

Evaluating Environment and Development: Lessons from International Cooperation

Juha I. Uitto

Evaluation Office, United Nations Development Programme (UNDP), USA

Abstract

Environmental evaluation is discussed in the context of national development in developing countries. The paper makes a case for evaluation to move beyond evaluating individual interventions to assessing the contributions to sustainable development at the national and international levels. It highlights challenges relating to evaluating environment in national development, as well as aggregation and attribution of results from programmes aimed for demonstration and policy influence. The discussion is focused around two concrete cases involving programmes by UNDP and the Global Environment Facility.

Keywords

evaluation, international development, sustainability, environment, attribution, results aggregation

Introduction

Questions related to environment and development remain topical today more than two decades after the Rio de Janeiro Earth Summit¹. Both national and international actors in governmental and nongovernmental fields, as well as in academia, are searching for insights into how sustainable development can be advanced and environmental concerns incorporated into the development agenda more effectively. There is an environment-poverty nexus at the heart of sustainable development that is still often neglected in development endeavours. Environment as a global public good tends to get short shrift, as in the short term it is seen as an externality and there are perceived trade-offs between economic development and environmental protection. Although these trade-offs in reality often are false, this perception lingers. The economies of virtually all countries, but in particular poorer ones, rely on sustainable management of the environment and natural resources. Similarly, poor people tend to be more dependent on the environment for their livelihoods and vulnerable to changes in it than their richer compatriots. There is evaluative evidence that when these interlinkages between the environment and poverty are recognized, both benefit (EO, 2010). While development programmes often ignore the environment, environmental programmes also tend to operate in parallel of mainstream economic development.

Evaluation can play an important role in demonstrating results and analysing what works and why. However, to do so, evaluation must move beyond assessing individual interventions in isolation and contribute to the understanding of how environmental concerns can be better incorporated into development efforts in the national context. This does not imply abandoning

project evaluation, which plays an important role for accountability and improvement of specific interventions, but it is always important to keep the big picture in mind.

This article seeks to illustrate the complex interlinkages between environment and development and the implications for evaluation at the programmatic level. It aims to illustrate the challenges to be addressed through two real-life cases drawn from international programmes, which have both been designed to integrate environmental and developmental considerations. The challenges are partly specific to evaluating programmes that focus exclusively on environmental objectives, but are exacerbated in evaluating programmes in which the environment is not the only target. The cases also demonstrate issues pertaining to efforts to evaluate environment and development work beyond individual projects. Particular challenges related to measuring and attributing the success of integration of environmental concerns into national development strategies to programmes whose goal is primarily demonstrating good practices and to influence policy are highlighted. Another particular challenge identified relates to aggregation of results from a large number of projects or sub-programmes to national and global levels.

The first part of the article discusses in general terms related to environmental evaluation in light of recent literature and how these affect evaluation of the environment in development programmes. Next, the article makes a case for the need to move to a higher level where evaluations are conducted against national development trends instead of focusing narrowly on individual programmes and projects. The article then presents two specific cases drawing upon evaluations conducted by the Evaluation Offices of the United Nations Development Programme (UNDP) of the work of the organization at the country level and the Small Grants Programme (SGP) of the Global Environment Facility (GEF). These cases illustrate concretely the issues raised by the author. The final part of the article moves towards a conclusion with an attempt to stimulate discussion in the evaluation community concerned with evaluating sustainable development at the national and global levels.

Challenges to Tackle

Particularities of Environmental Evaluation

As in other fields, there is an increasing demand for demonstrating impacts and learning what works and why in the field of environment. Large investments are made in programmes and projects addressing local, national and global environmental problems as a primary target or in the context of addressing economic and social development. Yet, the measurement of results and evidence-based decision-making tend to lag behind for several reasons. Some of these reasons may be political and ideological (e.g., based on different valuations of environment and economic development by different groups of stakeholders), while lack of transparency and inadequate sharing of data may sometimes play a role (Keene and Pullin, 2011). Most often, it would seem that the reasons pertain to methodological challenges in measuring results in complex situations (Forss, Marra and Schwartz, 2011).

The close interlinkages between environmental and development issues, complicated by significant differences as outlined below, clearly pose specific challenges to the evaluation community. There are certain evaluation challenges that are specific to the environmental area. Some of these stem from the particular characteristics of environmental issues, as well as the programmes and projects designed to address them. Birnbaum and Mickwitz (2009: 3-4) categorize them as follows:

1. Differing and frequently long time horizons;
2. Disparities in scaling;
3. Data quality and credibility issues;
4. The problem of research designs for assessing attribution in environmental policies and programmes.

