Article

American Journal of Evaluation
1-19

DOI: 10.1177/1098214019865936
journals.sagepub.com/home/aje
SAGE

A Joint Evaluation With Lessons for the Sustainable Development Goals Era: The Joint GEF-UNDP Evaluation of the Small Grants Programme

Carlo Carugi[1] and Heather Bryant[2]

Abstract

The integrated nature of the Sustainable Development Goals (SDGs) calls for greater synergy, harmonization, and complementarity in development work. This is to be reflected in evaluation. Despite a long and diversified history spanning over almost three decades, joint evaluations have fallen out of fashion. Evaluators tend to shy away from joint evaluations because of timeliness, institutional and organizational differences, and personal preferences. As the SDGs call for more joint evaluations, we need to get them right. This article supports the appeal for more joint evaluations in the SDGs era by learning from the existing long and diversified experience. This article shares lessons from a joint evaluation that is relevant in the context of the SDGs for the United Nations Evaluation Group, the Evaluation Cooperation Group, and the wider international evaluation community.

Keywords

evaluation in international settings, joint evaluation, sustainable, development goals, evaluation practice, horizontal evaluation

The 2030 Agenda for Sustainable Development, adopted in 2015, recognizes that the multiple aspects of development expressed in its 17 Sustainable Development Goals (SDGs) and 169 targets are integrated and that the three dimensions of sustainable development—the economic, the social, and the environmental—are indivisible (United Nations General Assembly [UNGA], 2015). Development interventions are being implemented, and evaluations conducted in contexts that are recognized as increasingly complex, with multiple partners involved in multiple network configurations, where harmonization and complementarity are ever more critical. This complexity has implications for evaluation: Evaluations need to recognize the interconnectedness of multiple development

[1] Global Environment Facility, Washington, DC, USA
[2] United Nations Development Programme, New York, NY, USA

Corresponding Author:
Carlo Carugi, Global Environment Facility, 1818 H St., NW, Washington, DC 20433, USA.
Email: ccarugi@thegef.org

objectives, analyze potential synergies and trade-offs between goals and interventions designed to contribute thereto, and integrate the principles of the SDGs, such as "no one left behind."

Joint evaluations, where partners come together to combine resources and complementary perspectives, are one way to strengthen evaluation in the context of the SDGs. Joint evaluations are not new: Their advantages and challenges have been the subject of several analyses and guide over the past decades (Feinstein & Ingram, 2003; Organization for Economic Cooperation and Development: Development Assistance Committee [OECD-DAC], n.d., 2005, 2013; Swedish Agency for Development Evaluation [SADEV], 2008; UN Evaluation Group [UNEG], 2014). However, data suggest that there is still resistance to undertaking joint evaluations given their high transaction costs and the difficulties in addressing the requirements and information needs of the different potential partners. While the challenges of joint evaluations are real, the SDGs call for a renewed investment in joint evaluations and investigation into ways to effectively overcome the barriers, so that the benefits might be fully realized.

The Independent Evaluation Offices (IEOs) of the Global Environment Facility (GEF) and the UN Development Programme (UNDP) conducted an evaluation of the Small Grants Programme (SGP) between 2013 and 2015. The SGP is a program that develops community-led and community-owned strategies and technologies for reducing threats to the global environment while addressing livelihood challenges. Although carried out prior to the formal adoption of the SDGs, this evaluation provides an example of a joint undertaking that integrated analysis of the economic and social with the environmental and therefore provides several lessons of interest for evaluation in the SDG context.

This article briefly traces the history of joint evaluations to highlight the initial rise in interest about 30 years ago by the evaluation community and donors, explained by a number of widely perceived benefits such as the potential for broader scope than single-agency evaluations, increased multidimensionality in the analysis, and a single set of coherent messages addressed to multiple partners, among others. This was followed by a progressive decline in popularity due to many challenges experienced by those involved in implementing them. This article discusses a renewed impetus for joint evaluations brought about by the global commitment to the SDGs and then presents the experience and lessons of the Joint GEF-UNDP Evaluation of the SGP. Building on this experience, we have identified ways to overcome the implementation challenges. We describe how, despite these challenges, joint evaluations offer significant advantages that can lead to important institutional and strategic changes with a view to strengthening development results in today's SDG context. We believe our experience offers insights that are applicable beyond development evaluation. These can be of interest to other evaluation settings such as horizontal (multidepartmental) evaluations (Montague & Teather, 2007) as well as to efforts to move evaluation toward public interest matters (Chelimsky, 2014).

History of Joint Evaluations

While evaluation is arguably as old as the human race itself, dating, as Shadish and Luellen (2005) argue, "from the time humans first made a judgment about whether building campfires and using weapons helped them to survive" (p. 183), systematic evaluation research is a 20th-century development. Evaluation of social programs first became common in education and public health, with, for example, assessments of literacy and occupational training programs in the first decades of the 20th century. Following World War II, with the independence of former colonies and the expansion of social and international development programs, development evaluation began to emerge as a distinct field. In 1960, the OECD and the OECD-DAC were created. Evaluation has been a major theme in the OECD-DAC's work from the beginning: Evaluation was mentioned in the first DAC Chairman's report in 1962 (OECD, 2013). The late 1960s and 1970s witnessed evaluation efforts led by two Nordic countries, the Netherlands and Germany, and in 1982, a DAC Expert Group on Aid Evaluation was created to strengthen the exchange of information and experience between national evaluation units and to promote and support developing countries' evaluation capacities.

Joint evaluations, defined in an OECD-DAC review of 30 years of development evaluation, as "evaluations involving multiple donors or both donor and recipient countries" became an important topic and were discussed "at almost every meeting" of the DAC evaluation network in the 1980s (OECD-DAC, 2013, p. 17). In the 1990s, a number of "flagship" joint evaluations were carried out; the DAC Network cites among others an evaluation of the World Food Programme in 1994 led by Canada, the Netherlands, and Norway and the evaluation of the international response to conflict and genocide in Rwanda led by Denmark and including several OECD members, UN Agencies, and others (OECD-DAC, 2005). In the late 1990s, the DAC Working Party on Aid Evaluation began focusing on joint sector evaluations in selected developing countries and observed that "in the medium term this will be particularly relevant for monitoring progress towards the achieving the goals of 'Shaping the 21st Century'" (OECD-DAC, 1998, p. 5), an OECD-DAC precursor to the Millennium Development Goals (MDGs). In this light, the Working Party developed several tools to facilitate joint evaluations, including a report on "Donor experience with joint evaluations: A typology and lessons learned" (OECD-DAC, 1998). Several reports on joint evaluations followed: In 2000, the same group published a report on "Effective practices in conducting a multi-donor evaluation" and in 2005 on "Joint evaluations: Recent experiences, lessons learned and options for the future" (OECD-DAC, 2005). In 2008, the SADEV identified "Challenges, opportunities and approaches for increasing joint donor programming of evaluations."

