



The full road to impact: the experience of the Global Environment Facility Fourth Overall Performance Study

Rob D. van den Berg & David Todd

To cite this article: Rob D. van den Berg & David Todd (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, DOI: [10.1080/19439342.2011.562302](https://doi.org/10.1080/19439342.2011.562302)

To link to this article: <http://dx.doi.org/10.1080/19439342.2011.562302>



Published online: 26 Sep 2011.



Submit your article to this journal [↗](#)



Article views: 71



View related articles [↗](#)

The full road to impact: the experience of the Global Environment Facility Fourth Overall Performance Study

Rob D. van den Berg^{a*} and David Todd^b

^a*Global Environment Facility Evaluation Office, Washington, DC, USA;* ^b*International Development, Environment and Disasters (IDEDES), Barbados*

The Evaluation Office of the Global Environment Facility (GEF) has developed an approach to impact evaluation that aims to go beyond causalities and counterfactuals into replication, up-scaling and longer-term lasting systemic changes. Starting from the OECD Development Assistance Committee glossary definition of impact and building on a theory-based approach to evaluation, several methods and tools have been developed that cover various aspects of the long road from outcomes to impact. This article presents both the development over time of these methods and tools and methodological considerations. Two full-scale impact evaluations were undertaken: the first on protected areas in Eastern Africa and the second on the GEF portfolio of projects supporting the removal from production and consumption of ozone-depleting substances in countries with economies in transition. In addition several other impact studies were undertaken. The evaluations of various aspects of the progress from outcome to impact were gathered and as far as possible aggregated in the Fourth Overall Performance Study of the GEF, which was presented in 2009 to the replenishment process of the GEF. An overview is presented of progress toward impact of the GEF portfolio of finished projects on climate change mitigation and on stemming the loss of biodiversity. The Office will continue to develop and use the variety of impact evaluation tools at its disposal.

Keywords: impact environment; evaluation; theory-based approach

1. Background

When the Global Environment Facility (GEF) Evaluation Office started developing its approach to impact evaluation in 2005, extensive methodological discussions took place in the office to determine how the issue should be tackled. At that time the most commonly shared theoretical perspective on evaluations was objective-based evaluation through a theory-based approach, as pioneered by Chen (1990); of which the realist evaluation approach developed by Pawson and Tilley (1997) is the most relevant to the kind of strategy and interventions of the GEF. However, in the broader development community at this time, there was also a growing emphasis on the value of establishing attribution through the use of counterfactual assessments. In keeping with the differing emphases in the evaluation community, the Evaluation Office decided to focus on how a range of approaches could be applied to the specific evaluandum of the GEF. This multi-stranded perspective

*Corresponding author. Email: rvandenberg@thegef.org

was put into practice through two major field-based evaluations of impact and a series of desk studies, field visits and quasi-experimental studies, as well as a first tentative field study of a randomised controlled trial. This article aims to highlight the discussions on several aspects of impact that took place in the GEF Evaluation Office; and how these led to a distinctive mixed-methods approach, which aimed to generate a comprehensive body of data through the use of methods tailored to address the broad range of activities and approaches supported by GEF financing.

The first overarching issue to be faced was the definition of impact. Many evaluation offices start from the definition in the glossary published by the OECD Development Assistance Committee (DAC) (2002, p. 24) that impacts are the ‘positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended’. This definition identifies a plurality (impacts) and fits most closely with a contribution rather than an attribution perspective – an intervention is here studied for any linkages to developments outside the scope of the project; geographically, in terms of time scale and in the substance of the intervention.

A more recent definition of impact, which supports a different perspective, is that espoused by the International Initiative for Impact Evaluations (3IE) initiative (3IE 2010a). This defines impact as the ‘net change in outcomes for a particular group of people that can be attributed to a specific program using the best methodology available, feasible and appropriate to the evaluation question that is being investigated and to the specific context’ (3IE 2010a, p. 6) This definition is focused on one particular perspective on impact and is also in its purest form (of a randomised controlled trial) confined to the boundaries of the project, both in time and space, but not necessarily as regards substance. It is interesting to note that in its guide for grantees, 3IE (2010b) stresses the need for a mixed methods approach, including qualitative and process-oriented methods to be able to decide additional questions, such as why an intervention succeeds or fails.

When looking at these two major approaches to the definition of impact and what they imply for how the GEF should undertake its impact evaluations, it seemed clear that both address important aspects of impact. It seems vital to go beyond project boundaries and check whether and how interventions made a difference in the larger scheme of things: as well as whether there were any unintended consequences, which led to negative results that may offset the positive achievements. For the GEF the big question is whether global environmental trends are moving in the right direction and whether the support that the GEF provides to countries helps and supports them in tackling global environmental issues. On the other hand, it is also clear that if a supported intervention does not seem to have ‘worked’ and does not provide clear evidence that it actually made a difference, then the targeted global trends have not been effectively influenced by GEF support.

Given these considerations, the Office decided to focus mainly on the OECD DAC definition of impacts, without losing sight of more specific concerns about the extent to which the evidence base is sufficient to show that interventions have made a difference. As noted by Howard White (2009), the two approaches to impact evaluation are different, but not mutually exclusive. The OECD DAC definition focuses on ‘long-term effects’. These are fundamental to an organisation such as the GEF that is dealing with global environment change, which usually occurs over long periods of time. The definition does not explicitly mention attribution. In interventions supported by the GEF, many prior evaluations by the Office had shown that it is rarely the case that the intervention is the only factor causing a longer-term change. The results of the GEF can therefore best be described in terms of partial attribution, or contribution, since these terms are effectively interchangeable. In terms

of the ‘philosophical’ dimensions of causality, GEF interventions are often necessary for a desired change to happen, but rarely sufficient on their own to ensure that it does.

The next step was to identify what kind of impacts the GEF actually aimed to achieve. For this it was essential to first understand how the broad aims of the GEF are intended to be addressed. This led to an initial study of ‘global environmental benefits’ in its largest focal area of biodiversity, undertaken by Nair (2007), to better identify what kind of interventions the GEF was funding and how these could be expected to ‘work’ and to achieve broader ‘impacts’.

One important additional element concerned how best to identify the effects of any implementation failure. Earlier evaluations conducted by the Office indicated that sometimes interventions can be shown to ‘work’; but if certain follow-up activities do not take place, or if these interventions are copied to other situations without sufficient translation to local circumstances, the ‘working’ stops and the progress toward broader impacts is halted. To begin to address this issue, further evaluative work was done on the ‘catalytic role’ of the GEF. This uncovered various types of implementation failure; from mistakes in replication, to issues of scale and geographic scope. It can be argued that the identification of such implementation failure should be an essential element in the evaluation of impacts; and could be a useful initial step, since a counterfactually oriented evaluation does not make sense if the intervention has not been applied correctly in the first place.

Lastly it was recognised that whether or not an intervention was making a big or a smaller difference, or no difference at all, it was important to present this in the broader context of emerging trends in development, the environment and, more particularly, the global environmental benefits that were intended to be achieved. This often led to a historical perspective of presenting the timelines of and events in sustainable development in countries and regions, providing a comprehensive narrative of ‘what happened’, to better understand what role the GEF support played.

These early explorations led to the recognition that impacts in the GEF need to be evaluated at several levels and time scales (see Vaessen and Todd 2007):

- Attributable impact in the form of direct cause and effect, as established through experimental means, either in laboratory circumstances or field tests or through randomised controlled trials, quasi-experimental studies or modelling.
- Impact progress: whether all elements are in place to lead to impact, as established through evaluation of implementation failure and processes in place to ensure scaling up to the required level. This would include ‘necessary but not sufficient’ conditions that form stop or go elements for progress toward impacts.
- Historical impact: evaluating what happened rather than focusing on causal linkages.
- Impacts as defined by OECD DAC: evaluating what happens after the project ended, inside and outside the boundaries of the project; focusing on long-term effects, including unintended consequences and impacts of other developments, either natural or man-made, that influence the results of the interventions.

Given these various levels and aspects of impacts, the Evaluation Office decided to adopt a theory-based approach as a first step to uncover what each intervention was supposed to achieve in its own setting and as a platform from which to go forward to assess higher-level impacts. This step was expected to lead to the identification of questions for verification and evaluation, using the most appropriate methodology for each.

2. Overview: applying the theory-based approach

The theory-based approach of the GEF Evaluation Office builds on the ubiquitous development assistance project design tool, the logical framework or 'log-frame'. More recent cohorts of projects in the GEF portfolio have log-frames, which outline their intended inputs, activities, outputs and outcomes, as well as the risks that need to be taken into account. When the theory-based approach was applied to three Protected Areas biodiversity projects in Eastern Africa (CDC 2007a–d, Global Environment Facility Evaluation Office [GEFEO] 2007, 2008a, b), it quickly became clear that the log-frames in themselves did not provide sufficient information to construct a full 'theory of change' for the projects. Questions concerning how these projects were supposed to achieve their intended long-term objectives could best be answered in consultation with former staff of the project. All three projects had ended, so formally there was no longer a project team with which to interact, although it did prove possible to contact many of the original implementing team members. Furthermore, since all three projects concerned protected areas, the current management of these areas could be considered the successors to the projects and could therefore be involved in the reconstruction of their theories of change. As well as management and staff of the protected areas, former technical advisors and consultants were also invited to participate. In an initial workshop held at one of the project sites in East Africa, these stakeholders reconstructed the theories of change in an interactive process developed to assess the extent and nature of data needed to evaluate to what extent they had succeeded.