The different time horizons for observing changes in natural and social systems, as well as the often short policy and programme cycles, pose challenges to evaluators. Similarly, the spatial scales of environmental problems—such as deforestation or climate change—often do not match the scales of political jurisdictions making evaluation at appropriate scales challenging. Data availability and credibility challenges relate partly to the above. Long time series data are regularly not available for the environment, in particular at the local level. On the other hand, given that environmental units often do not coincide geographically with jurisdictional ones, data where they exist are often fragmented. It has been recognized that due to a combination of these factors, environmental programmes in complex situations do not lend them easily to experimental or quasi-experimental designs (Vaessen, 2011). Recognizing these issues, Rowe (2012) has suggested a two-system evaluand approach that takes into account the different temporal and spatial frames of the human and natural systems.

Nevertheless, rigour in evaluations can be enhanced through the formulation of theories of change, including causal hypotheses with explicit assumptions, and making observations regarding baselines and key indicators (Ferraro, 2009; Vaessen and Todd, 2007). Focusing specifically on evaluation of environmental conservation programmes, a team of researchers makes a case for rigorous design alternatives using qualitative, non-experimental approaches (Margoluis et al., 2009).

Climate change poses new challenges to evaluation, including shifts in the objects of evaluation, new metrics, and greater focus on risk, uncertainty and complexity (Picciotto, 2009). Both mitigation of climate change and adaptation to its impacts and reducing people's vulnerability, especially in the developing countries, are closely related to economic development. Their evaluation requires different approaches (van den Berg and Feinstein, 2009). Systematic evaluation can and should play a critical role in identifying what approaches are most effective and efficient in terms of costs and benefits to address them (White, 2009). This will require the building of an adequate body of evaluative evidence that can be used to synthesize lessons.

A stocktaking conducted a few years ago concluded that while the evaluation community was starting to react to the specific issues pertaining to climate change adaptation, monitoring and

evaluation frameworks had not been systematically applied thus far. Similarly, very few ex-post evaluations have been undertaken, although their role for learning, accountability and transparency is acknowledged (Feinstein, 2009). Evaluation is needed to bring new knowledge to this emerging field. Denton (2009) has argued that evaluation (especially when it is participatory) can help build the knowledge and capacities of communities to adapt and to reduce their vulnerabilities.

Evaluation – Beyond Projects

Apart from the particular challenges of evaluating environmental programmes and projects, there is a strong argument for moving beyond evaluation of individual interventions. In this scenario, environmental evaluation becomes part of a broader discussion about development effectiveness. Picciotto (2007) argues that the main unit of account for evaluation should shift from the project level to the country level with new metrics derived from the Millennium Development Goals. He further calls for making development evaluation a country-based and country-owned process instead of a donor imposition. As indicated in the introduction, it is this author's view that elevating the focus of evaluation is not a substitute to evaluating individual interventions, but important for framing the questions in terms of whether such interventions make adequate contributions at the macro level. For international organizations supporting countries in their own development efforts, evaluation plays a central role in assessing the relevance and results of the programmes and projects to national development. At the same time, these organizations have a responsibility to contribute to broader goals beyond individual countries, as well, and sometimes there are tensions between local, national and global priorities. The global environmental goals are a case in point.

By the same token, it has been convincingly argued that evaluation must also address the issue that has been termed the 'micro-macro paradox' in economic literature, i.e. that while many individual interventions seem to succeed in reaching their objectives, there is little to show by way of solving the larger problems at the national or global levels. Global environmental degradation and climate change fall squarely into this category. An earlier evaluation of GEF work at the local level concluded that in many projects where the global-local linkages were not adequately addressed, this resulted in reduced benefits at both levels (GEF, 2006: 7).

Van den Berg (2011) argues that evaluation has the responsibility of moving beyond assessing the success of individual interventions to look at the big picture and contribute to the broader understanding of the reasons for the micro-macro paradox, which often lie in the realm of treating the environment as global public goods and, thus, as an externality to economic development. He suggests fine tuning how we define 'relevance' and 'impact' as evaluation criteria to ensure that the relevance of any intervention includes whether it actually made a difference in addressing the larger issue. Similarly, evaluation should not only address the direct impact of an intervention but its 'final or ultimate' impact. Obviously, defined in such a way, relevance and impact are closely related to each other and the ultimate question whether the intervention being evaluated has actually contributed to changing the negative trends in society, economy and environment. However, there is still a distinction between the two

criteria and at the intervention level relevance can be evaluated when full impact is not yet determinable.

