

The background features a large, abstract circular graphic composed of numerous overlapping, semi-transparent blue segments of varying shades, creating a dynamic, layered effect. The segments are arranged in a roughly circular pattern, with some extending towards the edges of the frame. The overall aesthetic is modern and technological.

**Part V**  
Toward  
Sustainability  
of Impact



---

## Chapter 18

# New Frontiers for Evaluation in a Fast-Changing World

Chris Barnett and Rachel Eager

**Abstract.** *The world has changed significantly in the past decade: rising inequality, conflict and insecurity, mass migration, terrorism, and climate change all present major global challenges. The responding United Nations' 2030 Agenda for Sustainable Development sets out a holistic approach, cognizant of the interconnectedness between society, economy, and the environment. In addition, many private investors and businesses are increasingly aware of their long-term interdependence on natural systems. These changes provide both an opportunity and a significant evaluative challenge for those who have traditionally operated within the aid/development sphere. This chapter considers new frontiers for the evaluation profession in terms of methodology, engagement with new actors, and how best to provide evaluative evidence within complex and rapidly changing contexts. Responding to global challenges requires more than just methodological improvement and innovation. There is a need for a bolder evaluation agenda, recognizing the evaluators' role in contributing to change: acting not just as providers of evidence, but to proactively engage in an ethical obligation to society, stimulating deliberation and re-examination of evidence by a broader range of citizens—citizens who can be emboldened to use such evidence to improve their situations and hold others to account.*

---

Chris Barnett, Centre for Development Impact and Itad, [Chris.Barnett@itad.com](mailto:Chris.Barnett@itad.com);  
Rachel Eager, Centre for Development Impact and Itad, [rachel.eager@itad.com](mailto:rachel.eager@itad.com).

The world has changed significantly in the past decade, with rising inequality, conflict and insecurity, mass migration, terrorism, and climate change all presenting major challenges for our collective future. New influential trading powers like China and India challenge the economic status quo, while the choices of voters in the United States and the United Kingdom suggest a re-emergence of protectionist and populist policies in response to increasing globalization and a new multipolar political order. The old East-West, North-South, developed-developing divides look increasingly irrelevant. Meanwhile, over the past few decades, evaluation as a field has developed largely in response to public policy and government intervention, by assessing effectiveness and impact on behalf of governments and their taxpaying constituents. In international aid particularly, this has been underpinned largely by assumptions from an old-world order: it has been focused on public expenditure commissioned by donors for recipient countries, and dominated by large-scale interventions managed mostly by international nongovernmental organizations (INGOs), agencies of the United Nations (UN), or “managing agents” (Western private companies).

In this chapter, we consider how the field of evaluation might best respond to a changing world, and in doing so set out new frontiers for the coming decades. The first part describes major global trends, with billions of people continuing to live in poverty, and with growing inequality between the richest and the poorest; threats to the environment and our ecosystems; and insecurity, migration, and conflict. In response to these challenges, the international community has set out an ambitious vision for the future: the 2030 Agenda for Sustainable Development, which includes the Sustainable Development Goals (SDGs); the Addis Ababa Development Financing Action Agenda; the World Humanitarian Summit; and the Paris Climate Change Agreement. In parallel, there has been a rising tide of private capital and businesses consciously aiming to become a force for good (Social Impact Investment Taskforce 2014a). This includes various forms of socially and environmentally responsible private and blended capital (e.g., impact investing and venture philanthropy), as well as new modalities and changes in the ways of doing business such as the Blueprint for Business.<sup>1</sup> In the second half of the chapter, we explore the implications of these changes for evaluation in terms of both methodology and new partnerships with new actors. Finally, we argue that these global challenges require something more fundamental than just changing the way in which evaluators respond and adapt to a changing context. Rather, we argue that they imply the need for a bolder evaluation agenda, in which evaluators contribute to the change itself and take up a more value-driven mantra: not only to provide evidence-based assessments, but also to engage in an ethical obligation to society to make evidence available in such a way that it can be deliberated upon and reexamined by a broader cross-section of the population. At a time when many of the most vulnerable feel disenfranchised by global trends, it is perhaps even more

---

<sup>1</sup>[www.blueprintforbusiness.org](http://www.blueprintforbusiness.org).

pertinent for evaluators to embrace technology, new ways of working, and new partnerships in order to focus increasingly on the ultimate “client” (the poor and marginalized), by empowering citizens to better use evidence to inform, challenge, and call to account politicians, policy makers, development professionals, and private companies.