Within the UN, in 2006, the Secretary General convened a panel on UN systemwide coherence, and in his remarks on the panel's report, observed that, "It will be essential that we swiftly modernize and achieve full compatibility on processes for resource planning, human resources, common services and evaluation, as these are important drivers of coherence in the United Nations system" (UNGA, 2006, para. 17). The MDGs and Poverty Reduction Strategy Papers also emphasized the need for coordinated and joint approaches to programming and by extension joint evaluations. In 2010, the UNEG decided to establish a Joint Evaluation Task Force and in 2014 published a "Resource pack on joint evaluations." This document "was a reflection of the fact that a majority of UNEG members were involved in joint evaluation work and that these engagements were growing" (UNEG, 2014, p. 10). The UNEG (2014) Resource Pack defines joint evaluations as

> a joint evaluative effort by more than one entity of a topic of mutual interest, or of a programme or set of activities which are co-financed and implemented, with the degree of "jointness," varying from cooperation in the evaluation process, pooling of resources, to combined reporting. (p. 12)

These various papers on joint evaluations, published over a period of 15 years, enumerate the benefits and the challenges of joint evaluations. Several of them emphasize the importance of joint evaluations in a context of common, multidimensional development goals that can only be achieved through partnerships (OECD-DAC, n.d.) and in particular the MDGs (Feinstein & Ingram, 2003; OECD-DAC, 2005; UNEG, 2014). Joint evaluations enable a broader scope and number of evaluation questions to be addressed given combined resources (OECD-DAC, n.d.) and help facilitate a sense of the bigger picture within which a program or a set of joint activities are situated (UNEG, 2014). Joint evaluation arrangements have the potential to yield higher quality, more credible evaluations (Feinstein & Ingram, 2003; OECD-DAC, n.d., 2005; UNEG, 2014). Analyses, findings, conclusions, and recommendations of joint evaluations are based on broader knowledge and a wider range of inputs and contributions (OECD-DAC, 2005), joint rather than individual scrutiny and quality assessment procedures (UNEG, 2014). There are higher chances that a variety of methods for data collection and analysis will be used, as different agencies may have different types of data available and experiences using different approaches (OECD-DAC, 2005; UNEG, 2014). Similarly, interviews can be conducted with a variety of stakeholders, allowing for better triangulation (UNEG, 2014). Several of these sources emphasize that joint evaluations typically carry more weight and are less easy to ignore as they reflect the views of multiple partners (Feinstein & Ingram, 2003; OECD-DAC, 2005, UNEG, 2014). It is also argued that joint evaluations reduce transaction costs for the evaluand and stakeholders consulted, as the number of missions in a joint evaluation is less if each donor carried out its own evaluation (Feinstein & Ingram, 2003; OECD-DAC, 2005; UNEG, 2014). Joint evaluations also reduce multiple messages and the risk of conveying conflicting evaluation messages to partner countries (OECD-DAC, 2005; UNEG, 2014). Joint evaluations result in a reduced financial burden through cost sharing (Feinstein & Ingram, 2003; OECD-DAC, n.d.). Another benefit, it is also postulated, is that joint evaluations foster mutual capacity development and peer learning, as agencies learn from each other and share evaluation techniques, and partners come to understand better the different perspectives, mandates, and approaches of each of the institutions involved (Feinstein & Ingram, 2003; OECD-DAC, n.d., 2005; UNEG, 2014).

However, these papers also highlight a number of challenges of joint evaluations. Developing comprehensive and yet manageable terms of reference that meet each country's or agency's needs may take time and require compromises (OECD-DAC, n.d.; SADEV, 2008; UNEG, 2014). Joint evaluations are typically characterized by complex management, communications, and governance arrangements and generally require more time to carry out, which in turn may result in being more costly than individual agency evaluations (Feinstein & Ingram, 2003; OECD-DAC, n.d., 2005; SADEV, 2008; UNEG, 2014). Methodologies can also become an area of contention, and reaching agreements on methods may also extend the time frame of the preparatory phase (OECD-DAC, 2005; SADEV, 2008). Other challenges may include power differentials between partners and different needs and accountability requirements of the partner organizations (UNEG, 2014). These challenges can be quite significant, and the OECD-DAC (2013) notes that

> recently there has been a bit of a backlash to joint evaluations because of the high transaction costs [for the evaluation units] often associated with coordinating multiple partners and the increased accountability pressure on member evaluation units to assess the contribution of their country's own development assistance. (p. 24)

Joint Evaluations in the UN Context

As seen earlier, many authors have emphasized joint evaluations as a modality to address common, multidimensional objectives that can only be achieved in partnerships. This is particularly relevant to the UN system: Since the late 1990s, the UN started pursuing more coherence in the operations of its several agencies, programs, and funds through various forms of partnerships at the country level. In theory, a corresponding move toward more partnerships in evaluation would be expected. Data suggest however that, in practice, joint evaluations in the UN context are not that numerous. They are undertaken primarily in the context of a joint program or are evaluations of a country's UN Development Assistance Framework (UNDAF). The UNEG evaluation database (http://www.uneval.org/evaluation/reports) includes 6,777 evaluation reports from 1997 to 2017 from 25 agencies (as of February 22, 2018). The database includes the option of indicating whether an evaluation is joint or not. However, for most evaluations (5,235), no information is provided, limiting the analysis of trends in joint evaluations in the UN system. Of the 185 marked as joint evaluations, the first such evaluation dates from 2006, with a peak number of 51 joint evaluations recorded in 2011, followed by a decreasing trend, with only 4 recorded in 2015 and 1 each in 2016 and 2017. Forty-six percent of the evaluations in the database are UNDP evaluations.

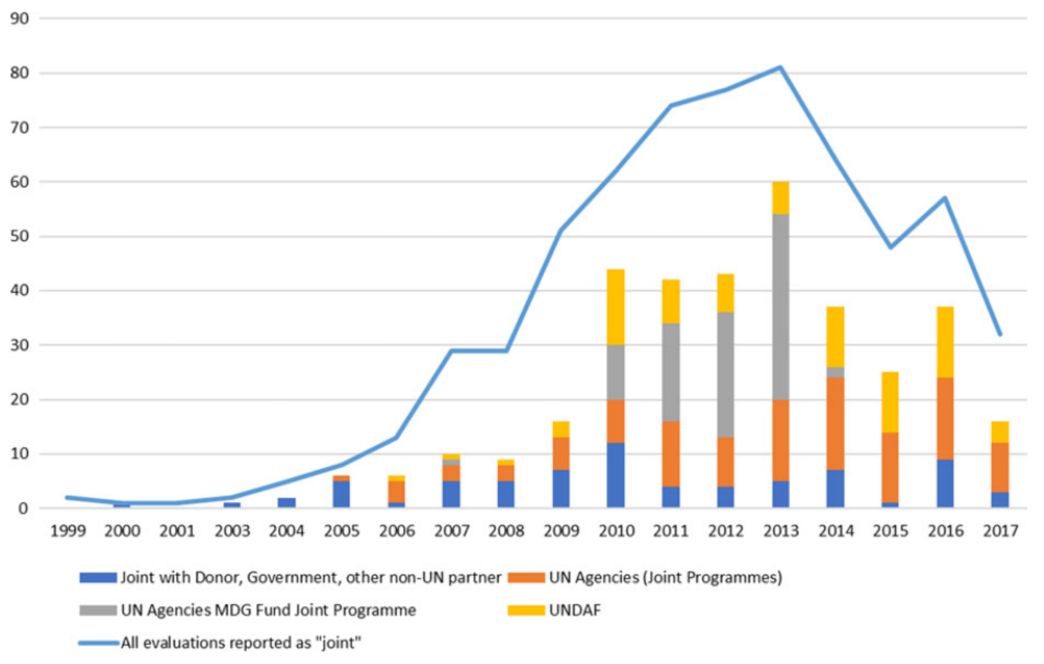

Figure 1. Joint evaluations listed in United Nations Development Programme's Evaluation Resource Center. The bars represent evaluations considered to be "joint" evaluations after review; the blue line represents evaluations reported by the commissioning unit as being "joint."