Recognizing the micro-macro paradox, the Evaluation Office of UNDP already a decade ago reassessed its approach to country programme evaluations. It became obvious to evaluators that, in some cases, evaluations reported that UNDP programmes and projects were fully on track in achieving their designed results, while all the macro indicators in the country in question showed a declining trend. In response, the Evaluation Office modified its approach to evaluate UNDP's contributions to national development results and benchmarking such contributions against the overall trends in national development. While moving the benchmark to a higher level from the performance of individual projects and programmes was a clear improvement, the country level evaluations have faced other problems, not least the difficulty of attributing the achievement of results to UNDP's programmes. A particular challenge has been to assess the organization's contributions in the environmental field in relation to the overall country programme in which environmental concerns and projects feature to a varying degree. Examples that follow will highlight specific challenges related to this issue.

Evaluation of Environment in International Development: Two Examples

This section highlights two particular cases drawing from the work of the Evaluation Office of UNDP. The first focuses on evaluating when the environmental goals are not the primary objective of the intervention, but are seen as important for their interconnectedness with other development objectives at the country level. The second addresses the challenge of aggregating results from a number of small interventions towards a programmatic goal. In both cases, issues relating to the nature of the results that are being evaluated, and their ownership, are raised.

Evaluating Environment in UNDP's Programming at Country Level

As the development network of the United Nations, UNDP aims to help "build nations that can withstand crisis, and drive and sustain the kind of growth that improves the quality of life for everyone"². Specific challenges faced by UNDP relate to integrating its environmental work with the broader development framework and poverty reduction strategies that are its main focus and mandate (Stewart, Uitto and Wells, 2009). Tensions between these objectives in UNDP programming were highlighted in a global evaluation conducted by the Evaluation Office (EO, 2008). It was found that these tensions were frequently caused by conflicting institutional incentives and organizational structures. These stemmed from and reflected tensions in institutional alignment and accountabilities, in particular between contributing to national development goals and achieving global environmental objectives.

In particular, the tendency by UNDP country programmes to substitute environmental work with direct linkages to national and local development issues in the developing countries, such as sustainable management of natural resources or local energy development, with projects funded by the Global Environment Facility was found to have the effect of distancing

environmental programming from the poverty reduction focus. The GEF mandate emphasising global environmental issues is complementary to UNDP's explicit poverty focus. As an implementing agency for projects funded by the GEF, UNDP is expected to mainstream global environmental concerns into the national development strategies and programmes that are its primary concern. While the funding for projects focusing on global environmental issues, such as biodiversity conservation or climate change mitigation, is valuable and necessary, the lack of funding for addressing local environmental priority issues was found to form a gap that was often difficult to fill.

The presence of this gap in integration poses particular challenges for evaluating UNDP's work at the country level. While poverty eradication is an overarching goal for UNDP, environment and sustainable development features as one of its four focus areas, based on the understanding of the close interlinkages between poverty, natural resources management, environment, and energy. UNDP also recognizes that the poor people in particular depend on natural resources and the environment for their day-to-day lives.

There is a well-established and functioning system for evaluating GEF-funded projects in UNDP, as well as in other agencies that implement projects on behalf of the Facility. Terminal evaluations are mandatory for all projects and follow a well-defined evaluation protocol. Understandably, these evaluations assess the achievement of the specific goals and objectives that the projects were set up for; mostly focusing on contributing to the global environmental benefits that are the mandate of the GEF. The question then is whether this is adequate in the UNDP context, as the organization's mandate is explicitly to support countries in reaching their own nationally defined development results. It would not appear appropriate to judge the performance of GEF-funded projects in isolation from UNDP's core mandate, especially as they form the bulk of the organization's environmental programming.

Apart from the project evaluations, country-level and other broader programmatic evaluations attempt to take an integrated look at UNDP's country-level programming in which all aspects of UNDP's work in a country—including policy advice, advocacy, capacity development, project work and other types of support—are assessed utilizing the evaluation criteria of relevance, effectiveness, efficiency and sustainability against the national development goals. Country-level evaluations have documented a number of cases where environmental projects have been successful in their own sphere, but where they have been designed and implemented on a parallel track separately from the rest of the country programme³.