## GROWING INEQUALITY

There is increasing recognition that inequality is one of the major challenges of our time. In recent years, seminal work by leading economists has highlighted a growing divide between the world’s richest and poorest citizens (Atkinson 2015; Bourguignon 2015; Picketty 2013). Indeed, inequality within countries continues to rise (Oxfam 2014), with the United Nations Development Programme (UNDP) reporting that income inequality in developing countries increased by 11 percent between 1990 and 2010 (UNDP 2013). Furthermore, the location of the poorest is changing, and the distribution is no longer concentrated in low-income countries. Research shows that 72 percent of the world’s poor (960 million) now live not in poor countries, but in middle-income countries (MICs), a dramatic shift from two decades ago, when the vast majority (over 90 percent) lived in low-income countries (Sumner 2012). Increasingly, the problems of poverty are a challenge for MICs that are less dependent, and may not be at all dependent, on development assistance. As such, poverty reduction in many MICs becomes less about having enough resources and more about having the political will to address issues of redistribution. This has implications for development assistance, which is increasingly being used to target the most stubborn problems in the least developed countries, especially among fragile and conflict-affected states (Picciotto 2015a). Not only is this a challenge to the traditional North-South model of development assistance: it also affects evaluators and the role of evaluation. While evaluators need to retain some focus on development assistance, increasingly there is a new role emerging: to support national-level (and country-led) policy objectives, and to consider policy coherence internationally. This requires not simply evaluating aid-driven (or micro level) interventions in isolation, but also assessing the effects of other policies—of trade, investment, environmental protection, foreign policy, immigration, and so on—to either mutually support (or to undermine) poverty, inequality, and sustainability objectives (Picciotto 2005).<sup>2</sup> Or, as van den Berg and Cando-Noordhuizen (2017) have put it, “Evaluators need to point out to policy makers and decision makers that what they promote with one hand, is more than sufficiently undone with a very active and much bigger other hand.”

---

<sup>2</sup> See, e.g., the blog discussion by Heider (2017).

## **PUSHING PLANETARY BOUNDARIES**

At the same time, our understanding of the environment is changing. New thinking and new science suggest that there are planetary boundaries, i.e., that there are thresholds within which there exists a safe operating space for humanity. For instance, Rockström et al. (2009) suggest a framework of nine planetary boundaries (including stratospheric ozone depletion, ocean acidification, and biogeochemical flows) within which humanity can continue to develop and thrive for generations to come. Work on climate change also suggests that growing global resource demands will continue to exacerbate the likelihood of increased flooding, heat waves, earthquakes, and other natural disasters, the consequences of which are linked in many countries to food security concerns and increased levels of conflict (UN 2015c; World Economic Forum 2016). Economic, social, and environmental developments are increasingly viewed as interlinked, and pose challenges for policy makers, the private sector, and evaluators. The latter can no longer operate purely within defined sectoral boundaries, and even social systems are too limiting: there is an increasing need to consider the relationship between social and natural systems, as well as longer-term time frames—that is, no longer 3–5 years, but 20, 50, or 100 years (Rowe 2012).

## **CONFLICT, SECURITY, AND SAFETY**

Alongside a growing gap between the richest and poorest, and a greater vulnerability to risks and disaster-related losses especially for the poorest, deterioration in indicators of peace in the Middle East and North Africa have been so severe they have masked increases in other areas. The global peace index has highlighted deterioration in the impact of terrorism and political instability indicators as key contributors (Institute for Economics and Peace, 2014). The severity of the situation in the Middle East has wider implications. Global levels of displacement are now higher than ever before, with 65.3 million people living in exile at the end of 2015—a population of forcibly displaced people that is greater than the entire population of the United Kingdom (UNHCR 2015). Large-scale global migration is now rated as one of the most likely and impactful global risks (World Economic Forum 2016). The intertwined dynamics of terrorism, conflict, and political instability within a small number of countries has consequences at a global level, with the economic impact of violence estimated at 13.3 percent of world GDP (Institute for Economics and Peace 2014). Evaluation, particularly when undertaken in fragile contexts, increasingly has to respond to dynamic situations, where the politics of diplomacy, military intervention, and peace-building activities often intersect with more traditional forms of humanitarian and development assistance (Broegaard, Bull and Kovsted 2014).