The establishment of these project-specific theories of change, incorporating assumptions on causal linkages and on actions of a broad range of stakeholders, led to the discovery of three important factors that helped in the further development of the impact approach in the GEF Evaluation Office. Firstly, it became clear that the pathways to impacts were varied and complicated and followed multiple and different routes for various outcomes. However, a series of concrete intermediate states could be identified, which signified systemic changes that would bring outcomes forward toward impacts.

Secondly, it became clear that the assumptions on causal mechanisms initially identified in project log-frames were located relatively early in the results chain. Whilst these early levels of the results chain might lend themselves to quasi-experimental analysis using counterfactuals and control groups, this would not uncover whether the higher-level objectives were achieved, but only whether a relatively small part of the overall theory of change could be proven to 'work'. A secondary but related issue was that on the activity level there were many parameters, all of which would require their own control groups, producing an unmanageable set of variables.

Thirdly, it became clear that the mechanisms that would actually ensure impacts (defined as 'long-term' results) were often largely out of the direct control of the project. They were not included in the projects themselves, but were mechanisms that were supposed to 'work' if the project created the right enabling circumstances for them. For example, Protected Areas are supported by the GEF not as a goal in themselves, but as a means to improve the status of global biodiversity. In Eastern Africa, the natural habitat of species is declining rapidly due to population pressure and a range of other factors. Clearly, the higher level impacts that the GEF aims to catalyse, require a range of social, institutional, biological, natural and physical processes that cannot be directly controlled by any GEF-supported intervention. What the GEF support can do is create an environment, which will enable these processes to start operating. Thus the GEF interventions are

likely to be ‘necessary but not sufficient’ to achieve the desired long-term changes, which constitute impact.

Building on these early lessons from the implementation of impact evaluation at field level in East Africa, the GEF Evaluation Office proceeded to test the broad approach in a completely different context: GEF support for the phase out of Ozone Depleting Substances in Countries with Economies in Transition (GEFEO 2009b). At the same time, a substantial process was undertaken to develop novel approaches that could ‘scale up’ the understanding of progress towards impacts across the GEF portfolio, using existing data sources and a new abbreviated method of field-level impact evaluation. These developments are described in more detail in this paper, as well as in a set of supporting documents published on the GEF website (www.gefeo.org).

3. The setting: the Global Environment Facility

The GEF was established in 1991 as an independent financial mechanism to help developing countries to play a role in protecting the global environment, whilst at the same time promoting national-level approaches to sustainable development. The Facility now has 182 participating member countries, including 32 that currently contribute to the GEF Trust Fund. It is governed by a Council, which includes representatives of donor and recipient countries, as well as a few that fall into both categories. A broad range of partnerships has been established to implement GEF-assisted activities, incorporating national governments, international agencies, non-governmental organisations, the scientific community and the private sector.

The GEF is linked to a range of multilateral agreements and institutions. It is a financial mechanism for:

- The Convention on Biological Diversity.
- The UN Framework Convention on Climate Change.
- Stockholm Convention on Persistent Organic Pollutants.
- The UN Convention on Combating Desertification.

The GEF also collaborates closely with other treaties and agreements to reach common goals in the areas of international waters and ozone layer depletion.

According to figures reported in the Fourth Overall Performance Study of the GEF (GEFEO 2010a), the GEF has allocated more than \$10 billion to over 1700 projects in 140 countries, matched by external contributions totalling more than twice this amount. This level of funding makes the GEF the largest single source of international public support for the global environment. The approximate distribution of GEF expenditure among its main Focal Areas is shown in Figure 1.

The GEF does not directly implement the projects to which it contributes. It operates through implementing agencies, which may delegate execution of the project to other agencies or to partner governments. Historically, three agencies formed the GEF partnership, each with a notionally distinct set of characteristics, which were intended to be complementary. The United Nations Development Programme has been particularly engaged with capacity development and technical assistance activities, whilst the United Nations Environment Programme has taken the lead on transboundary, regional and global projects, as well as normative and research projects. The World Bank has been the major player with regard to investment projects. More recently, seven additional agencies have

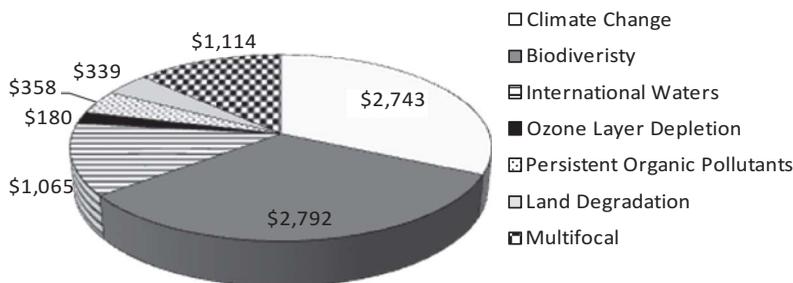


Figure 1. GEF portfolio by Focal Area supported, June 2009 (in US\$ million).

been enabled to directly access GEF funds and, over time, this will change the balance of partnerships. The new agencies are: the Food and Agriculture Organisation of the United Nations, the UN Industrial Development Organisation, the International Fund for Agricultural Development, as well as the African, Asian and Inter-American Development Banks, and the European Bank for Reconstruction and Development.

The ultimate objective of the GEF is to contribute to a broad range of global environment benefits including:

- Reducing the risks of climate change.
- Stemming biodiversity loss.
- Preventing ozone layer depletion.
- Safeguarding international waters.
- Eliminating persistent organic pollutants.
- Preventing land degradation.

Few, if any, of these changes can be achieved through direct means, and most require intermediary steps in terms of changed human behaviour towards the environment, such as:

- In biodiversity, interaction between local populations and natural resources.
- In climate change, energy use and efficiency.
- With regard to international waters, reduction of overfishing, nutrient spills and marine pollution.
- Concerning land degradation: changes in land and water use, particularly in farming systems.

The complex areas in which the GEF seeks to intervene involve it in an extremely broad range of issues and activities. It focuses on multidimensional ‘problems’, which require multi-disciplinary ‘solutions’, which incorporate human, natural, chemical, technical and economic aspects. Some of its activities are location specific, notably in such areas as biodiversity, international waters and land degradation. Its activities target a broad range of socio-economic categories, including indigenous people, migrants, farmers and entrepreneurs; and often aim to catalyse technological interventions, including agricultural waste management, energy efficiency and alternative energy sources. Many of these interventions are of a ‘demonstration’ nature, with replication and scaling up left to ‘market forces’. These complexities present a structural challenge to gather evidence on results, as analysed in Vaessen and van den Berg (2010).

4. Developing impact evaluation in the GEF: preliminary stages

The approach paper developed by Todd and Brann (2007) for the Evaluation Office, in preparation for its first impact evaluation, explored a range of potential methodologies, each of which had advantages and constraints in view of the specific characteristics of the GEF activities. A theory-based approach was identified as the most viable and was further developed by Foundations of Success (2007) on the basis of detailed desk research into the biodiversity portfolio of the GEF.

The desk research showed that GEF strategies, programmes and projects are based on assumptions concerning why and how they are supposed to generate the anticipated benefits. For some activities, these assumptions are relatively explicit in formal documents, but in many others they are not; in which case they need to be carefully drawn out before they can be evaluated. Project theories can be assembled on the basis of programme and project documents, to include the original strategy on which the intervention is based, as well as its intended outputs, outcomes and impacts.

5. The pilot impact evaluations

On the basis of the approach developed, three Impact Evaluations were undertaken on projects in East Africa: The Lewa Nature Conservancy Project (Kenya), the Bwindi-Mgahinga Project (Uganda) and the Cross-Borders Project (Kenya, Tanzania and Uganda) (see CDC 2007b–d, GEF Evaluation Office 2007, 208a, b). The purpose of these pilots was to test a theory of change approach to understanding impacts in the biodiversity Focal Area and to assess its potential for broader application across the GEF portfolio.

The objective of the Bwindi-Mgahinga Project was to ensure the long-term status of a population of mountain gorillas, an endangered species on the IUCN Red List. The gorillas were largely resident within the boundaries of two inter-linked protected areas, but their population had been in decline for many years. In order to create a theory of change to describe how the intervention intended to achieve its objectives, the evaluation team first analysed the key threats to the status of the gorilla population. Several were identified, including encroachment on the habitat by surrounding communities, danger of forest fires and poaching. For each threat, a problem tree analysis was undertaken, to establish its underlying reasons and what steps could be taken to address them. This provided the raw material for the construction of a results chain, against which the intervention's success in addressing the underlying reasons for the decline in the gorilla population could be evaluated.

As an example of the problem tree analysis, the threat to the gorilla population presented by poaching is presented in Figure 2.