A more detailed analysis of UNDP evaluations in UNDP's publicly accessible Evaluation Resource Centre (https://erc.undp.org/) provides some additional insights. In UNDP's database, 636 evaluations (of a total 4,201 or 15%) are listed as "joint" evaluations (as of February 7, 2018). An examination of these evaluations shows that 42% of the sample are not truly "joint" evaluations, but, for the most part, traditional UNDP-led project or outcome evaluations, where the national government partners are to be consulted, do not appear to be otherwise engaged in the evaluation process. Thus, approximately 8% of UNDP's evaluations have been joint, which is similar to findings of other analyses (Andersen & Broegaard, 2012). Most of the evaluations actively involving another partner are joint evaluations with other UN Agencies (80% of the evaluations assessed to be joint evaluations). Of these, many are in fact evaluations led and required by the MDG-Fund Secretariat of joint programs (25%) or evaluations of UNDAFs (23%). Only 20% of the UNDP evaluations that were assessed as joint were conducted jointly with a donor, a government counterpart, or other multilateral partner (e.g., the terms of reference clearly state that UNDP and the partner/government were jointly commissioning the evaluation or the partner was clearly mentioned along with UNDP in various steps of the evaluation management process). In terms of trends over time, joint evaluations peaked in 2013, due in part to a peak in MDG-Fund-led final evaluations (Figure 1).

The UNDP data appear to conform to the OECD-DAC observation of a backlash or reduction in the number of joint evaluations, following the upsurge between 2010 and 2013. As can be seen in Figure 1, much of this upsurge was in fact due to the MDG-Fund Secretariat's requirement for evaluations. The data also suggest that despite the potential advantages of joint evaluations, these remain exceptional cases. However, there are broader changes in context that suggest a potential resurgence of interest in joint evaluations. In 2012, the UNGA requested the UN Secretary General

Figure 2. The sustainable development goals. *Source*. United Nations Sustainable Development Platform: https://sustainabledevelop.ment.un.org/?menu=1300.

to develop a policy for independent systemwide evaluation of operational activities for the development of the UN system (UNGA, 2013, para. 181), leading to a review of the systemwide evaluation initiative in 2017 and a proposal for a new independent systemwide evaluation office (UNGA Economic and Social Council, 2018, para. 305). UNEG reflections on the new proposals cite member agencies' experiences with joint evaluation as relevant to and important for building systemwide evaluation capacity (UNEG, 2017).

With the advent of the SDGs and new commitments such as the "funding compact" between Member States and entities of the UN Sustainable Development Group, wherein the latter commits to "accelerating results for countries through more collaboration" including through joint activities "notably in the areas of analytical work, the formulation of operational support and policy options for SDGs acceleration and evaluations" (Lebada, 2019), there is also additional impetus for joint evaluation. The Evaluation Cooperation Group (ECG) (whose members are multilateral development banks) is also giving increased attention to joint evaluations, discussed at the ECG fall 2018 meeting.

The SDGs and Evaluation

The SDGs were introduced in 2015 during the 70th session of the UNGA, convened for the adoption of the post-2015 development agenda (Figure 2). The 17 SDGs identified in the "2030 Agenda for Sustainable Development" build upon the experience of the precursor MDGs. Many papers as well as blogs from nonprofit advocacy groups have summarized the differences between the MDGs and the SDGs, for example, in terms of their formulation and their scope (see Boucher, 2015; Coonrod, 2014; Kumar, Kumar, & Vivekadhish, 2016; Woolbridge, 2015). From the perspective of evaluation, there are also differences. The MDGs emphasized monitoring and tracking progress against 21 targets, and there was little emphasis on evaluation. The Millennium Declaration itself does not mention evaluation and uses the word "review" only once in its penultimate paragraph. An evaluation of monitoring and evaluation of the MDGs found that "in contrast to monitoring, the intergovernmental process for following up on the United Nations Millennium Declaration did not envision an evaluation framework" and "it proved very challenging to access MDG evaluations in a meaningful, systematic or digestible manner" (UN Economic and Social Council, 2015, pp. 20-21). In contrast, the 2030 Agenda includes a whole section on follow-up and review, including evaluation ("follow-up and review processes at all levels will be guided by the following principles: . . . they will be rigorous and based on evidence, informed by country led evaluations . . . " [UNGA, 2015, para. 74]).

Another difference between the MDGs and the SDGs with implications for evaluation is that in the latter, sustainability and the environment are understood as an integral part of development. While the MDGs included the goal to "ensure environmental sustainability," the SDGs consider development and sustainability as inextricable concepts underpinned by an understanding that sees the environment, economy, and society as embedded systems rather than separate competing pillars (Woolbridge, 2015). Implementing the SDGs invites the creation of "an integrated, holistic, multi-stakeholder approach" (Reynolds, Blackmore, Ison, Shah, & Wedlock, 2018). This implies the need for systems thinking in practice, a tradition that draws on systems theories, tools, and techniques able to facilitate better conversation and cooperation between agencies. If systems thinking is applied to SDG implementation, it should be applied to evaluation as well. Joint evaluations could facilitate the necessary multidimensional thinking by bringing together more than one agency—and thus more than one perspective—at the evaluation phase.

Yet another difference between the MDGs and the SDGs is that the MDGs, which emerged from a series of global development conferences in the 1990s, were about the achievement of improved human development outcomes in developing countries (World Health Organization, 2015, p. 193), whereas "all countries and all stakeholders, acting in collaborative partnership, will implement" the SDGs (UNGA, 2015, preamble). MDG Progress Reports were regularly published in developing countries; now voluntary national reviews are produced by developing and developed countries alike. Finland, for example, has identified focus areas to pursue the 2030 Agenda and has produced an evaluation of its sustainable development policy (Lähteenjoja, Berg, & Korhonen-Kurki, 2019). With respect to evaluation, this means that lessons from what has been referred to as "development evaluation" (evaluation of development programs), such as the present joint evaluation, may well have lessons for evaluation more broadly.

With the UN's identification of the 2030 Agenda as a "universal, integrated, transformative, and people-centered" agenda and the recognition of the interlinkages between the dimensions of sustainable development, the UN Development Group (UNDG) presented the principles for integrated programming in the implementation of the SDGs. At the core of these principles is "leave no one behind," underpinning (1) human rights, gender equality, and women's empowerment; (2) sustainability and resilience; and (3) accountability (UNDG, 2016). In its 2016 policy and program brief, the UNDP indicates the need for integrated programming in the implementation of the SDGs, calling for "a need to go beyond silos" (UNDP, 2016, p.6). UNDP also stresses the need for a "people-centered multistakeholder approach" and the operationalization of a strategic plan that mainstreams the three dimensions of sustainable development, namely, the environment, economy, and society. For example, an intersectoral and multistakeholder implementation approach would be expected in water and climate interventions, where the thematic areas themselves call for such an approach (Tosun & Leininger, 2017). The introduction of integrated programming in the implementation of the SDGs calls for the courage to take innovative actions to launch critical new joint evaluations.

A New Impetus for Joint Evaluations

Being both dynamic and complex, the SDGs make results and impacts harder to evaluate (Heider, 2015; Schwandt, Ofir, Lucks, El-Saddick, & D'Errico, 2016; Thomas, 2017; UNEG, 2016). The goals and targets are interdependent and changing over time, and there are inevitable tensions between a triangle of desired outcomes: growth, prosperity for all, and sustainability (Heider, 2015). For example, with respect to Goals 2 (zero hunger) and 3 (good health and well-being), the latter cannot be achieved without access to food. SDG 2 targets increased agricultural production, but how this is done will have a major influence on soil and water quality, land use, and ecosystem health and functioning, which are in turn key determinants of health (International Council for Science, 2017).