In other cases, there have been missed opportunities for addressing priority environmental issues with linkages to national and local development and poverty reduction, because the availability of external funding has directed country programme priorities to different priorities and funding has not been available from UNDP's own resources to fill the gaps. Such was the case, for example, in India where opportunistic programming based on fund availability had led to a fragmentation of the country programme (EO, 2012). This situation has in many countries been by the internal organizational structures promoting silos in which programme staff working in the poverty reduction, democratic governance, environment and sustainable

development, and crisis prevention and recovery areas do not adequately share information or cooperate with each other (EO, 2010).

These examples do not by any means represent the entire picture of UNDP work and there is evidence that the importance of mainstreaming environment into overall programming has gained increasing recognition, at least in part due to the series of evaluations highlighting the issue. Yet, they do demonstrate the point that it is not sufficient to focus on evaluating individual projects—however significant many GEF-funded projects may be—in isolation of the broader goals that they are intended to contribute to.

It is also important to be explicit about what we evaluate against. In the case of UNDP and the Global Environment Facility, there is a close relationship and cooperation between the two entities, but their goals and mandates are not identical even though both are very important and worthy. It is therefore not adequate to evaluate GEF-funded projects implemented by UNDP only in relation to how well they contribute to the GEF goals of generating global environmental benefits. The expected value added by UNDP goes beyond being a competent implementing agency for the GEF projects and should be manifested in the complementarities between the global, national and local benefits achieved through integrated programming.

Small Grants Programme: Aggregation Challenges

The Small Grants Programme, launched in 1992 and implemented by UNDP on behalf of all the GEF partners, provides direct support to local communities to take action in the areas of biodiversity conservation and sustainable use, climate change, international waters, land degradation and sustainable forest management, and chemicals management. Over its two first decades of operation, SGP has funded over 17,000 projects in more than 120 countries. The direct funding from GEF for the projects has amounted to some US\$380 million, with over US\$500 million in co-financing. With such scale of funding, there is an obvious interest to determine the results of the programme on the ground. From an evaluation point of view, several interrelated challenges arise.

In 2007-2008, a major joint evaluation by the GEF and UNDP evaluation offices attempted to take a comprehensive look at SGP and its results. The joint evaluation emphasized that the previous three evaluations of SGP “were primarily oriented toward improving SGP operations and design and toward distilling lessons” and were “not able to assess which global environmental benefits had been achieved, whether the programme was cost effective, or whether there were trade-offs between the SGP and other GEF projects” (GEF, 2008: 5). The joint evaluation consequently set out to assess the SGP’s impacts in achieving global environmental benefits, its effectiveness and efficiency, and the factors affecting the results. The joint evaluation engaged a large team of evaluators collecting both qualitative and quantitative information on the programme performance and results at programmatic and national levels. The evaluation included twenty randomly selected country case studies. The joint evaluation took as one of its starting points the GEF mandate to achieve global

environmental benefits, while recognizing SGP's role in promoting people's well-being and livelihoods.

There has been a long-standing debate about the relationship between the global and the local in GEF programming. The Facility was originally established in the context of the 1992 Earth Summit to finance the 'incremental cost' of national actions to make them beneficial for the global environment beyond the strictly defined national interests of the developing countries. In its initial phases, GEF could only finance projects limited to the preservation of globally significant biodiversity, mitigation of climate change, phase-out of ozone depleting substances, and management of international waters. Since the early stages of GEF, the policy has evolved and a broader understanding of the importance of linking the global and national concerns has emerged. Another joint evaluation of GEF partners looked specifically at the role of local benefits in achieving global environmental benefits and concluded that the local and global benefits were closely interlinked (GEF, 2006). GEF also broadened the range of activities it could finance, but the focus on global environmental benefits remained.

The question of local versus global is particularly important in the context of the Small Grants Programme. While all of SGP's thousands of projects must fall within the GEF focal areas, it can be argued that not all of them are explicitly focused on addressing global environmental issues—or if they are, the direct impacts on issues such as reduction of greenhouse gas emissions are quite small at the level of individual projects. The mode of operation of the programme is based on community empowerment, participation and capacity development. The programme aims to help communities to move towards sustainable development through small projects that encourage the conservation and sustainable use of environment and natural resources. Replication and scaling up of successful small-scale initiatives are explicit objectives of SGP, which also aims at influencing policy at the national level by bringing such successful approaches to the focus of partnership with governments and nongovernmental actors.

The governing bodies and funders of the GEF Small Grants Programme have stated the need for aggregating the impact of the programme globally. The question is: what level of attention should be paid to the direct global environmental benefits generated by the individual projects as opposed to other, intermediate results? While the programme as a whole has developed an elaborate monitoring and evaluation system—and there are periodic independent evaluations built into the system at the time of each replenishment of the SGP fund—the individual projects or even country-level programmes are hardly ever evaluated. Instead, there is a cumbersome self-reporting system against the GEF focal area results frameworks that stretches the capacity of most national and local stakeholders. One can also question the feasibility of aggregating direct global environmental results from thousands of small projects most with budgets in the US\$20,000-35,000 range.