The cause and effect chain was then converted to a means-end chain, illustrating how the threat posed by poaching was to be overcome by the intervention, as shown in Figure 3.

The steps in building the theory of change were therefore as follows. First, the major strategies were identified, which the project had designed to reduce the threats to and enhance the status of the intended Global Environment Benefit, in this case the gorilla population. Second, the means-ends linkages designed to deliver the strategy were specified. Third, key assumptions underlying the means-ends linkages were outlined. It was then possible to assess whether and how the intended impacts had been achieved. Data were gathered from documentary sources, through workshops with local level stakeholders and by means of fieldwork conducted in and around areas targeted by the project. In terms

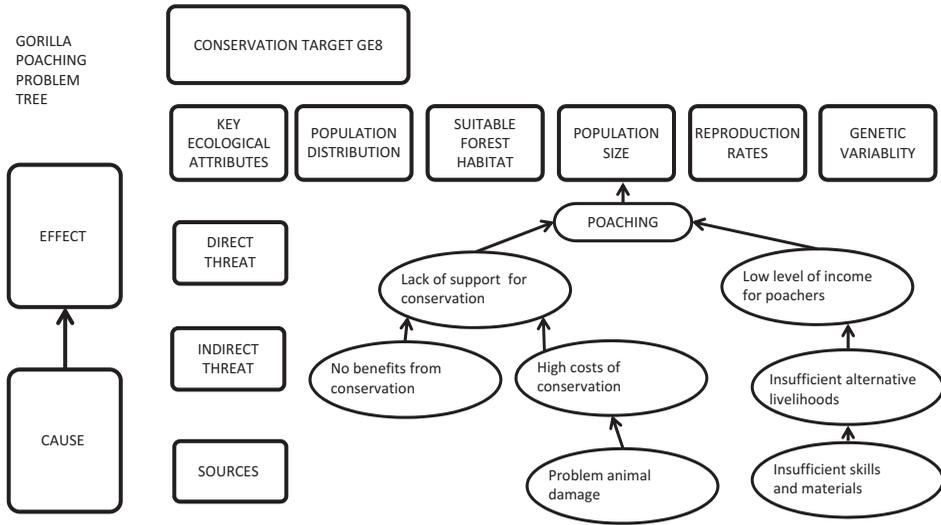


Figure 2. Problem tree of cause and effect of poaching on the gorilla population.

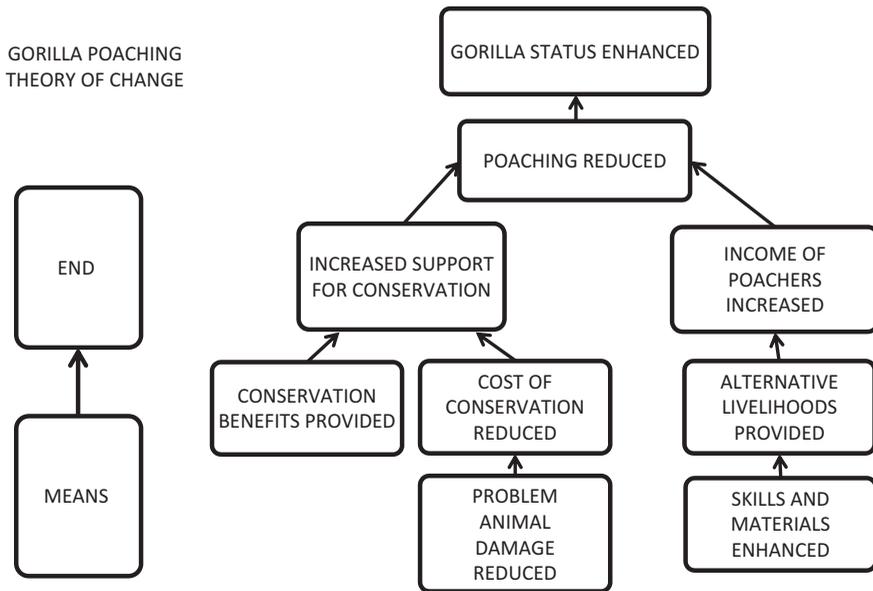


Figure 3. Means-end chain of project intervention targeting poaching.

of the analytical approach adopted, it was intended to first test the strategy, to verify the outcomes with and without it in different areas.

The possibility to undertake such comparison was limited, owing to inadequate datasets, particularly for comparable gorilla populations outside the Protected Areas. The main quantifiable datasets were available through project sources and therefore mainly enabled a straight before/after comparison to be made of trends in the gorilla population. However, there were also historical data on the national gorilla population, of which those

in the project area accounted for around one-half, which could provide some comparative information. These showed a very clear historical trend nationally of a rapidly declining population of this species prior to implementation of the project. Then the key linkages were tested to verify whether the means-ends relationships held true in the field situation. Next, the key assumptions made by the project were tested to see whether they proved realistic, or at the opposite extreme, whether any were sufficiently far from the case to seriously damage the project's achievements (so-called 'killer assumptions').

The evaluations of the two other projects went through similar analysis and were also able to draw well-founded conclusions on whether and how the projects had 'worked' in terms of progress towards their intended long-term impacts. It was clear that the approach produced a credible analysis of what had been achieved through GEF assistance and how. Six main conclusions were derived from the evaluations and presented to the GEF Council:

- There are measurable and recorded improvements in the status of two key threatened species in Bwindi-Mgahinga and Lewa (the mountain gorilla and black rhino, respectively).
- Two of the three protected area projects have contributed to a sustained reduction in threats to key conservation targets.
- The third protected area project has not been able to continue its threat-reduction mechanisms effectively after GEF support ended.
- Impacts were achieved in two of the three protected area projects because an explicit plan for institutional continuity was built into the projects from the start.
- The Bwindi-Mgahinga and Lewa Projects have both contributed toward substantial additional global environment benefits through catalytic effects.
- The Bwindi-Mgahinga Project has not yet satisfactorily resolved some unanticipated negative consequences on the Batwa indigenous people, whose way of live has been changed by their exclusion from the Protected Areas.
- The last conclusion was underscored by a separate case study undertaken to specifically study the effects of the Protected Area on the Batwa (Namara 2007).

6. Lessons learned from the pilot impact evaluations

Although the theory of change approach proved highly effective, some challenges were also encountered. In the first place, projects are rarely systematically constructed around a theory of change designed to fulfil the necessary means-end relationships to accomplish their intended objectives. They are rather organised around the logical frameworks (log-frames) that almost all donor and implementing agencies require. These log-frames revolve around the practical accomplishments the project is expected to deliver, and not around the underlying development processes necessary to drive change. Monitoring is therefore also focused on the elements of the log-frame, so that in many cases insufficient information is gathered on the underlying theory of change elements. This meant that the approach needed to be adapted to make maximum use of available data and fieldwork findings to build a convincing evaluation of impacts partially or fully achieved. The modified impact evaluation approach developed consisted of three distinct analytical elements:

- *Project logical framework analysis*. This focuses on the delivery of the project outcomes listed in the log-frame and draws substantially on the terminal evaluation (or

its equivalent) of the project, as well as the review of this terminal evaluation by the GEF Evaluation Office or the independent evaluation office of the implementing GEF agency. It usually measures the effects, or behavioural changes the project delivers, rather than ultimate project impacts. For many early GEF projects a retrospective log-frame had to be constructed, since the original project proposal documents did not contain one. Log-frame analysis is relatively weak in understanding the underlying processes driving change and in addressing issues of sustainability and replication, which are fundamental to the understanding of impact.

- *Targets-threat analysis.* This is used in addition to or as a proxy for the direct measurement of intended impacts. The approach draws on the Nature Conservancy's Conservation Action Planning process. It is usually the case with GEF interventions that the intended impact on the global environment will not be realised until several years after formal project closure. However, since the means of achieving this impact is the reduction or removal of threats to the project's target (for example, by removing the poaching of gorillas), then the achievement of threat reduction is a key indicator, which can be used to detect and verify progress towards the impact. Furthermore, in many instances removal of the threat would be an impact in itself, since this will effectively solve the problem in the longer run. An example is the removal of ozone-depleting substances (ODS), which will through natural processes in the upper atmosphere lead to restoration of the ozone layer.
- *Outcomes to impacts theory of change analysis.* This key step is used to map and validate the linkages between project outcomes and ultimate project impacts. It provides a new method for understanding whether project outcomes (assessed by the logical framework analysis) are leading towards the ultimate impacts (assessed through the targets-threats analysis). Systematic validation of the results chain linkages outlined in the theory of change and of their underlying assumptions, as shown in Figure 4, provides indirect evidence of progress towards impacts and can also address measures taken to ensure sustainability as well as replication.