To capture the changing and integrated programming around SDGs, evaluation needs to go beyond examining results in a single area and carefully look at trade-offs and synergies between the three dimensions of sustainable development to assess how the different outcomes can be valued. This is not impossible to do. Evaluation of natural resource interventions is complex by definition, as it necessarily involves both human and natural systems but is not necessarily more difficult than evaluation in human systems. Joint knowledge production processes with decision makers and stakeholders increases salience and legitimacy in response to challenging evaluation settings (Rowe, 2012). While international development institutions and partner country governments alike still tend to deal either specifically or predominantly with one of the three sustainable development dimensions, a more useful way to ensure a multidimensional approach to assessment that is suited to the interrelatedness of SDGs' objectives is through joint evaluations.

In addition, and importantly, the concept of "partnership" permeates the entire SDG Agenda 2030. Partnerships are a goal in themselves (Goal 17), products of cooperation and solidarity, and a precondition to tackle the complex and multidisciplinary challenge represented by the SDGs. According to UNEG, complex partnerships and collaborations in implementation will necessarily require "coalitions for evaluation." UNEG also suggests that in the context of the UN support to Agenda 2030 implementation, joining forces among its members is expected to strengthen assessment of the UN contribution to the progress toward the SDGs at national, regional, thematic, and global levels (UNEG, 2016). While the UNEG report may be aspirational, the 2019 "funding compact" mentioned above includes new commitments to joint activities including evaluation. Last but not least, evaluation resources are scarce. Coordination and collaboration in evaluation work across different institutions risks entailing high transaction costs (for the partnering evaluation units). To fully capture the potential for higher cost-effectiveness through joint evaluations, adequate coordination mechanisms should be designed to operate in a light, transparent, and efficient manner.

The above discussion is particularly relevant to the GEF, a financial mechanism for international cooperation, which addresses global environmental concerns while supporting national sustainable development initiatives. The GEF operates through a large partnership involving 183 countries, 18 agencies, and a large network of civil society organizations and private-sector entities. In such a large partnership, a special challenge is posed in adequately monitoring and evaluating the achievements of the GEF, taking into account the activities of all partners in a consistent and coherent manner. Furthermore, as the GEF is moving toward more integrated programming to serve its mandate of achieving global environmental benefits in terms of reduced environmental stress and/or improved environmental status, there is scope for a corresponding move in its evaluation function toward more joint evaluations within its partnership. The recent Joint GEF-UNDP evaluation of the SGP, conducted in partnership by the IEOs of the GEF and UNDP, is the latest in a series of concrete examples of such a move. This evaluation provides examples of both main advantages and challenges highlighted earlier. In line with the SDGs philosophy, it deals with environment and socioeconomic issues interconnectedly.

The 2015 Joint GEF-UNDP Evaluation of the SGP

The GEF created the SGP in 1992 with the aim of developing community-led and community-owned strategies and technologies for reducing threats to the global environment while

Box 1. Joint GEF-UNDP Evaluations of the Small Grants Programme.

> The 2008 joint SGP evaluation was crucial in shaping the way forward for the SGP. Overall, the evaluation found that the SGP is a cost-effective way for the GEF to generate global environmental benefits while addressing country priorities and responding to the needs of local populations. It also found that the SGP management model had reached its limits and was not suitable for a new phase of growth: The SGP governance and audit procedures needed to be strengthened. The 2015 joint SGP evaluation assessed the extent to which the most important recommendations from the 2008 evaluation had been implemented and the extent to which these recommendations and Council decisions remained pertinent. The evaluation also reported on the SGP's current role and results in terms achieving global environmental benefits while addressing livelihoods, poverty, and gender equality; broader adoption of grant-level results; the SGP's strategic positioning; and monitoring and evaluation.

addressing livelihood challenges. Implemented by UNDP, the SGP awards small grants—up to a maximum of $50,000—to needy communities to support the use of practices and technologies that benefit the global environment. Since start-up, the SGP has provided over 18,000 such grants to communities in more than 125 countries. Evaluating the SGP called for an integrated analysis of the program's economic, social, and environmental dimensions made possible by a joint evaluation approach involving the main institutions, the GEF and UNDP. To note, together with the World Bank and the UN Environment Programme, UNDP is one of the three historic Agencies endowed with responsibilities of implementing projects and programs supported by the GEF. It is the largest Agency in terms number of projects and the second largest in terms of GEF funding (GEF IEO, 2017).

The GEF Monitoring and Evaluation (M&E) Policy establishes that monitoring and evaluation is a shared responsibility within the GEF partnership (GEF IEO, 2010). This approach makes full use of the combined capacities of the partnership and the respective comparative advantages of each GEF partner. The multiplicity of stakeholders also places a premium on learning and improvement by continuously sharing knowledge from M&E, both within and among the GEF partners and with external stakeholders.

As an example of such shared responsibility, the SGP has been jointly evaluated by the IEOs of the GEF and UNDP twice, in 2008 (GEF IEO & UNDP IEO, 2008) and in 2015 (GEF IEO & UNDP IEO, 2016a; Box 1).

The IEOs of the GEF and UNDP have conducted several joint evaluations. Since 2006, the GEF IEO has conducted three important joint evaluations with other partners (COWI & International Institute for Environment and Development, 2009; GEF IEO, 2006, 2014). Likewise, the UNDP IEO has been involved in six joint evaluations with various partners (UNDP IEO, 2000, 2002; UNEG, 2008, 2009; United Nations Industrial Development Organization (UNIDO) & UNDP IEO, 2006, UNIDO & UNDP, 2009). In addition, the GEF and UNDP IEOs collaborated in 2008 on the abovementioned first Joint GEF-UNDP Evaluation of the SGP, on four country portfolio studies/country program evaluations (El Salvador and Jamaica in 2011 and East Timor and Sierra Leone in 2013), and in 2016 on the "Biodiversity impact evaluation—Support to protected areas and protected area systems" (GEF IEO & UNDP IEO, 2016b). These collaborative efforts were designed to maximize the complementarity between the two institutions, the GEF assessing the environment aspects with an eye on socioeconomic co-benefits and UNDP focusing on sustainable development while addressing environment and energy dimensions of the programs evaluated. This complementarity is an important enabling condition for exploiting the full potential of joint evaluations, although not a prerequisite.

Lessons on Factors That Enhance Success, Challenges That Arise, and Benefits of Joint Efforts

This decade-long experience of each of the agencies with joint evaluations as well as direct collaboration in several GEF/UNDP joint evaluations facilitated the decision to undertake, and the approaches adopted in, the 2015 joint SGP evaluation by the two institutions. Reflecting on the SGP experience, there are many observations and lessons to share, including factors that lead to successful implementation of a joint evaluation, challenges that arise, and the benefits of joint efforts. These elements are presented below in the approximate order they are encountered in the evaluation process, beginning with the more practical and often challenging aspects of implementing a joint evaluation, such as partnership arrangements, timing, and management structure and then moving to the more substantive aspects including scope, methods, and analysis, where the advantages and benefits of joint evaluations become more apparent. Some of these considerations are relevant to joint evaluations in general and others to joint evaluations in the SDGs context. In addition, we believe that many of the points below are applicable beyond development evaluation and are potentially useful to many evaluation settings.

Deciding on a Joint Evaluation

The decision to carry out a joint evaluation is made within a political or institutional context. In the case of the SGP, the decision to conduct a joint evaluation was straightforward. The GEF funds the SGP, and UNDP implements it, making the two institutions natural partners for a joint evaluation. Both the UNDP and GEF evaluation offices are independent and have considerable scope in determining their work plans. A precedent had been set with the 2008 joint evaluation. For the two institutions, this specific context was conducive to the decision to conduct a joint evaluation.