## **THE POST-2015 ERA: RESPONDING TO GLOBAL CHALLENGES**

The above-mentioned challenges of poverty, inequality, insecurity, and the environment are increasingly being recognized, with the re-emergence of “sustainable development” as a unifying concept—that is, the process of

meeting human development goals while sustaining the very natural systems that are needed to provide the resources and ecosystem services for humanity (both society and the economy) to thrive. For the coming decades, the international community has set out an ambitious vision: the 2030 Agenda for Sustainable Development (in September 2015), which includes the SDGs; the Addis Ababa Action Agenda on development financing (in July 2015); the World Humanitarian Summit (in May 2016); and the Paris Climate Change Agreement (in December 2015). The SDGs in particular represent a major shift toward global responsibility, unlike the Millennium Development Goals (MDGs), where the focus was on developing countries. These agreements also offer a more holistic vision that acknowledges the interconnectedness between objectives (societal, economic, and environmental), and advocates partnerships that go beyond governments, and that include the private sector.

The Agenda for Sustainable Development presents a vision for eradicating poverty and tackling inequality, and addressing the need for economic, social, and environmental sustainability with a breadth of focus that recognizes the interconnected nature of the issues faced. While the MDGs focused on identifying and filling gaps, the SDGs ask a more holistic question about how we can stimulate sustainable progress across a much broader range of complex and interrelated goals (Weisen and Prokop 2015).

Evidence and learning are also integrated into the 2030 Sustainable Development Agenda, which is monitored via a results framework composed of the 17 ambitious SDGs, and 169 quantitative and qualitative target indicators. These goals are aspirational and universal, setting a framework within which countries define their own targets based on national priorities and contexts. In contrast to the tracking of the MDGs, where disparities in performance were masked by aggregate-level reporting, the 2030 Agenda calls for the monitoring of targets at all levels, and an increased focus on data disaggregation by a range of characteristics including income, race, age, gender, disability, and ethnicity (Weisen and Prokop 2015). Systematic follow-up and review processes are also prominent within the agenda, positioning evidence, data, and evaluative thinking as critical to the achievement of the strategy. Review processes are expected to operate at the national, regional, and global levels, in order to “promote accountability to our citizens, support effective international cooperation...and foster exchanges of best practices and mutual learning” (UN 2015a, clauses 72–73). Follow-up and review processes will be informed by the tracking of progress toward goals, and rigorous country-led evaluation, and are expected to make “a vital contribution to implementation and will help countries to maximise and track progress...in order to ensure that no one is left behind” (UN 2015a, clauses 72 and 74g).

The Agenda poses many challenges for evaluation. First, it requires country-led evaluation which requires strengthening national evaluation programs through enhanced capacity-building support for developing countries. Although as highlighted by the International Institute for Environment and Development (IIED), little guidance is given on how countries may set their own national agendas, and evaluation will need to address issues such as whether or not progress made is equitable, relevant, and sustainable (Ofir et al. 2016; Schwandt 2016). Second, lessons from the monitoring and

evaluation (M&E) of the MDGs suggest that the MDG approach oversimplified the development narrative, essentially creating sector-based silos. In response to this, the interrelated nature of the SDGs demonstrates the need for a more holistic development—and therefore evaluation—approach. Third, moving away from the public sector, donor-centric MDG approach, there is a clear recognition of the need for diversification in funding mechanisms. The new agenda emphasizes the role of multiple change agents, and recognizes the need for contextual flexibility and the disaggregation of data (Ofir 2015).