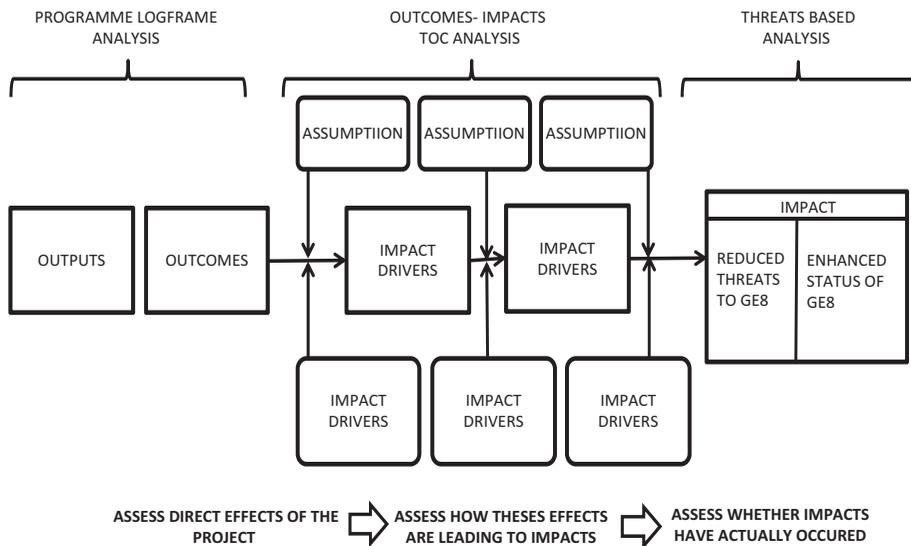


Figure 4. Revised theory-based impact evaluation framework.

7. Building on the pilot results: impacts of GEF support for the phase out of ozone-depleting substances

The next step in the development of the GEF Evaluation Office's approach to impact evaluations was to apply it to a completely different area of GEF support, to see whether it was as effective there as it had been for protected areas (see GEFE0 2009a–c). This stream of activities was in the area of the phase out of ODS. The strategic objective of GEF Ozone Layer Depletion Focal Area is to protect human health and the environment by assisting countries in phasing out the consumption and production, and in preventing releases, of ODS while enabling alternative technologies and practices according to countries' commitments under the Montreal Protocol. The expected long-term impact of the GEF interventions is to contribute to the return of the ozone layer to pre-1980 ozone levels, which is expected by 2065.

The GEF focuses on providing support to Countries with Economies in Transition (CEITs) that are signatories to the Montreal Protocol but are not eligible for funding under the Multilateral Fund of the Montreal Protocol, which targets only developing countries. Since the early 1990s, the GEF has allocated nearly US\$183 million to 18 countries, through 21 national and five regional projects.

The overall objective of the Impact Evaluation was to evaluate the impact of the GEF finance to its portfolio of ozone projects on the phase out of ODS in CEITs. The evaluation combined three approaches to investigate impact from several perspectives, using a mix of quantitative and qualitative methods of data collection and analysis: an overall theory of change approach; in-depth field case studies to assess whether the theory of change analysis had accurately described the process; and before and after measures of ODS consumption and production in CEITs for an internal comparison among the GEF-supported countries, as well as an external comparison with a matched sample of Multilateral Fund-supported countries, to compare ODS consumption and production and cost-effectiveness of the phase-out process.

The theory of change approach was based on an initial meta-analysis of GEF ODS strategies, project documentation and available evaluations. The majority of the projects lacked a log-frame as they were developed between 10 and 15 years ago, when log-frame analysis was not a GEF requirement. Consultations were then held with the GEF Secretariat, Implementing Agency staff, Evaluation Offices, national government stakeholders and enterprises. The function of the consultation was to provide an opportunity for stakeholders to give inputs at an early stage prior to the theory of change being applied and tested in the field case study approach.

In-depth case studies were conducted in four CEITs: the Russian Federation, Ukraine, Kazakhstan and Uzbekistan. A further 10 field case studies were conducted as part of the parallel United Nations Development Programme–United Nations Environment Programme terminal evaluations, which addressed similar issues in the other Eastern European, Baltic and Central Asian countries. Four countries were examined through desk review alone.

The Impact Evaluation drew six major conclusions:

- GEF support for the phase out of consumption and production of ODS in countries with economies in transition has made a contribution to global environmental benefits.
- Legislative and policy changes supporting ODS phase-out provided a foundation for success and ensured sustainability.

- The private-sector commitment to ODS phase-out was a critical driver for the success of the GEF investments in countries with economies in transition.
- Illegal trade threatens to undermine gains in ODS reduction in the non-European Union countries with economies in transition.
- Halon recovery and banking has been neglected in the non-European Union countries with economies in transition.
- In some countries the National Ozone Units ceased to function after GEF support ended, and this may prevent measures being put in place to address the remaining threats to the ozone layer.

8. Quasi-experimental impact evaluations

The impact evaluations on Protected Areas and on the ODS portfolio used a variety of actual or proxy measures to establish quantitative with/without evidence. In both cases challenges were posed by missing data and the imprecise nature of comparisons which could be drawn. For example, on ODS the comparison between countries with economies in transition supported by the GEF and developing countries receiving support from the Multilateral Fund of the Montreal Protocol was found to be very loose because both the nature of the problems addressed and of the support provided differed too much to allow for robust statistical comparison. The likelihood that establishing with/without comparisons would prove difficult was recognised at an early stage, and the Evaluation Office therefore embarked upon a separate strand of work to undertake quasi-experimental impact evaluations, with strong support from the Science and Technology Advisory Panel of the GEF. Since the data in areas that were being investigated did not permit specific counterfactual analysis between GEF project areas and non-GEF areas, the evaluators first sought data-rich situations elsewhere in focal areas covered by the GEF, which could provide examples of how with/without comparisons could be established if better data were available.

Given the emphasis on existing datasets, these studies were therefore not specifically focused on GEF interventions, but sought to establish key patterns of results associated with Protected Areas in specific countries. Two studies in Costa Rica (Andam *et al.* 2007, 2009) showed that Protected Areas have over time led to effective reduction in trends of deforestation and that this has also led to aggregate social benefits for communities surrounding these areas. The latter social impact finding suffered from one substantial area of missing data, inherent in the secondary statistical sources on which the analysis was based; namely the absence of time-series data on social inequality. A parallel study of Protected Areas in Thailand (Sims 2008) produced a similar aggregate income finding, but specified on the basis of time-series Gini-coefficient data that social inequality had also increased. This finding is in agreement with many case studies of the social impacts of Protected Areas, including some of those included in an earlier GEF Evaluation Office study of the role of local benefits in GEF-supported projects (GEFEO 2006). Furthermore, a case study undertaken to look at the effects of the Protected Area on indigenous peoples in Uganda (Namara 2007) also revealed that statements on whether local communities receive benefits or detriments need to take account of socio-economic, cultural and ethnic differences among those affected.

An initial exploration of the possibilities and challenges of experimental impact evaluation was undertaken through an examination of some aspects of a GEF payment for environmental services project and a draft report of its findings is available on the GEF website (GEFEO 2009d). Further work is being undertaken on this study to finalise its analysis and conclusions.

9. Broadening the impact concept: development of the Review of Outcomes to Impacts

As a result of the experience in Protected Areas and the start-up phase of the ODS impact evaluation, it was concluded that the approach can be applied to all focal areas covered by the GEF. At the same time the Office was developing its approach to the Fourth Overall Performance Study of the GEF, for which the terms of reference, established by the GEF Council in September 2008, urged a central focus on impacts. The question was then how to introduce elements of the impact evaluation approach into other focal areas without undertaking full-blown impact evaluations.

One element that had emerged both in the evaluation of the impacts of the three Protected Areas in Eastern Africa and in the ODS evaluation was that project documentation tended to have a lot of information on theory of change assumptions and on progress toward impact, without presenting this in a systematic or comprehensive manner. To uncover this material, a new method was developed by the GEF Evaluation Office together with the Conservation Development Centre (2007a–d, 2009) of Kenya. This new method was named the Review of Outcomes to Impacts (ROtI).

Two versions of the ROtI were developed: the desk ROtI, which is quick and cheap, using only available project data, particularly the terminal evaluation of a finished project or its equivalent; and the field ROtI, which requires limited fieldwork and therefore costs more time and money than the desk equivalent, but which provides more detailed information on the level of progress towards impacts. In the Fourth Overall Performance Study, the full-scale impact evaluation on ODS would also contribute to provide insights into whether and how the GEF support helps to achieve impact. The contribution of the three different approaches used by the Evaluation Office is illustrated in Figure 5.