Partnership Arrangements

The SGP experience suggests that institutions working as equal partners enhance joint evaluation success. It has been observed that there are often power differences at play in joint evaluations, where one agency may end up dominating the evaluation process, because it has invested more funding, has a stronger position in the management structure, or because it has an agenda it wishes to pursue (UNEG, 2014). In the case of the 2008 joint evaluation, the GEF led the evaluation (with the final report credits citing the Director of the GEF Evaluation Office and the Task Manager [from the GEF Evaluation Office]), with UNDP staff mentioned as members of the evaluation team. The 2015 evaluation was carried out by the two offices as equal partners, as illustrated by the management structure (see below) and the final report, which credits both evaluation office directors and the two task managers, one from each office. The different mandates of the two institutions did mean the two had slightly different interests in the evaluation, which did lead to significant debates during the evaluation (see below on the scope of the evaluation), but these were debates between equal partners.

Timing

Following the decision to conduct a joint evaluation, the next practical consideration for a successful joint evaluation is timing. In the case of the joint evaluation of the SGP, on the UNDP side, a terminal evaluation of the current SGP operational phase was required as the current phase (Operational Phase 5 [OP5], 2011–2015) was coming to an end, and the evaluation would serve the purpose of informing the next SGP replenishment. On the GEF side, the GEF IEO needed to include an assessment of the SGP in its fifth Overall Performance Study of the GEF (OPS5) and inform the GEF-6 replenishment negotiations. In general, the timing was appropriate for the two institutions, if not perfect, for a joint evaluation. The timing of the evaluation was dictated by institutional programming time frames, and not the state of the program implementation or its results, a challenge not limited to development evaluation (see, e.g., Chelimsky, 2014, with reference to the U.S. context). More specifically, timing considerations led to a decision to conduct the evaluation in two phases. The first phase served the GEF IEO's need for information on the adoption and implementation of the 2008 recommendations within a fairly short time frame in order to contribute to the OPS5 report. The second phase was an opportunity to build on the first phase findings by addressing additional issues at the heart of the UNDP mandate, particularly related to livelihoods, poverty reduction, and gender/inclusion.

With respect to another aspect of timing, the literature also cites typically longer time frames for joint evaluations. Conducting the joint SGP evaluation in two phases, with reports at each phase, introduced some complications, including reaching agreement on the form the final report should take, which also led to an extended time frame to complete the evaluation. Figure 3 illustrates the time line for the completion of this joint evaluation. The data collection and analysis phases (6 months from approval of the terms of reference to submission of the draft report to the evaluation's joint steering committee) were no different than they would have been in a single-agency evaluation. However, joint steering committee review of the drafts, agreement on various issues, report revisions, and final committee approval of the reports in order to disseminate took 4 months in the first phase (delaying initiation of the second phase) and 7 months in the second phase, considerably longer than the usual time to finalize a single agency report. At the time of the original terms of reference, it was expected that the final second phase report would be delivered in May 2014, but it was only in February 2015 that the final report was sent to GEF and UNDP for management response. With reference to the SDG context and multisectoral interventions, the extended time line of joint evaluations as referenced in the literature and as experienced in the present case may present challenges to relevance of the final evaluation results, as conditions and interactions between dimensions are constantly changing on the ground.

Cofinancing

An important success factor of this joint evaluation—of potential interest to practitioners if not specific to the SDGs context—was the cofinancing. The evaluation was cofinanced by the IEOs of GEF and UNDP, and by the SGP itself, from the SGP budget allocated to the OP5 terminal evaluation. A key aspect of the cofinancing for the evaluation was that the two offices agreed to share costs equally while avoiding complicated fund transfers between two institutions, each with its own financial, administrative, and operational rules and procedures. The budget for the evaluation was designed with each institution taking responsibility for the funding of different components, with regular consultation to ensure the cost split remained equal. The trust built through the previous experience of joint work between facilitated the budget-sharing arrangement.

Management Structure

Designing an appropriate management structure and developing adequate communication mechanisms are important for successful implementation of a joint evaluation. As mentioned earlier, joint evaluations have typically been characterized by complex management and governance arrangements. This was also the case of the present joint evaluation, with a three-tier management structure, including a top-heavy apex body. The challenges of the structure were partially mitigated by the communication strategies adopted by the evaluation managers. The management architecture and roles and responsibilities were as follows:

(i) The Steering Committee, co-chaired by the two offices' directors and including senior staff members from each office. The Committee was a rather crowded decision-making forum, the convening of which required careful coordination of the members' respective busy agendas to avoid delays. The Committee responsibilities included reviewing and approving the terms

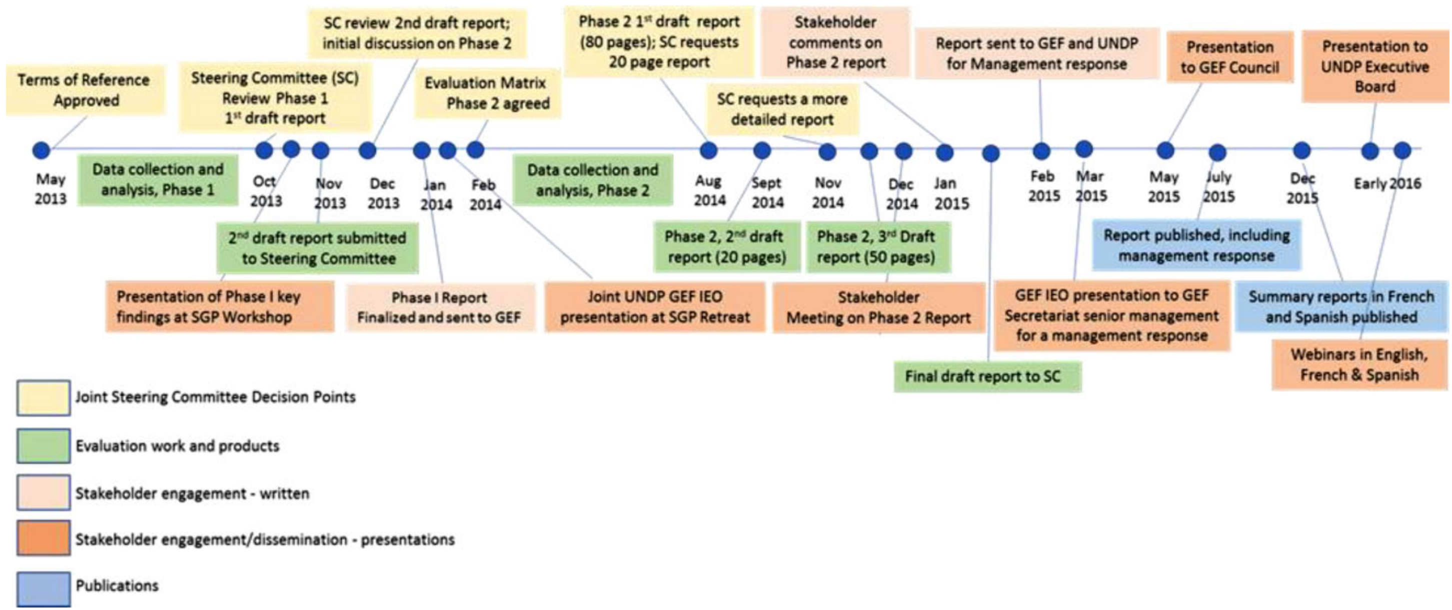

Figure 3. Time line of the Joint Global Environment Facility–United Nations Development Programme evaluation of the Small Grants Programme.
Source. Elaboration based on official records.

of reference, the joint management arrangements, the selection and hiring of consultants, and the evaluation report. It ensured that sufficient and timely resources (human and financial) were made available for the evaluation. The Committee defined the scope jointly in a way to respect the constraints each office had to face (such as the timing, mentioned above), while maintaining the core scope elements that justified a joint evaluation.

