L.; Olsson, P. and Young, O. (2006) 'Scale and Cross-scale Dynamics: Governance and Information in a Multilevel World', *Ecology and Society* 11.2: 8
- CEPAL (2009) *Situación y Perspectivas de la Eficiencia Energética en América Latina y el Caribe*, Documento de proyecto, Santiago: Comisión Económica para América Latina (CEPAL)
- Chen, Huey T. (2012) 'Theory-driven Evaluation: Conceptual Framework, Application and Advancement', *Evaluation von Programmen und Projekten für eine Demokratische Kultur*: 17–40
- Considine, M. (2005) 'Partnerships and Collaborative Advantage: Some Reflections on New Forms of Network Governance', background paper for conference on Governments and Communities in Partnership: From Theory to Practice, *Australian Policy Online*, http://web.archive.org/web/20051217064117/http://www.public-policy.unimelb.edu.au/conference06/Considine_Background_Paper.pdf (accessed 9 November 2014)
- de Buen, O. and Carpio, C. (2010) *Estudio de la Situación de los Programas de Sustitución de Lámparas en el Sector Residencial y en el Sector del Alumbrado Público de América Latina*, Washington DC: Inter-American Development Bank
- Duhigg, C. (2012) *The Power of Habit: Why We Do What We Do in Life and Business*, London: Random House LLC
- Feeny, D.; Berkes, F.; McCay, B.J. and Acheson, J.M. (1990) 'The Tragedy of the Commons: Twenty-two Years Later', *Human Ecology* 18.1: 1–19
- FIDE (2007) *Informe de Operaciones 2006, Ahorro de Energía Eléctrica en México*, México DF: Fideicomiso para el Ahorro de Energía Eléctrica (FIDE)
- Fitz-Gibbon, C.T. and Morris, L.L. (1996) 'Theory Based Evaluation', *Evaluation Practice* 17: 177–84
- Folke, C.; Pritchard Jr., L.; Berkes, F.; Colding, J. and Svedin, U. (1997) 'The Problem of Fit between Ecosystems and Institutions: Ten Years Later', *Ecology and Society* 12.1: 30
- GEF IEO (2014) *Impact Evaluation of GEF Support to Climate Change Mitigation: Transforming Markets in Major Emerging Economies*, Washington DC: Global Environment Facility Independent Evaluation Office
- GEF IEO (2013) *Impact Evaluation of GEF Support to the South China Sea and Adjacent Areas*, Washington DC: Global Environment Facility Independent Evaluation Office
- Gladwell, M. (2000) *The Tipping Point: How Little Things Can Make a Big Difference*, Boston MA: Little Brown
- Holling, C.S. (2001) 'Understanding the Complexity of Economic, Ecological, and Social Systems', *Ecosystems* 4: 390–405
- Mackenzie, M. and Blamey, A. (2005) 'The Practice and the Theory: Lessons from the Application of a Theories of Change Approach', *Evaluation* 11.2: 151–68
- Marbek Consultants (2006) *Mexico – Ilumex Project: Post-Implementation Impact Assessment*, Washington DC: World Bank GEF
- Mayne, J. (2011) 'Contribution Analysis: Addressing Cause and Effect', in R. Schwartz, K. Forss and M. Marra (eds), *Evaluating the Complex*, New Brunswick NJ: Transaction Publishers
- McQuatters-Gollop, A.; Gilbert, A.J.; Mee, L.D.; Vermaat, J.E.; Artioli, Y.; Humborg, C. and Wulff, F. (2009) 'How Well do Ecosystem Indicators Communicate the Effects of Anthropogenic Eutrophication?', *Estuarine, Coastal and Shelf Science* 82.4: 583–96
- Mee, L.D. (2010) 'Between the Devil and the Deep Blue Sea: The Coastal Zone in an Era of Globalisation', *Estuarine, Coastal and Shelf Science* 96.1: 1–8

- Morra Imas, Linda G. and Rist, Ray C. (2009) *The Road to Results: Designing and Conducting Effective Development Evaluations*, Washington DC: World Bank, <https://openknowledge.worldbank.org/handle/10986/2699> (accessed 9 November 2014)
- OECD (2002) *Glossary of Key Terms in Evaluation and Results Based Management*, Paris: OECD, www.oecd.org/development/peer-reviews/2754804.pdf (accessed 9 November 2014)
- OED (2002) *High-efficiency Lighting Project in Mexico, Précis*, Washington DC: World Bank Operations Evaluation Department
- OLADE (2012) *Iniciativas Nacionales Cuba*, www.olade.org.ec/portal-eficiencia-energetica/experiencias-regionales-cuba (accessed 28 November 2012)
- Ostrom, E. (2010) 'Beyond Markets and States: Polycentric Governance of Complex Economic Systems', *American Economic Review* 100: 1–33
- Ostrom, E. (2009) 'A General Framework for Analyzing Sustainability of Social-Ecological Systems', *Science* 325.5939: 419–22
- Ostrom, E. (2008) *Polycentric Systems as One Approach for Solving Collective-Action Problems*, SSRN, <http://ssrn.com/abstract=1304697> (accessed 9 November 2014)
- Patton, M.Q. (2011) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, New York NY: Guilford Press
- Pawson, R. and Tilley, N. (2004) *Realist Evaluation*, British Cabinet Office, www.communitymatters.com.au/RE_chapter.pdf (accessed 9 November 2014)
- Ramalingam, B. and Jones, H. (2008) *Exploring the Science of Complexity: Ideas and Implications for Development and Humanitarian Efforts*, London: Overseas Development Institute
- Ramirez, A. (2012) *Perspective of Long-term Impacts of ILUMEX*, Mexico DF
- SENER (2012) *Inicia Segunda Etapa del Programa Luz Sustentable*, Global Industries, www.globalindustries.mx/index.php/explore/features/dropline-menu/126-inicia-segunda-etapa-del-programa-luz-sustentable (accessed 28 November 2012)
- SENER (2011) *México Obtendrá Recursos por Certificar Reducción de Emisiones por el Cambio de Lámparas Ahorradoras*, www.sener.gob.mx/portal/Mobil.aspx?id=2059 (accessed 28 November 2012)
- SENER (2010) *Eficiencia Energética de Lámparas para uso General. Límites y Métodos de Prueba*, Mexico DF: Secretaría de Energía, Diario Oficial de la Federación: 16
- Stein, D. and Valters, C. (2012) *Understanding Theory of Change in International Development*, London: Justice and Security Research Programme
- UNEP (2009) *Reversing Environmental Degradation Trends in the South China Sea and Gulf of Thailand – Terminal Evaluation*, Nairobi: United Nations Environment Programme
- Van den Berg, R. and Todd, D. (2011) 'The Full Road to Impact: The Experience of the Global Environment Facility Fourth Overall Performance Study', *Journal of Development Effectiveness* 3.3: 389–413
- Vo, S.T. and Pernetta, J. (2010) 'The UNEP/GEF South China Sea Project: Lessons Learnt in Regional Cooperation', *Ocean and Coastal Management* 53.9: 589–96
- Vogel, I. (2012) *Review of the Use of 'Theory of Change' in International Development*, London: Department of International Development
- White, H. (2009) *Theory-Based Impact Evaluation: Principles and Practice*, Working Paper 3, New Delhi: International Initiative on Impact Evaluation, 3ie