The combination of a limited number of full impact evaluations and a moderate set of field ROtIs, as well as a full complement of desk ROtIs covering the entire cohort of projects, was projected to provide the best available combination of high-quality data

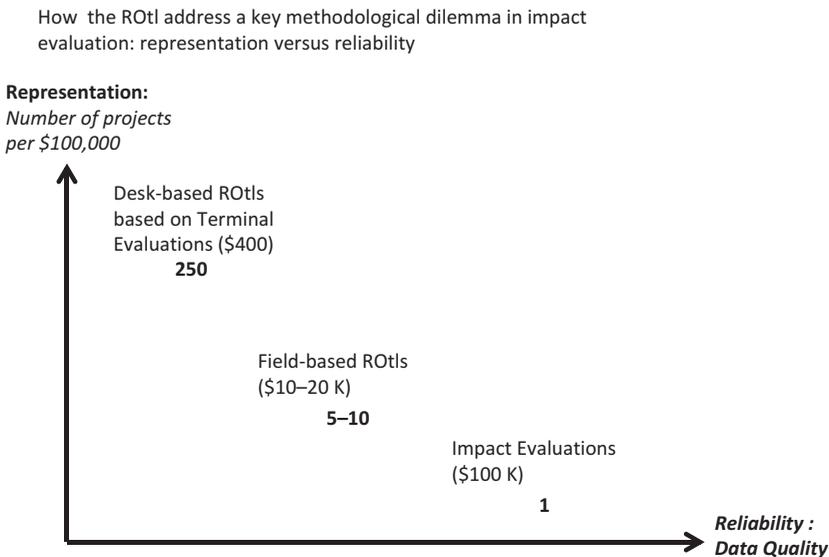


Figure 5. Three impact evaluation approaches: representation versus reliability.

with breadth of coverage. The ROTI methods were developed, piloted and verified as part of the Fourth Overall Performance Study process. They emerge from those developed for the full impact evaluations of the Evaluation Office. They are briefly explored below.

10. Desk and field ROTI methodologies

10.1. The desk ROTI

There are three stages in a desk ROTI: Stage 1, developing the project's Theory of Change on the basis of project documents; Stage 2, assessing progress along the Outcomes to Impacts pathways outlined in the Theory; Stage 3, rating the project. It is important to stress that the ROTI review, although primarily a desk exercise, is based on independent or independently verified evaluations that included field work. In other words: this is not paper on top of paper, but analysis of what objective and independent evaluations of GEF projects have revealed through a systematic framework. Looking at existing evidence from a new perspective allows comparison between achievements and judgments on progress toward impact that was not inferred by terminal evaluations themselves, or not in the same terminology or with the same rigour.

Stage 1 of the desk ROTI, outlining the project's Theory of Change, involves determining the project's intended impacts, verifying the project's logical framework and elaborating the project's intended Outcomes–Impacts pathways. The main data sources are the original project funding document, the Terminal Evaluation and any other readily available documents, such as the GEF Evaluation Office review of the terminal evaluation. The essence of the desk ROTI is to produce a rapid and cost-effective assessment of a large number of projects and any efforts to follow up additional sources would detract from this advantage. The only occasion when it is necessary to spend time locating documents is when one of the fundamental sources, such as the original funding proposal, is not immediately available.

Stage 2, assessing the Outcomes–Impacts pathways, builds on the previous stage and involves exploring progress along each of the key pathways, focusing on the basic elements of the theory, which are presented schematically in Figure 6.

In this stage, the evaluator undertakes the following analysis:

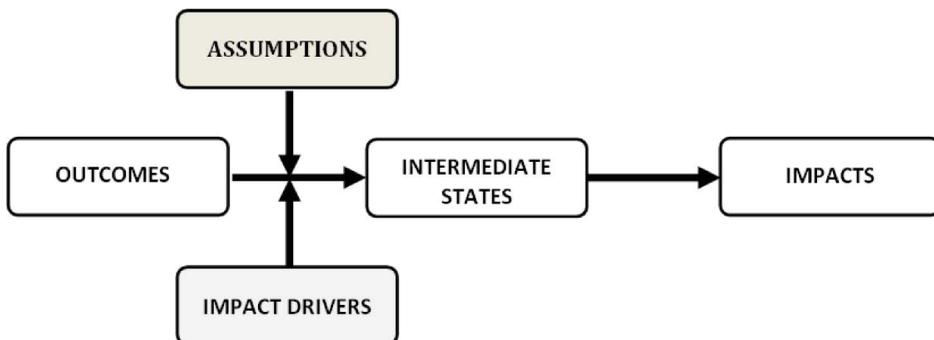


Figure 6. Key elements in a project Theory of Change.

- Assesses measures taken during the project to ensure that Intermediate States are reached.
- Assesses whether Impact Drivers are active by project end, to carry progress beyond the project funding period.
- Assesses whether external assumptions necessary to convert the project’s outcomes to the ultimate impacts are likely to hold.

The ground is then clear to undertake the project impact rating, Stage 3 of the desk ROTI process. There is a simple (at least in principle) rating system to assess project achievements at the level of outcomes, intermediate states (which provide an indication of the likelihood of achieving impacts) and any impacts actually achieved by project completion. It is beyond the scope of this article to explore the rating system in detail, but this is fully described in the ROTI Handbook. The purpose of the rating system is to enable aggregation of assessments of the results of different types of intervention. This takes impact evaluation to a new level, by enabling evidence to be assembled on an entire cohort of projects, in the case of the Fourth Overall Performance Study covering all projects for which terminal evaluations were received from mid-2005 to mid-2009; a total of 205 projects.

10.2. The field ROTI

The field ROTI commences with the same documentary analysis as in a desk ROTI. However, in this case the analysis is used to develop a focused set of questions for the fieldwork component. This systematically examines every link in the results chain to see what has actually happened since project completion. The steps to complete a field ROTI are illustrated in Figure 7.

The field ROTI is most cost-effective when the available set of key stakeholders in the project and its aftermath can be gathered together for a group discussion, which usually

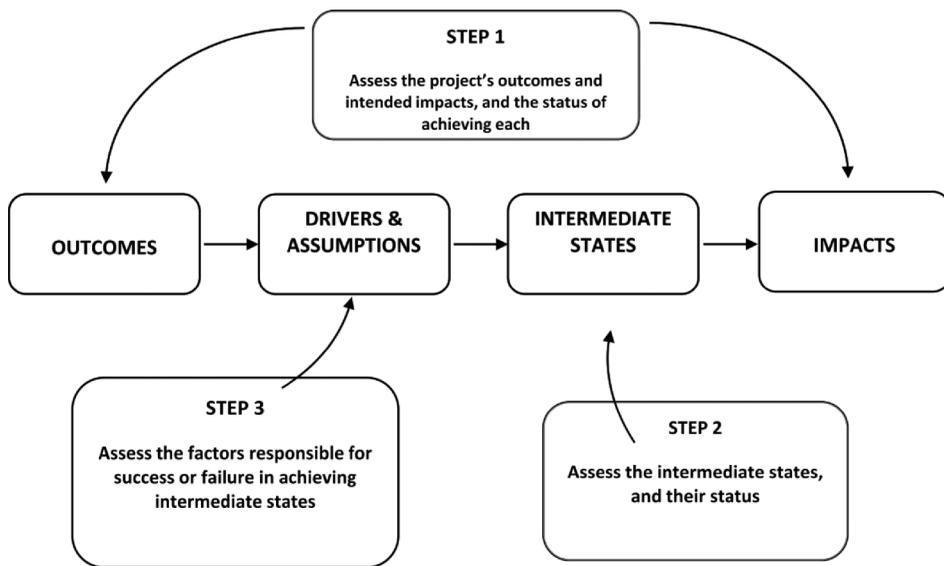


Figure 7. Key steps in the field ROTI exercise.

initially takes from one-half to one day and then requires substantial follow up of the issues and results it raises. The group discussion enables the three steps to be pursued in one session, drawing on a broad range of perspectives. This provides a clear picture of what the project is perceived to have achieved, whether there is a common understanding of this, or whether there is a range of differing perspectives. The discussion group can then be followed up with individual interviews of people unable to attend, or who are flagged as important stakeholders during the course of the discussion. Where time and budget allow, field visits to project-affected sites can be undertaken and additional data on progress since completion is gathered from local sources.

The rating system for the field ROTI is more detailed and precise than that of the desk ROTI, although broadly comparable. Fieldwork enables a closer examination of whether the theories on which the project was implicitly or explicitly designed were correctly conceived and delivered. It is also possible, if the time scale is right, to collect data locally on whether impacts were actually achieved. During the Fourth Overall Performance Study, the two types of ROTIs were conducted as independent exercises; but in more recent developments, the field ROTIs were introduced and incorporated into the Country Portfolio Evaluations that the Office undertakes, thus integrating impact evaluation elements into an overall evaluation of a portfolio of the GEF in a given country.

11. Reporting on impacts: the Fourth Overall Performance Study

11.1. Impact Evaluation results

As indicated above, the Fourth Overall Performance Study presented a substantial body of evaluative evidence on impacts, as part of its overall reporting on the results of GEF activities. The findings of the full Impact Evaluations of three Protected Area Projects and of support for the phase out of ODS were incorporated, as well as the quasi-experimental impact evaluations undertaken on behalf of the Evaluation Office.

The desk ROTI analysis of 205 projects for which terminal evaluations were received by the Evaluation Office in the period of mid-2005 to mid-2009 enabled an aggregate perspective on how well the portfolio was progressing towards achievement of its intended longer-term impacts (see Table 1). Many of the projects analysed had achieved intermediate states that denoted a systemic change, which would over time enable the intended global environmental benefits to appear. Of the 205 rated projects, 80 showed intermediate states that would enable solid progress toward impact, whereas 64 needed further action, and 61 showed no such perspective. Direct and short-term impacts were also demonstrated in all three categories, which could be scaled up to contribute towards Global Environment Benefits. One-third of the projects that currently show no progress from outcomes towards intermediate states, nevertheless have shown that impact could be achieved through the

Table 1. Progress toward impact by numbers of GEF completed projects.