(ii) The Management Team, formed by two evaluation comanagers responsible for the overall development and conduct of the evaluation (the authors of this article). The comanagers took care of the selection, hiring, and supervision of consultants in accordance to mutually agreed terms of reference and institutional procedures; coordinated the evaluation activities carried out by both offices, quality-controlled evaluation deliverables, and processes; and ensured the timely delivery of evaluation products. The quality of the collaboration between comanagers is key to the success of joint evaluations. In the 2015 evaluation, this collaboration was smooth and effective, as witnessed by the decision to share our experience through this article. But often, there are cases where partners operate in settings with power differentials or simply do not get along very well due to incompatible personalities. This potentially results either in conflict and delays despite the joint evaluation arrangements. Frequent communication, including short chats over Skype on specific issues, significantly contributed to effective comanagement.

(iii) The Evaluation Team, including the management team complemented by one lead consultant, nine regional or national consultants, and research assistants from both UNDP and GEF evaluation offices assigned to the evaluation. In addition to a long and diversified experience in evaluation (including joint evaluation), both evaluation comanagers had complementary skill sets. The UNDP IEO task manager has experience in poverty, livelihoods, gender, and inequalities; the GEF IEO task manager in environment, natural resource management, and sustainable development. The lead consultant was an expert in evaluation and climate change.

Scope

Determining the scope of an evaluation—balancing the potential interest in answering a wide range of questions with the available resources—is critical for any evaluation. One of the main benefits of joint evaluations cited in the literature is the potential to broaden the scope and the number of questions to be answered given combined resources. However, it is still necessary to ensure the scope is kept within a manageable frame. This was reaffirmed with the joint SGP evaluation, where the scope was broader than it might have been had only one agency conducted the evaluation. The partnership arrangements, discussed above, also played a role. The 2008 evaluation, while joint, was led by the GEF IEO; the 2015 evaluation, where the two offices worked together from beginning to end on a more even footing, broadened the scope beyond environmental benefits to include a more in-depth look at SGP's role in improving livelihoods, reducing poverty, and contributing to gender equality and women's empowerment, themes emphasized by UNDP. Thus, the evaluation looked not only at the environmental but at the interactions between the environmental, social, and economic, the three dimensions at the heart of the 2030 Agenda.

This broadening of the scope beyond environmental benefits did however spark numerous debates throughout the conduct of the evaluation as there were differing interests among the different program and evaluation stakeholders. For example, the SGP program manager (UNDP) saw one of the values of the program as its ability to "reach the most vulnerable," where as many environment specialists and GEF stakeholders challenged the claim, some arguing that it is virtually impossible to work with the most vulnerable as they do not have the necessary capacities to engage in development initiatives or are themselves a threat to the environmental benefits the program seeks to secure. Some GEF stakeholders questioned UNDP's interest in examining SGP's socioeconomic benefits in detail. The relevance of these questions and tensions is illustrated by the fact that one of the evaluation recommendations refers to the need to clarify the SGP's long-term vision, including the balance between global environmental benefits and socioeconomic objectives.

A risk of joint evaluations with their potentially broader scope is that the number of evaluation questions becomes unmanageable, as each agency has its own requirements, needs, and interests. In the case of the joint SGP evaluation, the scope was also overly ambitious in the beginning, but two phases allowed a refocusing of the second phase on the most critical evaluation questions. For example, the original terms of reference included a key question for the second phase asking, to what extent the SGP, through its work with nongovernmental and civil society organizations alike, facilitated civic engagement in the local and/or national policy arena, especially in postconflict and fragile states. UNDP was interested in this aspect of the program, given its mandate on governance and peacebuilding issues (which, in terms of the five "Ps" of the 2030 Agenda, would have added "peace" to the elements of "people, planet and prosperity" already captured by the focus on the social, environmental, and economic aspects of SGP). However, the question was not of primary interest to GEF, and the evaluation management team also came to realize this additional area of inquiry was not possible given the resources and time frame of the evaluation. Thus, both diverging interests of the two partners and practical considerations led to the dropping of the question in revised terms of reference for the second phase of the evaluation.

Multidimensional Analysis

Another feature and important benefit of this joint evaluation was the strengthened multidimensional analysis composed of environmental and socioeconomic research issues that drew on the strengths of the two partner evaluation offices highlighted earlier. This justifies discussing this evaluation as a particularly relevant example supporting the call this article makes for more joint evaluations in the SDG context. The GEF aims at achieving global environmental benefits. While not directly targeting socioeconomic co-benefits, it welcomes when the initiatives it supports manage to achieve them too. UNDP focuses more on poverty reduction, livelihoods, and gender, while the environment is a theme that cuts across UNDP's main strategic and operational pillars. The mandates of the two organizations overlap, but they are not identical, which can lead to differences in approach or emphases to programming and by extension to evaluation (Uitto, 2016). This joint evaluation, however, in terms of evaluation purpose, objectives, and key questions, was designed with a high degree of complementarity. Gender equality and livelihoods are both components of the SDG sustainability and resilience principles. A particularity of this evaluation was that its scope included an assessment of progress made in adopting the recommendations of the previous joint evaluation conducted in 2008, which had to do with the expansion of SGP and the "upgrading" of SGP country programs to GEF full size projects, as well as a follow-up assessment of SGP achievements in terms of global environmental benefits. The 2015 evaluation added a stronger emphasis on livelihoods, poverty reduction, and gender results.

Methods and Tools

Another advantage to joint evaluations is the potential for methodology development and the use of a greater variety of methods for data collection and analysis (UNEG, 2014) brought to the table by the partner agencies. In the case of the SGP joint evaluation, given that some elements of the evaluation were following up on or comparing with the 2008 evaluation findings, many of the evaluation methods and tools were built upon and adapted from the ones developed and used in the first joint evaluation. Modifications were done to fit these 2008 tools to the new evaluation questions. New tools were also introduced. For example, 12 country studies were conducted, with field verification of 144 small grants projects. The team used the same tool that had been developed for the 2008 joint evaluation to assess grant project effectiveness in terms of economic benefits (also allowing for comparison with the 2008 evaluation results). Inspired by UNDP's "gender marker," the team developed new rating tools to assess the contributions of the grant projects to gender equality and women's empowerment as well as to the reduction of poverty, inequality, and exclusion, allowing for more in-depth analysis of these aspects. More generally, data were collected through several additional complementary methods and tools: a global online survey reaching out to almost 2,500 program country stakeholders; an in-depth literature review of SGP documentation at all levels, from global to national; a meta-analysis of 30 randomly selected SGP country program strategies and 50 evaluations covering SGP interventions; a portfolio review of performance and financial data for 18,000 small grant projects; and several key stakeholder interviews and focus groups at central level as well as in the countries visited during the evaluations. These tools generated a huge amount of both quantitative and qualitative data that could be used for generating solidly triangulated findings using techniques developed by the GEF team (Carugi, 2016) in response to each evaluation question. Quantitative data (mostly from the portfolio review but also from the country studies) were used primarily to establish observable facts across the SGP portfolio. The online survey provided qualitative information from program countries. Other qualitative data (collected from all tools except the portfolio review) focused on identifying and understanding the factors affecting results.