## **MOBILIZING RESOURCES FOR THE POST-2015 AGENDA**

Financing this ambitious global agenda requires a significant increase in resources allocated, and a mobilization of resources far beyond that of development assistance. The United Nations Conference on Trade and Development (UNCTAD) has estimated the total investment needs in developing countries to amount to \$3.9 trillion annually. Current investment reportedly stands at \$1.4 trillion, highlighting a substantial investment gap of \$2.5 trillion per year (UNCTAD 2014, 145). In many MICs, the public revenues generated are sufficient to meet costs: however, insufficient funds are being allocated to basic services. The current political climate in both Europe and the United States increasingly challenges the commitment to deliver 0.7 percent of gross national income as official development aid (Nakhoda et al. 2016). In this resource-constrained context, the importance of private sector investment is heightened, changing the dynamics of development finance significantly. The role of the private sector is clearly acknowledged within the 2030 Agenda, and the Addis Ababa Action Agenda encourages philanthropic donors to continue their engagement through impact investments, with calls for increased transparency and accountability across the sector (UN 2015a, 2015c).

Meanwhile, there is a potential convergence between this globally agreed agenda and the individual interests of some in the private sector. Over the past decade, the private sector has been evolving considerably, with the emergence of businesses with a more deliberate social or environmental conscience. The long era of corporate social responsibility (CSR) continues, but this has sometimes been plagued by accusations of tokenism, or “greenwashing.” Indeed, claims of CSR are rarely evaluated (Picciotto 2015b), therefore little is known about the effectiveness of these approaches in creating social and environmental change (Flynn, Young, and Barnett 2015). But whereas CSR has sometimes been viewed as an add-on to the core business, there is now a growing movement that is advocating using business as a force for good—for example, initiatives such as Blueprint for Better Business, the World Forum on Natural Capital, and B-Corp. Building on the pioneering work of businesses such as The Body Shop as well as fair trade and other certifications, these movements aim to encourage businesses to define and operate with a purpose that serves society and the environment. B-Corp, for instance, now has more than 2,000 businesses that are certified to its standards of social and environmental performance, accountability, and transparency.



The investment field provides a good example of the multiplicity of challenges faced by more traditional forms of evaluation. In recent decades, new variants of purpose-driven, or mission-based, capital have emerged, aiming for both social and environmental returns along with their financial benefits. The boundaries between traditional public sector–driven international aid and private sector investment are becoming more and more blurred with the emergence and proliferation of a huge range of social and environmental investment funds, financial intermediation schemes, and insurance products aimed at achieving profit with purpose (Picciotto 2015a; Social Impact Investment Taskforce 2014a). Private sector resources are now being directed toward socially responsible purposes, with new investment modalities seeking to achieve both financial and social or environmental returns.

There is a broad range of approaches to socially or environmentally responsible capital, existing across a spectrum that ranges from philanthropic giving to traditional, profit-driven investment (Avantage Ventures 2011). As shown in figure 18.1, investment modalities can be broadly categorized into five groups: socially responsible investments; environmental, social, and governance investments; impact investing; program-related investing; and venture philanthropy. Toward the profit-oriented end of the spectrum, socially responsible investing focuses mainly on “do no harm” principles that involve avoiding investments in companies with ethically, socially, or environmentally questionable business practices. However, a new breed of environmental, social, and governance (ESG) investments are oriented more toward “doing good” by incorporating ESG factors consciously into their investment decisions, and aiming to improve the sustainability and ethical impacts of an investment (Avantage Ventures 2011).

At the other (socially oriented) end of the scale, there has been a rise in philanthropy, which has more direct, charitable-giving purposes. Venture philanthropists in particular provide flexible financial support and mentoring to social entrepreneurs and organizations that are aiming to drive innovation and social change, and to achieve operational sustainability. Financial support has traditionally been focused on grant giving, but has broadened recently to include other mechanisms, such as equity-like investments and loans.<sup>3</sup> Program-related investments go further still, and are made where there is a potential for return on investment within a specific period of time, allowing recipients to access capital at lower rates.