At a certain level, it may be possible to add up the areas protected for biodiversity conservation or the amounts of deforestation avoided. Evidence from evaluations suggests that SGP projects have allowed communities to move towards sustainable use of their resources and to embark on economic development activities that benefit both the communities and the environment.

At the country level, the programme strategy varies considerably. In a few countries, the strategy has been to concentrate the programme geographically or thematically in an effort to maximize anticipated impact (such a case would be Mexico where the programme operates only in the Yucatan peninsula). Most others have adopted an inclusive approach that approves any well-developed proposal following project eligibility criteria in any geographic or thematic area put forth by a civil society organization. Making the leap from these activities to, for example, conservation of global biodiversity makes several assumptions, not least that the communities operate in areas with globally significant biodiversity that would otherwise be threatened. In addition, aggregating protection of hectares of areas in highly varied types of ecosystems would be difficult, if not meaningless from a scientific point of view. Short of conducting extensive surveys with built in counterfactuals of levels of endangered species or ecosystems—hardly feasible given the vast geographic scope and the number of projects—it would seem impossible to determine the impacts of the programme on actual biodiversity conservation.

Aggregating only the direct environmental benefits of the individual projects may also significantly underestimate the value of the programme. The stated mission of the Small Grants Programme puts enhancing the wellbeing and livelihoods of especially poor and vulnerable communities at its centre⁴. It would appear fair to include the results of replication and scaling up of SGP projects and approaches into the valuation of its results, given that these objectives are explicitly built into the programme approach. This would give a better sense of the overall impact SGP has had beyond its direct investment. In this case, evaluation will face a challenge of attribution. First, while catalytic and learning dimensions of SGP could and should be captured by evaluations, by definition the actual scaling up or replication will be done by others. Consequently, the programme cannot be held accountable for the results achieved by activities beyond its control. Secondly, can we be certain that the expanded activities followed from SGP and would not have taken place without it? Establishing attribution and counterfactual would at the least require extensive and rigorous research in the field, which often will not be feasible under the circumstances. From the point of view of GEF and its partners, establishing a credible case for SGP contribution would seem adequate.

Arguably, the programme's most significant impact could be achieved through policy influence, which is another factor central to the approach. If SGP has been able to influence national policy through demonstrating the viability of environmentally sound practices and approaches that provide significant social and economic benefits to communities, this will have multiplier effects that may even reach beyond the country in question. Needless to say, attribution challenges become harder the further one is distanced from the direct activities, which take place at the community level. Furthermore, SGP (and UNDP managing the programme) operates in many countries where governments are sceptical about the civil society and efforts to promote inclusive development are faced with many obstacles. Environment is often seen as politically less sensitive than programmes that directly address challenges of democratic governance. There are cases where UNDP country-level leadership has reported using SGP as an entry point for discussion about inclusive governance and the role of civil society with

national governments that have not embraced these aspects before. These cases can be seen as positive and, at least partly, unintended results of SGP, albeit difficult to measure.

The evaluation objectives focused on assessing four dimensions (GEF, 2008: 23), namely the:

- relevance of SGP results to the GEF mandate and operations, and to country sustainable development and environmental priorities;
- effectiveness of the SGP in generating global environmental benefits;
- efficiency of the SGP in engaging community-based and nongovernmental organizations to address global environmental concerns;
- key factors affecting SGP results; and,
- monitoring and evaluation systems.

As evident from the above list, the emphasis was placed squarely on the global environmental dimensions. Even with regard to the assessment of the engagement with the civil society organizations, the focus was on their ability to deliver global environmental benefits. The evaluation did review 22 country programmes and as part of these reviews conclusions were drawn regarding replication, scaling up, and mainstreaming local community activities. It was also found that all country programmes contributed to local policy instruments, such as municipal environmental ordinances. More than half of the country programmes contributed to national policy formulation (GEF, 2008: 7). Other conclusions pertaining to contributions to local livelihoods and targeting poverty were generally positive. All in all, however, these social, economic and institutional results of the programme were not systematically pursued, as the evaluation focused primarily on the global environmental mandate of the programme.