OPS4: 205 desk ROTIs	Number of projects		Initial impact demonstrated	
	Number of projects	Percentage	Number of projects	Percentage
Solid progress (towards ultimate impact)	80	39	44	21
Progress needing further action	64	31	24	12
No progress (towards ultimate impact)	61	30	21	10

Source: Fourth Overall Performance Study (OPS4).

mechanism they employed (new technology, improved management, changes in land or fertiliser use, and so on).

The analysis of impacts in a large cohort of projects gave rise to the consideration of the concept of gradations of impact, which were clearly present. For example, a project that provided solar power to 500 households could be stated to have had an impact on the environment by reducing emissions produced by a given amount of power provided. However, the ultimate impact objective of the project may be to reach all rural households in a region with such a supply: an impact of a completely different scale. The fact that both achievements can be labelled 'impact' is therefore confusing. The ultimate impact objective of the project is the larger one, which requires the initial impact to be scaled up before it can be achieved. So the desk ROTI analysis showed that some projects achieved initial impacts and built upon these to pass through more intermediate states on the way to their intended ultimate impacts, which in the case of the GEF are categorised as Global Environment Benefits. On the other hand, there were many projects that achieved initial impacts, but which made no progress towards scaling these up, so that their intended ultimate impacts are unlikely to be realised. These differences are shown clearly in the overall desk ROTI ratings, as discussed below. We can therefore conclude that the concept of impact does indeed subsume gradations, from 'initial impact' through intermediate states towards the intended 'ultimate impact', which in the case of the GEF is described as a Global Environment Benefit.

In funding amounts, larger projects were found to make better progress towards their intended ultimate impact and to have a higher rate of demonstrated initial impact than smaller projects (see Table 2). However, when we examine projects that have made little or no progress towards their ultimate impact objective, we find that the smaller projects demonstrate more initial impact than their larger counterparts. This finding suggests a hypothesis that these smaller projects do not make progress towards their intended ultimate impact because of lack of funding. This hypothesis could be tested by the Evaluation Office in future impact work.

Field ROTIs, of which nine were undertaken specifically for the Fourth Overall Performance Study, and several others in the framework of the ODS impact evaluation, provided further evidence on the importance of the impact drivers and on the need for further action after project closure, in order to move towards ultimate project impact. Given the fact that neither the GEF nor the GEF agency implementing the project are involved after the project has terminated, any follow-up actions are the responsibility of the countries and local communities concerned.

The full report of the Fourth Overall Performance Study (GEFEO 2010b) contains a full analysis of the desk ROTIs, where possible triangulated with evidence available from field ROTIs, impact evaluations and thematic and country-level evaluations. In this article

Table 2. Progress towards impact by amount of funding of GEF completed projects.

OPS4: 205 desk ROTIs	Amount		Initial impact	
	(US\$ million)	Percentage	demonstrated	Percentage
Solid progress (towards ultimate impact)	385	44	255	30
Progress needing further action	278	33	170	20
No progress (towards ultimate impact)	195	23	25	3

Source: Fourth Overall Performance Study (OPS4).

we focus on some of the results, which emerged from the analysis of the Climate Change and Biodiversity focal areas.

11.2. *Climate Change results*

The Fourth Overall Performance Evaluation concludes that climate change funding has supported a solid level of achievement of progress towards the intended Global Environmental Benefits, both in terms of reduction or avoidance of greenhouse gas emissions and in sustainable market changes. Despite this achievement, the GEF contribution to reduction in greenhouse gas emissions is quite small in comparison with the emissions reductions required at the global level to ensure a more sustainable development path. This issue is explored fully in the Fourth Overall Performance Study (GEFEO 2010b, Chapter 2.1).

Projects that show higher progress toward global environmental benefits demonstrate more specific attention in design and/or implementation to: steps necessary to catalyse government commitment from national to local levels; coherent financial, policy, tariff and/or tax incentives to influence the market; commitment of resources necessary to scale-up project benefits; and measures to generate and encourage lasting commitment of key national stakeholders. Progress toward global environmental benefits also depends on long-term support from governments, the private sector and local communities after the project has terminated.

As shown in Table 3, about 38 per cent of the Climate Change project cohort has made strong progress towards their intended ultimate impact objectives, based on their combined ratings for the ‘targeted outcomes’ achieved, as measured by the ROtI method. At the other performance extreme, 22 per cent of projects have made little or no progress and are therefore considered highly unlikely to achieve their objectives for ultimate impact. Forty per cent of the projects were in the moderate progress range, which indicates that they have produced initial impacts with the possibility of progress toward impact, but have not begun to take the necessary steps to do so. Additional impact drivers will need to be actively engaged to move forward after project closure, but the means and institutions to supply these drivers were not planned or put in place by the project, so the future is uncertain.

When the ratings are placed in the context of funding provided, the picture changes substantially. Here, 55 per cent of the funds were spent on projects with strong progress toward ultimate impact and a further 33 per cent on projects with a medium level of progress. These latter projects need new impact drivers, to progress further, but these were not put in place by the end of project funding and remain in doubt. A very small portion (12%) of funds was expended on projects, which had made no progress towards their intended global environmental benefits by the time of their terminal evaluation.

Table 3. Climate Change portfolio progress toward ultimate impacts/Global Environmental Benefits.

	Little or no progress (%)	Moderate progress (%)	Solid progress (%)
Number of projects	22	40	38
Amount of funding	12	33	55

Source: Fourth Overall Performance Study.

The Fourth Overall Performance Study further found that the GEF has performed better than expected in direct emissions reductions and cost effectiveness for \$/tCO₂ in energy efficiency, has achieved less than one-half of intended emissions in renewable energy, and has partially attained emissions targets for ‘other’ projects. Older projects achieved higher ratings for outcomes, intermediate states and progress towards ultimate impacts. This may be indicative of increasingly complex and ambitious projects after the early years of the GEF; but may also point to the ‘low-hanging fruit’ nature of early projects funded by the GEF. Complex market transformations, which became the focus of much GEF funding after the initial years, are difficult to attain, more so in renewable energy than in energy efficiency. Failure to deliver results, in particular when it comes to market transformation processes, mainly relate to a few key external assumptions and impact drivers.

The analysis of OPS4 results highlighted a number of deficiencies in available data. Although many projects show positive developments in progress toward ultimate impacts and have demonstrated reduced greenhouse gas emissions, there is still a lack of systematic information on types of outcomes achieved and clear indicators with which to measure performance. Also absent are standardised emission reductions’ calculations and reporting at the project and portfolio levels. These performance monitoring systems will need to improve if the GEF wishes to report more accurately on its achievements.

11.3. Biodiversity results

ROtI desk reviews were conducted for 100 biodiversity projects. This represents a major new set of data (based on field-based final evaluations conducted by GEF implementing agencies) on results for the GEF’s biodiversity portfolio over the past four years, which complements several other Evaluation Office studies. The overall assessment of progress of the 100 completed projects from outcomes towards their intended ultimate impacts is shown in Table 4.

Desired intermediate states frequently found in GEF biodiversity projects are the effective management of protected areas, the generation of local benefits from conservation activities, and the reduction of threats through implementation of alternative income-generating activities. Effective management of protected areas can include community-based natural resource management, and sustainable use of resources in appropriately zoned areas. The GEF’s objective to play a catalytic role was found to be a key element of many projects’ strategies. Replication and scaling up can be considered either an impact driver or a desired intermediate state, depending on the anticipated time frame of replication and/or scaling up.

When examining the projects, which have achieved different levels of rating, an underlying question is what makes a project successful or not? A review of key met and unmet

Table 4. Biodiversity portfolio progress toward ultimate impacts/Global Environmental Benefits.

	Little or no progress (%)	Moderate progress (%)	Solid progress (%)
Number of projects	30	30	40
Amount of funding	33	26	41

Source: Fourth Overall Performance Study.

impact drivers and assumptions for each category of project achievement was undertaken and provides useful pointers, although the desk ROTI review methodology can only provide the level of information on individual projects that can be extracted from agency final evaluations.

Biodiversity projects that are highly likely to contribute to ultimate Global Environment Benefits have at least three main achieved impact drivers: stakeholder ownership and support, effective financial mechanisms and adequate information flows. In addition, it can be pointed out that such projects have appropriately addressed issues of scale.

A wide range of key assumptions hold true for successful projects; in other words, there are many risks that do not have to be faced, because of appropriate design and implementation. Successful projects do not always have smooth sailing, however, and the ROTI desk analysis identified some assumptions with which even successful projects must struggle. These are most often assumptions related to genuinely exogenous factors, such as socio-political stability within a country and macro-economic factors such as the relative return on investment of different land-use types, exchange rate fluctuations, and economically driven population flows.