The growing industry of purpose-driven investments has already leveraged a substantial amount of private capital for social and environmental “goods.” The potential of social impact investing to bring new capital to developing economies, and to advance development using market principles, is well recognized (Picciotto 2015a; Social Impact Investment Taskforce 2014b). Mapping market trends in global impact investing, the Global Impact Investing Network reports an increase in impact investing assets under management

---

<sup>3</sup> Source: Social Innovator, “Venture Philanthropy,” <http://www.socialinnovator.info/ways-supporting-social-innovation/third-sector/mission-related-investment/venture-philanthropy>.

FIGURE 18.1 Spectrum of social and financial objectives



SOURCE: Avantage Ventures 2011.

from \$25.5 billion in 2013 to \$35.5 billion in 2016, with survey respondents committing an average \$8.5 billion per year between 2013 and 2015 (GIIN 2016).

Progress in the field has been supported by the creation of new networks, approaches and guidance, and standards and metrics (Jackson 2013). However, to date there has been a limited focus on evidencing impact and systematic, independent evaluation, increasing concerns about the validity of current measures and approaches to assessing social and environmental impact (Picciotto 2015a). There is a significant risk that sectoral standards of impact assessment focus more on marketing claims of impact rather than demonstrable social change (O'Flynn and Barnett 2016); many key players are therefore calling for increased scrutiny of impact claims (Brest and Born 2013).

In a review of impact measurement initiatives emerging across the impact investing field, Flynn, Young, and Barnett (2015) identified a huge range of tools and approaches documented in the gray literature. The most prominent of these were the Impact Reporting and Investment Standards (IRIS) and the Global Impact Investing Rating System (GIIRS). The IRIS provides investors with a standardized menu of indicators ranging from jobs created through to sector specific changes; while the GIIRS sets out a rating system (guided by the IRIS) that can be used to assess companies, funds, and their portfolio companies in four key areas: governance, workers, community, and the environment.

While these approaches have moved the field forward in terms of metric standardization and reporting against output-level indicators, they have contributed little to our understanding of investment outcomes or impact. This can result, for example, in better measurement of the number of people employed by an organization but with limited, or no evidence, of the social value that these jobs create (Brest and Born 2013). Flynn, Young, and Barnett have concluded that while a wide range of tools and approaches exist, the predominant focus on metrics and rating systems concentrates on counting inputs and outputs rather than on establishing any meaningful understanding of social and environmental change through the measurement of intended outcomes (Flynn, Young, and Barnett 2015; Jackson 2013). Moving forward, it is clear that a variety of methods and approaches will be required in order to effectively evaluate the diverse range of impact investments occurring in such a broad spectrum of operating contexts. Many of the design approaches, and both the quantitative and qualitative methods employed in public sector development evaluation, will add value, especially when considering the participation of a wider range of stakeholders within the evaluation process (Jackson 2013).

### **THE NEW FRONTIERS: TAKING UP THE EVALUATION CHALLENGE**

To this point, this chapter has considered recent global trends such as growing inequality, environmental threats, mass migration, and insecurity and conflict, and the international response to them. Over the past decade,

however, evaluation, specifically in international development, has become a rather introverted field, with most effort focused on the inadequacies of methodology and the need for rigor in impact evaluation (Picciotto 2012). Certain methodologies have come to dominate the debate (spearheaded by work such as that of Duflo and Kremer 2003; Savedoff, Levine, and Birdsall 2006; and White 2009), particularly because they offer an apparent certainty to (mostly Western) donors under pressure to demonstrate accountability to their national parliaments. This pressure to prove “demonstrable impact” has trickled down through the system, from the policies and procedures of donor agencies, through staff priorities and capacities, and onward through funding mechanisms to multilateral agencies, NGOs, researchers, consultancies, and so on. Much of the focus has been on selecting the *best* method, and in many cases this rests on an assumption that particular designs and methods are superior (a “hierarchy of evidence”).<sup>4</sup> The term “rigorous impact evaluation” for instance, has become synonymous with experimental and quasi-experimental methods, while definitions of “impact” are often limited to counterfactual notions of causal inference (Stern et al. 2012). Even those who take a broader view of evaluating impact are still primarily concerned with selecting the best method; although admittedly rather than focusing on the intrinsic superiority of one method over another, their focus has been on the *appropriateness* of methods in line with the evaluation purpose, the evaluation questions, the context, and the characteristics of the intervention (Stern et al. 2012).