At the time of this writing, a new joint evaluation of the Small Grants Programme is underway (2013). A decision was made to largely focus the first phase of the evaluation on updating the previous evaluation. A second phase in 2014 intend to take a more systematic look at the results pertaining to replication, scaling up and policy influence of the programme.

Discussion

Evaluating the effects of environmental interventions is particularly challenging, given the complex context in which environmental programmes and projects operate, and the multiple intervening factors that affect results (Vaessen and Todd, 2008). The further beyond the direct intervention results evaluations want to move, the harder it becomes to determine attribution. This should not mean that it is impossible to conduct rigorous evaluations, but this may be more helpfully done by constructing careful theories of change that can be used to explain the contributions of the intervention.

It is important to clarify what we mean by results and impact. It is possible, and often meaningful, to define these concepts at different levels, ranging from the direct results that can be attributed to the intervention, to the next level of results that may reduce the environmental stressors, to the final ecosystem impact. Such approaches have been applied to

monitoring and evaluation of interventions in practice, for example to forest management or the management of transboundary water resources, and have been referred to variably as 'pressure—state—response' or 'process—stress reduction—environmental status' indicators (see, e.g., Wolfslehner and Vacik, 2008; Uitto, 2004).

In the cases described above, the definitions of results and impact focus largely in the middle ground and contain important dimensions of social, economic, organizational and institutional changes. Such results may be the primary target (such as of UNDP country programmes that have mainstreaming of the principles of sustainable development into national development plans as a goal) or they can be important instrumentalities to achieve environmental benefits (as is the case of the GEF Small Grants Programme). Either way, it is important that evaluations consider and capture them in a systematic way. If we accept the definition of 'relevance' suggested by van den Berg (2011), the relevance of interventions would then be evaluated against whether these changes lead to the ultimate impact related to environmental improvement and increased well-being, not just whether the activities have been aligned with national or global priorities.

Aggregation is particularly challenging in cases where an organization operates in a decentralized manner in many areas or countries. Programmes and approaches need to be adapted to local contexts in order to be effective. While programmes and projects in the various countries and situations may be designed to address the same goal—be it biodiversity conservation, climate change mitigation or any other topic—they seldom should follow exactly the same format. It is possible to aggregate outputs (such as number of government official trained, hectares placed under protective status, or number of environmental laws enacted), but aggregating the broader results and outcomes becomes challenging.

Aggregation can be facilitated by establishing a core set of common indicators at programme design and monitoring them throughout implementation. This is the approach taken both by UNDP in its attempt to aggregate the results of its work on the ground in more than 170 countries, as well as by the Small Grants Programme attempting to aggregate the results of some 17,000 small projects in over 120 countries. The results have been mixed and the challenges distinct. In the UNDP case, aggregation has mainly focused on activities and outputs based on self-reporting by the country offices. In the Small Grants Programme case, aggregation has mainly been against the GEF focal area results frameworks related to expected environmental impacts. Both efforts are based on self-reporting by country teams and in both cases a challenge relates to the diversity of programmes, which does not easily lend itself to reporting on actual results and impacts.

Aggregating the overall environmental impacts is much harder and will require long-term research that takes into account the variations in the local context. In some cases the impacts may be impossible to quantify meaningfully. For example, how does one add up the value of saving a specific hectare of savannah ecosystem versus saving the same amount of cloud forest?

Evaluation literature recognizes these challenges and has come up with potential solutions, which could be usefully applied to either of the above cases. They include approaches and methodologies towards synthesizing results from disparate interventions to make sense of their higher-level effects. Such methodologies include meta-analysis, systematic review and realist synthesis (see, Duvendack et al., 2012; Mallett et al., 2012; Pawson et al., 2004). Some issues to be dealt with in applying these approaches in the context of international development and environment programmes, such as those discussed in this article, pertain to availability and quality of evaluative evidence. The Evaluation Office of UNDP has recently embarked on a realist synthesis of its country-level evaluations conducted over the past decade. The long timeframe over which the evaluations were conducted and the evolving methodology pose some problems of comparability. Similarly, the evaluations cover less than a third of the countries in which UNDP operates. In the Small Grants Programme case the challenges would be even greater due to the lack of evaluations at the project or country programme level. Yet, it would seem worthwhile to explore these approaches further and to adapt them to the real-life situations.

Conclusions

In an era of dwindling resources and an increased emphasis on managing for development results, evaluation must play a central role in demonstrating results and providing evidence of what works and why. Evaluation can contribute to bringing about improved integration and mainstreaming of environmental concerns into development programmes through an evidence-based analysis of the results and by providing explanations to the state of affairs. It is safe to claim that the two independent thematic evaluations conducted on UNDP's environment and sustainable development work and its relationship to poverty programming generated important discussion in the organization and its executive board on how to better address the poverty-environment nexus.