An extension of these factors is the lack of emergence of unforeseen new threats. What makes biodiversity conservation so difficult is the ever-changing nature of any given set of environmental, socio-political, and economic circumstances in a geographic area. New threats can and do appear through the course of a project intervention. Such threats include infrastructure development, changes in global commodity prices that increase pressure on natural resources such as timber or precious metals, or drive agricultural expansion. Among the potentially most significant emerging threats to biodiversity at a global scale is climate change, which could, for example, shift ecosystem boundaries and disrupt the ecological rationale for the current delineations of protected area systems supported by the GEF. For this reason, sustainability of project results must be considered a dynamic state.

Projects shown to be unlikely to contribute to global environment benefits face multiple barriers to achieving impact drivers and meeting their original assumptions, which keep them from demonstrating the progress necessary to trigger a higher rating. Commonly unmet impact drivers include:

- Insufficient technical and institutional capacity.
- Ineffective or inappropriate policy frameworks; for example, related to land-tenure issues.
- Lack of mechanisms for replication/scaling-up, such as dissemination strategies.
- Insufficient financial sustainability; including relying on markets that are not adequately developed, or dependence on government funding, but with a low priority to receive such funds.
- Insufficient stakeholder ownership (ownership can be affected by any one of many potentially relevant stakeholder groups).
- Insufficient information/data to assess whether intended progress is actually being achieved.

In addition to the assumptions made by all projects, such as political stability, projects in the middle category (which are around one-half of all projects) often fall victim of the following unmet assumptions:

- Assumptions related to the linkage of community benefits to conservation results do not hold.

- Lack of existence (and maintenance) of adequate individual technical capacity.
- Inadequacy of intervention (breadth or scale) to address threats; political support or ownership does not materialise or is not maintained.
- Unavailability of financial options, either for community benefits or general sustainability of results.

12. Reflections on Impact Evaluation in the Global Environment Facility

On the basis of the very substantial body of work undertaken as part of the Fourth Overall Performance Study, it was concluded that the three levels (full impact evaluation, field ROTI and desk ROTI) each made a distinctive contribution; and that together they provided a comprehensive overview of impacts achieved and of what level of impacts is likely to be achieved in the future. Full impact evaluations provide detailed and reliable data on a few projects at substantial cost. The Field ROTIs provide moderate scaling-up possibilities, with a medium level of detail based on primary data collected during a limited period of fieldwork. The Desk ROTIs enable a rapid and low-cost method to assess the extent of progress towards intended impacts made by an entire portfolio of projects at closure, on the basis of which the likelihood that the portfolio will generate its intended ultimate impacts can be assessed.

During the course of developing and utilising the combination of methods designed to provide an overview of the impacts to which the GEF has contributed, the Evaluation Office encountered some external confusion as to the evidential base on which conclusions have been drawn. It appears that the full impact evaluations and the field ROTIs are readily understood, while the desk ROTIs are not. It should therefore be re-emphasised that the desk ROTI is, paradoxically, a field-based method. Each desk ROTI is undertaken on the basis of a systematic (using a practitioners' handbook) re-analysis of evidence produced by a field-based Terminal Evaluation undertaken or verified (in the case of the World Bank through a sample-based revisit procedure) by the independent evaluation office of the GEF Implementing Agency concerned. In all cases, therefore, the initial contribution of the GEF to outcomes and intermediate stages towards impact has been assessed and verified by field-based evaluation.

The nature of the impact of GEF-supported projects and interventions needs to be understood in line with the GEF's catalytic nature. The GEF does not intervene on its own, but together with international, national, and local partners. These partners are 'catalysed' through GEF support and continue working toward Global Environmental Benefits after this support has ended. Thus, the GEF *contributes* to the success of a project, but the ultimate impact of the project needs to be largely *attributed* to the partners that continue to work on the issues addressed by the project after GEF support ends. This is illustrated in Figure 8 from the Fourth Overall Performance Study Full Report (GEFEO 2010b). The impact work of the GEF is thus more aligned with the contribution approach as advocated by John Mayne (2001, 2004) and Rob D. van den Berg (2005) than with a narrow approach to impact as defined by Banerjee (2007) and 3IE (2010a), although it has to be noted that 3IE in its guide for grantees (3IE 2010b) fully recognises the need to understand the context in which the intervention is supposed to achieve impact.

The nature of the GEF's contribution was clearly demonstrated in the first impact evaluations undertaken by the Evaluation Office in Eastern Africa, where GEF support and GEF Agency involvement had ended three to five years earlier. In the two more successful projects, local communities, the management and staff of protected areas, the governments of Kenya and Uganda, and new donors continued to work on improving protected area

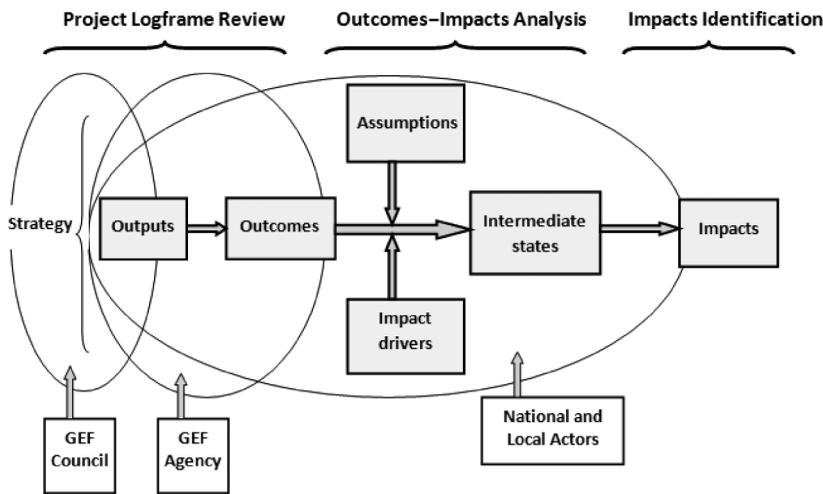


Figure 8. Overlapping contributions (circled) of GEF partners towards impacts: and the GEF Impact Evaluation methods (shown at top) used to assess them.

Note: See also Figures 4 and 6.

Source: Fourth Overall Performance Study (GEFEO 2010b).

management. When a measurable increase of key species in the two protected areas studied was discerned, this could be mainly *attributed* to the ongoing efforts of the partners that remained involved. The GEF was no longer active in the initiatives, but could be shown to have made essential *contributions* to starting up the process toward impacts. Furthermore, the evaluation showed that the sustainability of the impacts achieved would crucially depend on the national and local partners involved and not on the GEF or its Agencies. The evaluation demonstrated that the causal linkages from the project to the impact achieved can be followed and documented and the contribution of the GEF identified. This contribution will always need support from others over time and usually also during the course of GEF assistance, which rarely supplies all necessary inputs even during the project implementation period.

Based on the early experience of randomised trials and quasi-experimental studies in the overall GEF impact evaluation approach, two observations can be made:

- Randomised control trials mainly contribute towards our understanding of initial impacts, which are at a relatively low level. Such initial impacts do not in themselves predict progress toward global environmental benefits, due to the absence of the time and geographical dimensions which are usually necessary to move results from those achieved by project closure towards the intended ultimate impacts, which must be on a larger geographical scale and of long duration.
- In many areas of the GEF, the causal mechanisms in themselves are not in question. For example, the role of excess nutrient flows in rivers creating eutrophication in downstream water bodies has been empirically established, and the causal mechanism that leads to eutrophication has been researched and demonstrated in laboratories. Similarly, renewable energy technologies and technologies with lower greenhouse gas emissions are developed in laboratories under strictly controlled conditions; there is therefore no need to test these technologies through counterfactual evaluations.

Mainly where human behaviour is concerned would counterfactual evaluations be of interest. Such tests would, however, focus on limited elements of GEF strategies and national policies. Moreover, they would be focused on causal attribution, not contribution, and primarily on the role of the partners of the GEF rather than on the GEF itself. Evaluating the ultimate impact of the GEF thus requires a recognition that the final impact of follow-up activities of its support will be attributable to its partners: national governments, local authorities, local communities, industries, farmers, and civil society organisations.

The final issue that the Evaluation Office needs to address is how the results of its impact work are connected to, and in agreement or in conflict with, other evaluative evidence produced by the Office. The Senior Independent Evaluation Advisors to the Fourth Overall Performance Study, Robert Picciotto and Shekhar Singh (2010), noted in their comments that the desk ROtIs need to be connected to the outcome and sustainability ratings that are available in the terminal evaluations of finished GEF projects. These ratings after all, were 'always meant to estimate the likelihood of achieving positive global environmental impacts' (Picciotto and Singh 2010, 197). When all ratings were available, a check was performed to see whether the ROtI ratings on outcomes were consistent with the terminal evaluations on outcomes. It turned out that the ratings are consistent, although they measure slightly different aspects of outcomes. The terminal evaluation outcome ratings focus on the delivery of the project's intended outcomes, whereas the ROtI rates outcomes both on delivery and on the presence or absence of design elements that would enable progress toward the intended ultimate impact. Furthermore, the ratings use different scales. Further methodological development could lead to full consistency between the ratings.