Yet far less attention has been paid to the changing demands for evaluation—that is, who is now asking the impact questions, and what questions really need answering? Indeed, it is clear, as outlined in the preceding sections of this chapter, that evaluators are now operating in an increasingly rapidly changing and volatile environment; and they are facing complex and interrelated issues in which the traditional linear approaches are insufficient to describe the changes. Linear, cause/effect approaches to evaluation are insufficient to understand such complex interactions and the contextual variation that is influencing progress toward the SDGs (Befani, Ramalingam, and Stern 2015; Picciotto 2015b). Going forward, evaluation will need to draw on methods and approaches from further afield, including systems thinking and complexity science in order to support understanding of change in these circumstances (Barder and Ramalingam 2012; Ofir 2016). Monitoring and evaluation systems are required at the global, country, sector, and local levels, with a wide variety of evaluative approaches needed to assess achievements (Picciotto 2015b).

But are these new frontiers only about a methodological revolution—a new science for evaluation? In the next sections, we argue that while there are undoubtedly implications for methodology, much more is needed. Methodological innovation and adaptation is key to meeting the evaluation challenges

---

<sup>4</sup> Hierarchies are well established in the evidence-based policy tradition (e.g., evidence-based medicine), including the Cochrane Collaboration and Campbell Collaboration approaches to systematic reviews. In the development field, 3ie has taken this tradition on board.

of the SDGs: these methodological challenges will require flexibility, creativity, and innovation. No one approach or method will be sufficient, and there will be no “gold standard” (Picciotto 2015a). In addition, it will require building partnerships within and beyond the evaluation community in order to support both the country-led focus of SDGs through capacity building (Ofir 2016), as well as the new demands from a range of private and public actors. Indeed, the evaluation community should engage more effectively with the private sector, by promoting dialogue and mutual understanding, and increasing the demand for evaluative products. There is some level of urgency here in order to avoid these roles being filled by management consultancies, auditors, and accountancy firms, many of who have pre-existing relationships with investors or private companies, but limited experience in social (i.e., developmental) and environmental impact evaluation.<sup>5</sup>

Finally, we conclude that the evaluation field needs to take on a more value-driven approach. The increase in private sector social investments and the associated lack of public accountability mechanisms necessitates a response from evaluators in order to promote transparency in claims of impact; to support inclusive evaluative processes; to stimulate demand for (and use of) evidence in decision making; and to include deliberation and accountability for, and by, citizens.

## METHODOLOGICAL INNOVATION AND ADAPTATION

The methodological challenges facing the evaluation field in this new era are multiple and varied. The reframing of global development goals represents a conceptual shift in our approach to addressing issues of poverty and inequality, as we move from thematically silo-based thinking toward a greater understanding and acceptance of the complexity of the issues being faced. Understanding issues such as resilience, working in fragile and conflict-affected contexts, and meeting the requirements of private sector approaches will increase the complexity of evaluation in various ways, but all of them will challenge the reliance on results-based, linear, and experimental designs (Picciotto 2015a). We propose that four different responses will be required from the evaluation profession in the coming years:

- Methodological pluralism within coherent evaluation design
- Systems thinking and complexity science
- Increasing agility and flexibility
- Capitalizing on the data revolution

### Methodological Pluralism within Coherent Evaluation Design

Since the early 21st century, development evaluation has been characterized by a reliance on experimental and quasi-experimental approaches,

---

<sup>5</sup>As concluded in Clarke, Barnett, and van den Berg (2015).

with proponents declaring counterfactual logic the only valid approach to establishing causality (Picciotto 2012). However, limitations in experimental designs, and their inapplicability in a wide range of social and environmental contexts, have called this thinking into question, leading to increased interest in a broader range of evaluation designs and theoretical approaches to establishing causality (Stern et al. 2012). Today's evaluators have an increasing set of methodologies and tools at hand. However, faced with increasingly complex operating environments, multifaceted programs, and interrelated issues, the challenge in coming years will be to become more experienced in the use of a plurality of methods *within* broader (and sometimes nested) evaluation designs. Indeed, our acceptance of *mixed* designs, combining different approaches to establishing causality; and our innovating in the use of contrasting methodological approaches, will be fundamental to our ability to effectively capture and understand impact.