It is important to elevate evaluation so that it can contribute to assessing the ultimate relevance and impact of international programmes in environment and development to the problems of national and global significance. Credible, independent and transparent evaluations can make important contributions to the democratic debate on the value of such programmes. The role of evaluation is both to provide accountability for the intended results and the resources used, as well as to enhance our understanding of what works, why and under which circumstances. These roles are closely interrelated. Only through a credible and independent analysis can we arrive at a judgement of the results achieved and the factors contributing to them, which in turn will allow us to learn from past experiences. Learning from the past and improving future performance is key to achieving greater accountability.

The evaluation community must also have a role in educating the public, including the tax payers in the rich countries who fund much of the environment and development work in the developing countries, of what is possible and meaningful to measure. Often the critical outcomes to assess are at the intermediate level (i.e. beyond the outputs of individual interventions). It may be difficult to attribute the final impact on the environment or people's

wellbeing to a specific intervention. These impacts must, however, always be the ultimate benchmark against which we evaluate. To achieve positive changes in the environmental status, organizational and policy outcomes are essential. The two global evaluations of UNDP work in environmental management, energy and poverty linkages, as well as other evaluative evidence from the country level, showed that achieving effective integration of environment and development requires that institutional incentives, sanctions and indicators to track progress be put in place. At the country-level, there is a need for advocacy to promote such integration.

Environmental concerns should be brought to feature in national development strategies. Evaluations have also documented successful cases that can be disseminated and provide lessons across countries. Evaluations should help policy-makers to think about alternative strategies and policies that do not treat the environment as a mere externality to economic development and can contribute to sustainability at the local, national and global levels.

Acknowledgements

The article has benefited from useful comments on various versions by Osvaldo Feinstein, Michael Reynolds and Lee Alexander Risby, as well as three anonymous reviewers engaged by the journal. Discussions with Delfin Ganapin and Sulan Chen on evaluating the SGP are acknowledged with gratitude. While the article draws upon evaluations conducted by the Evaluation Office of UNDP, the opinions and interpretations are those of the author and do not reflect an official position of the Evaluation Office. The author was the evaluation manager of both of the global thematic evaluations of UNDP's environment and sustainable development work, as well as the India country-level evaluation.

Funding

The evaluations that the article draws upon were funded through the regular resources of the Evaluation Office of UNDP.

Notes

1. United Nations Conference on Environment and Development (UNCED), Rio de Janeiro, Brazil, 1992.
2. http://www.undp.org/content/undp/en/home/operations/about_us/ (downloaded 5 October 2013).
3. All country-level Assessments of Development Results are publicly available on the Evaluation Office website (<http://web.undp.org/evaluation>).
4. <http://sgp.undp.org> (downloaded 5 October 2013).

References

- Berg, R.D. van den (2011) Evaluation in the context of global public goods. *Evaluation* 17 (4): 405-415.
- Berg, R.D. van den and Feinstein, O. (eds.) (2009) *Evaluating Climate Change and Development*. World Bank Series on Development, Volume 8. New Brunswick and London: Transaction Publishers.
- Birnbaum, M. and Mickwitz, P. (eds.) (2009) *Environmental Program and Policy Evaluation: Addressing Methodological Challenges*. *New Directions for Evaluation* Number 122. San Francisco: Jossey-Bass.
- Denton, F. (2009) Challenges for evaluating adaptation to climate change within the context of Africa. In: Berg, R.D. van den and Feinstein, O. (eds.) *Evaluating Climate Change and Development*. World Bank Series on Development, Volume 8. New Brunswick and London: Transaction Publishers, 115-134.
- Duvendack, M., Garcia Hombrados, J., Palmer-Jones, R. and Waddington, H. (2012). Assessing 'what works' in international development: meta-analysis for sophisticated dummies. *Journal of Development Effectiveness* 4(3); 456-471.
- EO [Evaluation Office] (2008) *Evaluation of the Role and Contribution of UNDP in Environment and Energy*. New York: Evaluation Office of UNDP.
- EO (2010) *Evaluation of UNDP Contribution to Environmental Management for Poverty Reduction: The Poverty-Environment Nexus*. New York: Evaluation Office of UNDP.
- EO (2012) *Assessment of Development Results: India – Evaluation of UNDP Contribution*. New York: Evaluation Office of UNDP.
- Feinstein, O. (2009) Monitoring and evaluation of adaptation to climate change interventions. In: Berg, R.D. van den and Feinstein, O. (eds.) *Evaluating Climate Change and Development*. World Bank Series on Development, Volume 8. New Brunswick and London: Transaction Publishers, 237-239.
- Ferraro, P.J. (2009) Counterfactual thinking and impact evaluation in environmental policy. In: Birnbaum, M. and Mickwitz, P. (eds.) *Environmental Program and Policy Evaluation: Addressing Methodological Challenges*. *New Directions for Evaluation* Number 122. San Francisco: Jossey-Bass, 75-84.
- Forss, K., Marra, M. and Schwartz, R. (eds.) (2011) *Evaluating the Complex: Attribution, Contribution, and Beyond*. Comparative Policy Evaluation Volume 18. New Brunswick and London: Transaction Publishers.
- GEF (2006) *Evaluation of the Role of Local Benefits in Global Environmental Programs*. Washington, DC: Evaluation Office of Global Environment Facility.
- GEF (2008) *Joint Evaluation of the Small Grants Programme*. Washington, DC and New York: Evaluation Offices of Global Environment Facility and UNDP.
- Keene, M. and Pullin, A.S. (2011) Realizing an effectiveness revolution in environmental management. *Journal of Environmental Management* 92, 2130-2135.
- Mallett, R., Hagen-Zanker, J., Slater, R. and Duvendack, M. (2012). The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness* 4(3): 445-455).