This joint evaluation benefited from complementarity of skills and professional experience from UNDP—stronger than GEF in gender and livelihoods—and from the GEF, strong in environment. In addition to this complementarity, the experience offered an opportunity for peer learning, confirming one of the most common benefits of joint evaluations reported in the literature. Each component was designed together bringing diverse experience to bear on each section. The multidimensional analysis improved as a result, fitting perfectly to the nature of the evaluand, a global program tasked with achieving both environmental and socioeconomic benefits at the local level. In other words, the nature of the SGP calls for this kind of integrated analysis, which makes this experience informative in the context of the discussion on joint evaluation for the SDG 2030 Agenda.

Reducing Multiple Messages and Broadening Ownership

Recognized advantages of joint evaluations also include the potential for reducing multiple (and potentially conflicting) messages and broadened ownership of findings and recommendations (UNEG, 2014). The joint evaluation of the SGP examined several strategic and controversial (within the SGP universe) questions and, after extensive consultations including presentations of preliminary findings and recommendations to stakeholders, provided a single set of messages to both the GEF and its Council and UNDP (and by extension its Executive Board). The 2030 Agenda emphasizes the importance of multistakeholder partnerships to support achievement of the SDGs. Joint evaluations have potential to facilitate, as did the SGP evaluation, formulation of common messages to these multiple stakeholders to support adaptation and progress.

Use

Working as equal partners, as discussed above, contributed to evaluation use. In fact, the joint SGP evaluation led to structural changes in both organizations, a clear benefit. The evaluation recommended revitalization of the global SGP Steering Committee as a forum to provide high-level strategic direction and clarity of purpose to the program, for example, in terms of balancing the contributions to global environmental benefits and to other socioeconomic benefits. The GEF Council, to which the GEF IEO reports directly, issued a corresponding decision addressed to the GEF Secretariat. UNDP also adopted recommendations regarding management of different SGP country programs at different levels of maturity within a single unit to ensure coherence. Had the evaluation not been joint, detailed recommendations may not have been made to both institutions at the highest level. For example, UNDP alone could not have made recommendations to the GEF Council, and the GEF ensured that recommendations related to global environmental benefits were included, which might not have been the case had UNDP undertaken the evaluation by itself.

Conclusions

Joint evaluations have been the subject of several guides and analyses (Feinstein & Ingram, 2003; OECD-DAC, n.d., 2005, 2013; SADEV, 2008; UNEG, 2014). Their application in evaluation practice has seen highs and lows during the last three decades. In recent years, a backlash and resistance to joint evaluations due to high transaction costs and heavy coordination requirements has been observed, as has been illustrated with data from the UN and UNDP. However, 2015 represents a shift in the international development agenda with the advent of the SDGs. Their implications for multistakeholder approaches and integrated programming need to be incorporated into the evaluation agenda. As a result, more joint evaluations are needed and, if calls such as the 2019 UN funding compact for the SDGs are heeded, to be expected. In addition, with the engagement of all countries in implementing the SDGs, lessons from the field of development evaluation may have wider application for public policy evaluation. Thus, it is appropriate to give a fresh look at joint evaluations in the context of the SDGs.

The long and diversified experience in joint evaluations has produced lessons in terms of advantages and disadvantages as well as “do’s and don’ts,” many of which have been confirmed by the joint SGP evaluation discussed in this article. For example, joint evaluations are suited when all institutions involved have a stake in the scope covered, provided the scope remains sufficiently delimited, so that the key evaluation questions are kept to a manageable number. Institutional arrangements can easily become cumbersome, and care should be taken to keep institutional architecture of the evaluation as simple as possible, especially at the joint steering committee level. While not the reason for undertaking a joint evaluation, complementarities in expertise and professional experience of both institutions and professionals involved at various levels should be leveraged. In addition, joint evaluation task management must be done closely and in a balanced way. The joint SGP evaluation has demonstrated that engaging in short daily exchanges and communications between comanagers has proven a very effective form of joint task management. Joint evaluations will inevitably have transaction costs for the partnering evaluation units, but these can be minimized with appropriate architectures and regular communication.

In summary, the experience of the joint SGP evaluation supports the findings of earlier studies that the greatest advantages of joint evaluations include the potential for broader scope than single-agency evaluations, increased multidimensionality in the analysis which may be facilitated by complementary tools and methods brought by different partners, and a single set of coherent messages addressed to multiple partners. These benefits have even greater salience in the context of the SDGs, where interactions and linkages between goals and targets and the need for integrating different sectoral programs and to develop cross-sectoral policy are increasingly recognized (e.g., Mainali, Luukkanen, Silveira, & Kavio-oja, 2018) and where the lessons from the field of development evaluation may be relevant for public policy and program evaluation in other contexts. Joint evaluations facilitate the expansion of scope beyond a traditional single sector, bringing together partners with different knowledge and skill sets to foster cross-sectoral analysis, examine trade-offs and highlight synergies, and to craft meaningful messages relevant for multiple actors.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Carlo Carugi https://orcid.org/0000-0002-1761-2965

References

Andersen, O. W., & Broegaard, E. (2012). The political economy of joint-donor evaluations. *Evaluation*, *18*, 47–59.

Boucher, L. (2015). *Sustainable development goals vs. millennium development goals: What you need to know*. Washington, DC: Population and Education, a Program of Population Connection. *PopEd* Blog.

Carugi, C. (2016). Experiences with Systematic Triangulation at the Global Environment Facility. *Evaluation and Program Planning*, *55*, 55–66. https://doi.org/10.1016/j.evalprogplan.2015.12.001

Chelimsky, E. (2014). Public-interest values and program sustainability: Some implications for evaluation practice. *American Journal of Evaluation*, *35*, 527–542.

Coonrod, J. (2014). *MDGs to SDGs: Top 10 differences*. New York, NY: The Hunger Project.

COWI & International Institute for Environment and Development. (2009). *Joint external evaluation of the operation of the least developed countries fund for adaptation to climate change*. Copenhagen, Denmark: Evaluation Department, Danish Ministry of Foreign Affairs.

Feinstein, O., & Ingram, G. K. (2003). *Lessons learned from World Bank experiences in joint evaluation* (Room Document No. 3). Paris, France: Development Assistance Committee Working Party on Aid Evaluation.

Global Environment Facility Independent Evaluation Office. (2006). *Joint evaluation of the GEF activity cycle and modalities* (Evaluation Report No. 33). Washington, DC: Author.

Global Environment Facility Independent Evaluation Office. (2010). *The GEF monitoring and evaluation policy 2010* (Evaluation Document No. 4). Washington, DC: Author.

Global Environment Facility Independent Evaluation Office. (2014). *Joint GEF/Sri Lanka country portfolio evaluation* (Evaluation Report No. 96). Washington, DC: Author.

Global Environment Facility Independent Evaluation Office. (2017). *OPS6 report: The GEF in the changing environmental finance landscape* (Evaluation Report No. 110). Washington, DC: Author.

Heider, C. (2015). *Evaluation beyond 2015: Implications of the SDGs for evaluation*. Washington, DC: World Bank, Independent Evaluation Group (IEG). Blog What Works.

Global Environment Facility Independent Evaluation Office & United Nations Development Programme Independent Evaluation Office. (2008) *Joint evaluation of the Small Grants Programme* (Evaluation Report No. 39). Washington, DC: Author.

Global Environment Facility Independent Evaluation Office & United Nations Development Programme Independent Evaluation Office. (2016a). *Joint GEF-UNDP evaluation of the Small Grants Programme* (Evaluation Report No. 97). Washington, DC: Author.

Global Environment Facility Independent Evaluation Office & United Nations Development Programme Independent Evaluation Office. (2016b). *Biodiversity impact evaluation—Support to protected areas and protected area systems* (Evaluation Report No. 104). Washington, DC: Author.