A second comparison was conducted between the 'intermediate states' ratings of ROtI and the ratings for sustainability in terminal evaluations. This check showed more differences, because the perspectives of the ratings are fundamentally different: 'intermediate states' are rated on progress towards ultimate impact, whereas 'sustainability' is more simply concerned with whether outcomes will continue to exist. A comparison of the ratings shows this difference especially in the set of projects that face difficulties in moving towards ultimate impacts. For highly successful projects and projects that are likely to fail, the ratings are relatively consistent, but ratings for projects in between these two extremes are more nuanced in the ROtI system, which scores differences in the likelihood that ultimate impacts will be achieved. The ROtI rating system therefore offers a diagnostic tool on what is needed to get intermediate states moving forward to achieve the intended ultimate impacts.

The various methods and tools used and in some cases specifically developed by the Evaluation Office and its partners to evaluate impact try to address the full range of impact issues, from the causal relationship between an intervention and its effects, to the replication and up-scaling of such effects, and including the understanding of what this means in historical contexts. This more synthetic and multi-methods approach has been advocated by various recent authors, including Ravallion (2005, 2008), Pawson (2006) and Bamberger and White (2007), and highlighted in the guidelines published by the Network of Networks on Impact Evaluation (Leeuw and Vaessen 2009). The Fourth Overall Performance Study shows in principle how these various tools and methods can contribute toward understanding how an agency like the GEF is interacting with its partners to promote impact.

Of these methods, especially the ROtI desk and field reviews offer an approach that could be integrated into the terminal evaluations of project and programme interventions, and are thus within the (financial) scope of many evaluation offices and many agencies. The GEF Evaluation Office hopes to develop a terminal evaluation approach that would

include a review of progress toward impact without substantially increasing the budget of these evaluations.

To conclude, we believe that the GEF Evaluation Office has introduced new methods that enlarge the scope of impact evaluations, beyond a narrow concern for causal relationships at the micro level (which in itself is worthy of attention), toward a better understanding of how multiple strands of interventions and activities, and contributions from different partners, can lead to systemic change in the environment and how we interact with the environment, to achieve longer-term impacts on the global environment. The GEF Evaluation Office will continue to work on these issues and is now undertaking a full-scale impact evaluation of the GEF's support to improve the environmental status of the South China Sea.

References

- Andam, K.J., Ferraro, P.J., Pfaff, A.S.P. and Sanchez-Azofeifa, G.A., 2007. *Protected areas and avoided deforestation: a statistical evaluation*. Impact Evaluation Information Document No. 11. Washington, DC: GEFEO.
- Andam, K.J., Ferraro, P.J. and Holland, N., 2009. *Applying quasi-experimental methods to evaluate the social impacts of protected areas*. Impact Evaluation Information Document No. 13. Washington, DC: GEFEO.
- Bamberger, M. and White, H., 2007. Using strong evaluation designs in developing countries: experience and challenges. *Journal of multidisciplinary evaluation*, 4 (8), 58–73.
- Banerjee, A.V., ed., 2007. *Making aid work*. Cambridge, MA: The MIT Press, A Boston Review Book.
- Chen, H.-T., 1990. *Theory-driven evaluations*. Newbury Park, CA: Sage Publications.
- Conservation Development Centre, 2007a. *Case study methodology*. Impact Evaluation Information Document No. 6. Washington, DC: GEFEO.
- Conservation Development Centre, 2007b. *Case study: Bwindi Impenetrable National Park and Mgahinga Gorilla National Park Conservation Project*. Impact Evaluation Information Document No. 7. Washington, DC: GEFEO.
- Conservation Development Centre, 2007c. *Case study: Lewa Wildlife Conservancy Project*. Impact Evaluation Information Document No. 8. Washington, DC: GEFEO.
- Conservation Development Centre, 2007d. *Case study: Reducing Biodiversity Loss at Cross-Border Sites in East Africa Project*. Impact Evaluation Information Document No. 9. Washington, DC: GEFEO.
- Conservation Development Centre, 2009. *Towards enhancing the impacts of environmental projects: the ROI handbook*. Washington, DC: GEFEO.
- Foundations of Success, 2007. *Final report on a proposed approach to GEF impact evaluation*. Impact Evaluation Information Document No. 2. Washington, DC: GEFEO.
- Global Environment Facility Evaluation Office, 2006. *The role of local benefits in global environmental programs*. Evaluation Reports 30. Washington, DC: GEFEO.
- Global Environment Facility Evaluation Office, 2007. *GEF annual impact report 2007*. Evaluation Reports 46. Washington, DC: GEFEO.
- Global Environment Facility Evaluation Office, 2008a. *GEF annual impact report 2008*. Evaluation Reports 48. Washington, DC: GEFEO.
- Global Environment Facility Evaluation Office, 2008b. *GEF protected area projects in East Africa*. Impact Evaluation Information Document No. 12. Washington, DC: GEFEO.
- Global Environment Facility Evaluation Office, 2009a. *GEF annual impact report 2009*. Washington, DC: GEFEO, GEF/ME/C.36/2. Available from: www.gefeo.org [Accessed July 2010].
- Global Environment Facility Evaluation Office, 2009b. *GEF impact evaluation of the phase-out of ozone-depleting substances in countries with economies in transition: volume one: theory of change*. Impact Evaluation Information Document No. 17. Washington, DC: GEFEO.
- Global Environment Facility Evaluation Office, 2009c. *GEF impact evaluation of the phase-out of ozone-depleting substances in countries with economies in transition: volume two: country reports*. Impact Evaluation Information Document No. 18. Washington, DC: GEFEO.

- Global Environment Facility Evaluation Office, 2009d. *(RISEMP): assessing the potential for experimental evaluation of intervention effects: the case of the regional Integrated Silvopastoral Approaches to Ecosystem Management Project*. Impact Evaluation Information Document No. 15. Washington, DC: GEFEO.
- Global Environment Facility Evaluation Office, 2010a. *Progress toward impact: Fourth Overall Performance Study of the GEF: executive version*. Washington, DC: GEFEO.
- Global Environment Facility Evaluation Office, 2010b. *Progress toward impact: Fourth Overall Performance Study of the GEF: full report*. Washington, DC: GEFEO. Available from: www.gefeo.org [Accessed July 2010].
- Leeuw, F. and Vaessen, J., 2009. *Impact evaluations and development: NONIE Guidance on Impact Evaluation*. Washington, DC: NONIE.
- Mayne, J., 2001. Addressing attribution through contribution analysis: using performance measures sensibly. *The Canadian journal of program evaluation*, 16 (1), 1–24.
- Mayne, J., 2004. Reporting on outcomes: setting performance expectations and telling performance stories. *The Canadian journal of program evaluation*, 19 (1), 31–60.
- Nair, D., 2007. *Priorities and indicators for global environmental benefits from biodiversity: the current international architecture*. Impact Evaluation Information Document No. 5. Washington, DC: GEFEO.
- Namara, A., 2007. *Case study: impacts of creation and implementation of national parks and of support to Batwa on their livelihoods, well-being and use of forest products*. Impact Evaluation Information Document No. 10. Washington, DC: GEFEO.
- OECD Development Assistance Committee, 2002. *Glossary of key terms in evaluation and results based management*. Paris: OECD.
- Pawson, R., 2006. *Evidence-based policy: a realist perspective*. London: Sage Publications.
- Pawson, R. and Tilley, N., 1997. *Realistic evaluation*. London: Sage Publications.
- Picciotto, R. and Singh, S., 2010. Comments of the senior independent evaluation advisors. In: *Fourth Overall Performance Study: full report*. Washington, DC: GEFEO, 194–199.
- Ravallion, M., 2005. *Evaluating anti-poverty programs*. Washington, DC: World Bank, Policy Research Working Paper 3625.
- Ravallion, M., 2008. *Evaluation in the practice of development*. Washington, DC: World Bank, Policy Research Working Paper 4547.
- Sims, K., 2008. *Evaluating the local socio-economic impacts of protected areas: a system level comparison group approach*. Impact Evaluation Information Document No. 14. Washington, DC: GEFEO.
- 3IE, 2010a. *Improving lives through impact evaluation: a strategy for 3IE*. New Delhi: 3IE. Available from: www.3ieimpact.org/strategy [Accessed July 2010].
- 3IE, 2010b. *3IE impact evaluation practice: a guide for grantees*. Available from: www.3ieimpact.org/strategy [Accessed July 2010].
- Todd, D. and Brann, J., 2007. *GEF impact evaluation: approach paper*. Impact Evaluation Information Document No. 1. Washington, DC: GEFEO.
- van den Berg, R.D., 2005. Results evaluation and impact assessment in development co-operation. *Evaluation*, 11 (1), 27–36.
- Vaessen, J. and Todd, D., 2007. *Methodological challenges in impact evaluation: the case of the global environment facility*. Discussion Paper 2007.01. Antwerp: IOB.
- Vaessen, J. and van den Berg, R., 2010. Evaluative evidence in multi-level interventions: the case of the global environment facility. In: O. Rieper, F.L. Leeuw and T. Ling, eds. *The evidence book*. New Brunswick, NJ: Transaction Publishers, 67–94.
- White, H., 2009. *Some reflections on current debates in impact evaluation*. 3ie Working Paper 1. Available from: www.3ieimpact.org [Accessed July 2010].