- Margoluis, R., Stem, C., Salafsky, N. and Brown, M. (2009) Design alternatives for evaluating the impact of conservation projects. In: Birnbaum, M. and Mickwitz, P. (eds.) *Environmental Program and Policy Evaluation: Addressing Methodological Challenges. New Directions for Evaluation* Number 122. San Francisco: Jossey-Bass, 85-96.
- Pawson, R., Greenhalgh T., Harvey, G. and Walshe, K. (2004). *Realist Synthesis: An Introduction*. RMP Methods Paper 2/2004. ESRC Research Methods Programme, University of Manchester.
- Picciotto, R. (2007) The new environment for development evaluation. *American Journal of Evaluation*, 28(4), 509-521.
- Picciotto, R. (2009) Evaluating climate change and development. In: Berg, R.D. van den and Feinstein, O. (eds.) *Evaluating Climate Change and Development*. World Bank Series on Development, Volume 8. New Brunswick and London: Transaction Publishers, 19-45.
- Rowe, A. (2012) Evaluation of natural resource interventions. *American Journal of Evaluation*, 33(3), 382-392.
- Stewart, H.M, Uitto, J.I. and Wells, M.P. (2009) Contributing to global benefits or supporting local sustainable development: Evidence from a global evaluation of UNDP's program. In: Berg, R.D. van den and Feinstein, O. (eds.) *Evaluating Climate Change and Development*. World Bank Series on Development, Volume 8. New Brunswick and London: Transaction Publishers, 151-168.
- Uitto, J.I. (2004) Multi-country cooperation around shared waters: Role of monitoring and evaluation. *Global Environmental Change* 14, 5-14.
- Vaessen, J. (2011) Challenges in impact evaluation of development interventions: Randomized experiments and complexity. In: Forss, K., Marra, M. and Schwartz, R. (eds.) *Evaluating the Complex: Attribution, Contribution, and Beyond*. Comparative Policy Evaluation Volume 18. New Brunswick and London: Transaction Publishers, 283-313.
- Vaessen, J. and Todd, D. (2008) Methodological challenges of evaluating the impact of the Global Environment Facility's biodiversity program. *Evaluation and Program Planning* 31, 231-240.
- White, H. (2009) Using impact evaluation to increase the efficacy of climate change interventions. In: Berg, R.D. van den and Feinstein, O. (eds.) *Evaluating Climate Change and Development*. World Bank Series on Development, Volume 8. New Brunswick and London: Transaction Publishers, 105-113.
- Wolfslehner, B. and Vacik, H. (2008) Evaluating sustainable forest management strategies with the analytic network process in a pressure-state-response framework. *Journal of Environmental Management* 88, 1-10.

Juha I. Uitto is Deputy Director of the Evaluation Office of the United Nations Development Programme. Over the past dozen years, he has worked on a large number of programmatic and thematic evaluations with UNDP and the Global Environment Facility. He has previously worked as researcher and research manager at the United Nations University and the Nordic Africa Institute, and as a consultant in international development. He was educated at the Universities of Helsinki and Lund, and holds a PhD in Social and Economic Geography.