International Council for Science. (2017). *A guide to SDG interactions: From science to implementation*. Paris, France: Author.

Kumar, S., Kumar, N., & Vivekadhish, S. (2016). Millennium development goals (MDGS) to sustainable development goals (SDGS): Addressing unfinished agenda and strengthening sustainable development and partnership. *Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive & Social Medicine*, *41*, 1–4.

Lähteenjoja, S., Berg, A., & Korhonen-Kurki, K. (2019). *Sustainable development requires integration of the 2030 into policy-making* (Policy Brief 4/2019). Helsinki, Finland: Government's Analysis, Assessment and Research Activities.

Lebada, A. M. (2019). *Governments, UN finalize funding compact for SDGs*. Winnipeg, Canada: IISD. Retrieved April 8, 2019, from https://sdg.iisd.org/news/governments-un-finalize-funding-compact-for-sdgs/

Mainali, B., Luukkanen, J., Silveira, S., & Kavio-oja, J. (2018). Evaluating synergies and trade-offs among sustainable development goals (SDGs): Explorative analyses of development paths in South Asia and sub-Saharan Africa. *Sustainability*, *10*, 815.

Montague, S., & Teather, G. C. (2007). Evaluation and management of multi-departmental (horizontal) science and technology programs. *Research Evaluation*, *16*, 183–190.

Organization for Economic Cooperation and Development: Development Assistance Committee. (n.d.). *Effective practices in conducting a multi-donor evaluation*. Paris, France: Author.

Organization for Economic Cooperation and Development: Development Assistance Committee. (1998). *Annual report by the Chair of the DAC Working Party on Evaluation* (Report No. DCD/DAC/EV(99)1). Paris, France: Author.

Organization for Economic Cooperation and Development: Development Assistance Committee. (2005). *Joint evaluations: Recent experiences, lessons learned and options for the future* (DAC Evaluation Network Working Paper). Paris, France: Author.

Organization for Economic Cooperation and Development: Development Assistance Committee. (2013). *The DAC network on development evaluation—30 years of strengthening learning in development*. Paris, France: Author.

Reynolds, M., Blackmore, C., Ison, R., Shah, R., & Wedlock, E. (2018). The role of systems thinking in the practice of implementing sustainable development goals. In W. Leal Filho (Ed.), *Handbook of sustainability science and research* (pp. 677–698). World Sustainability Series. Cham, Switzerland: Springer.

Rowe, A. (2012). Evaluation of natural resource interventions. *American Journal of Evaluation*, *33*, 384–394.

Schwandt, T., Ofir, Z., Lucks, D., El-Saddick, K., & D'Errico, S. (2016) *Evaluation: A crucial ingredient for SDG success* (IIED Briefing). London, England: International Institute for Environment and Development.

Shadish, W. R., & Luellen, J. K. (2005). History of evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 183–186). Thousand Oaks, CA: Sage.

Swedish Agency for Development Evaluation. (2008). *Challenges, opportunities and approaches for increasing joint donor programming of evaluations* (SADEV Report 2008:8). Karlstad, Sweden: Author.

Thomas, V. (2017). Evaluating sustainable development. In R. van den Berg, I. Naidoo, & S. Tamondong (Eds.), *Evaluation for agenda 2030: Providing evidence on progress and sustainability* (pp. 9–21). Exeter, England: International Development Evaluation Association (IDEAS).

Tosun, J., & Leininger, J. (2017). Governing the interlinkages between the sustainable development goals: Approaches to attain policy integration. *Global Challenges*, *1*, 1700036.

Uitto, J. I. (2016). The environment-poverty nexus in evaluation: Implications for the sustainable development goals. *Global Policy*, *7*, 441–447.

United Nations Development Group. (2016). *Principles for integrated programming*. Retrieved February 22, 2018, from https://undg.org/programme/undaf-guidance/principles-for-integrated-programming/

United Nations Development Programme. (2016). *UNDP support to the implementation of the 2030 agenda for sustainable development* (UNDP Policy and Programme Brief). New York, NY: Author.

United Nations Development Programme Independent Evaluation Office & German Federal Ministry for Economic Cooperation and Development (BMZ). (2000). *The UNDP role in decentralization and local governance: A joint UNDP-Government of Germany Evaluation*. New York, NY: Author.

United Nations Development Programme Independent Evaluation Office & Ministry of Foreign Affairs of Japan. (2002). *Post-conflict assistance of the Government of Japan through UNDP in Kosovo and East Timor*. New York, NY: Author.

United Nations Economic and Social Council. (2015, March 18). *Thematic evaluation of monitoring and evaluation for the Millennium Development Goals: lessons learned for the post-2015 era* (Report No. E/AC.51.2015/3) Retrieved from https:/oios.un.org/page/download2/id/2

United Nations Evaluation Group. (2008). *Joint evaluation of the UNDG contribution to the Paris declaration on aid effectiveness (First phase)*. New York, NY: United Nations Development Programme.

United Nations Evaluation Group. (2009). *Joint evaluation of the role and contribution of the UN system in the Republic of South Africa*. New York, NY: Author.

United Nations Evaluation Group. (2014). *Resource pack for joint evaluations*. New York, NY: Author.

United Nations Evaluation Group. (2016). *Evaluation in the SDG era: Lessons, challenges and opportunities for UNEG*. New York, NY: Author.

United Nations Evaluation Group. (2017). *System-wide evaluation in support of UN reform: Towards collective accountability and learning. White Paper #1*. New York, NY: Author.

United Nations General Assembly. (2006, November 20). *Note by the Secretary-General [transmitting report of the High-Level Panel on UN System-wide Coherence in the Areas of Development, Humanitarian Assistance and the Environment]* (Report No. A/61/583). Retrieved from https://undocs.org/A/61/583

United Nations General Assembly. (2013, January 22). *Resolution adopted by the General Assembly on 21 December 2012* (Report No. A/RES/67/226). Retrieved from https://undocs.org/en/A/RES/67/226

United Nations General Assembly. (2015, October 21). *Transforming our world: the 2030 agenda for sustainable development* (Report No. A/RES/70/1). Retrieved from https://undocs.org/en/A/RES/70/1

United Nations General Assembly Economic and Social Council. (2018, January 19). *Implementation of General Assembly Resolution 71/243 on the quadrennial comprehensive review of operational activities for development of the United Nations system, 2018* (Report of the Secretary General. A/73/63-E/2018/8). Retrieved from https://undocs.org/A/73/63

United Nations Industrial Development Organization & United Nations Development Programme Independent Evaluation Office. (2009). *Joint terminal evaluation of the implementation of the cooperation agreement between the United Nations Industrial Development Organization and the United Nations Development Programme*. Vienna, Austria: Author. Retrieved from https://erc.undp.org/evaluation/evaluations/detail/4384

United Nations Industrial Development Organization Evaluation Group & United Nations Development Programme Independent Evaluation Office. (2006) *Joint assessment UNIDO-UNDP cooperation agreement pilot phase*. Vienna, Austria: Author. Retrieved from https://www.unido.org/sites/default/files/2007-11/58756_FINAL_EVAL_REPORT_20061020_Joint_assessment_UNIDOUNDP_Cooperation_Agreement_0.pdf

Woolbridge, M. (2015). *From MDGs to SDGs: What are the sustainable development goals* (ICLEI Briefing Sheet—Urban Issues, No. 01). Bonn, Germany: ICLEI—Local Governments for Sustainability.

World Health Organization. (2015). *Health in 2015: from MDGs, millennium development goals to SDGs, sustainable development goals*. Geneva, Switzerland: Author.