## Systems Thinking and Complexity Science

Evaluators will also need to borrow from other disciplines in order to meet the challenges raised in the post-2015 era. The fields of systems thinking and complexity science are increasingly drawn upon by evaluators who are engaged in the challenging task of *understanding* "what works" in complex, dynamic contexts. The principles of these approaches are well-established in many fields, but relatively new to development evaluation (Befani, Ramalingam, and Stern 2015). There is still a significant amount of work to do to understand the applicability and appropriateness of different methodologies, and to adapt and develop the tools used across this incredibly broad field, whether as a heuristic device, or as more complicated forms of social simulation and agent-based modeling. Examples of the application of systems thinking and complexity science within evaluations are limited. Further real-world testing of these approaches is critical to progress in this area (Befani, Ramalingam, and Stern 2015).

## Agility and Flexibility

Traditional public sector oriented evaluations can be a costly and time-consuming activity. Increased agility and flexibility will be essential in order to provide private actors with the information they require for decision making and learning. Within the impact investing market for instance, the current focus on lean systems presents a risk in terms of an overreliance on simplistic numerical summaries.<sup>6</sup> Outcome-level evaluation is expensive and time-consuming, which may be why there are few examples to examine (Brest and Born 2013). Evaluators must work to identify cost-effective tools and approaches that are able to meet this demand while providing robust,

---

<sup>6</sup> This is a trend that mirrors the development of the microfinance sector more than a decade ago, when what could easily be counted obscured the need to capture nonfinancial (social) returns. See Foose and Folan (2016).

high-quality evaluative evidence. This may be partially achieved through the blending of M&E, the employment of continuous data-capture approaches, and considering real-time monitoring and evaluative approaches (Greeley, Lucas, and Chai 2013).

## Big Data and Social Media

Evaluators will also need to capitalize on the flourishing information and communications technology market in order to support increased agility in data collection and analysis. Mobile phone and smart technologies can enable people to communicate more easily, engage in dialogue, and contribute their opinions to evaluations. Technological innovation has also made it possible to generate data on difficult-to-reach populations, such as those in fragile and conflict-affected environments (Bamberger 2016).

To date, there has been very limited use of social media and big data in international development evaluation, which has a greater focus on research, planning, and resourcing. When used appropriately, big data can increase the reliability of findings through the provision of huge data sets; facilitate data collection on sensitive topics and in difficult-to-reach situations; capture a range of stakeholder voices and empower vulnerable groups; support the evaluation of complex contexts and programs; and finally, play an important role in the dissemination of findings (Bamberger 2016).

The digital data revolution has significant potential to provide cost-effective, real-time data, and hence to increase the pace at which information can be generated for decision making and learning. However, there is presently insufficient understanding of and practical experience in using these approaches in development evaluation. Continued engagement from the profession will be required to capitalize on these new opportunities.

## GOING BEYOND METHODOLOGY: BUILDING NEW PARTNERSHIPS

Methodological pluralism alone, while important, will not lead to a significant increase in the relevance and utility of evaluation in meeting current global challenges. Methodology has dominated the last decade of development evaluation, but new demands will inevitably require evaluation to expand beyond the public sector and begin working more closely with the private sector, philanthropic foundations, and INGOs, with a stronger emphasis on dialogue within and between actors. Currently rooted in the public sector, the development evaluation profession is likely to have to expand its communication horizons; learn the language of the private sector and philanthropic organizations; and develop the skills and experience required to interact closely with key stakeholders outside the public sector. It will also be important to communicate the experience, knowledge, and approaches developed over the last decade in a manner that the private sector finds engaging, and that demonstrates a clear value proposition (Jackson 2013). Several such attempts have emerged in recent years—for example, the Social Impact Investment Taskforce, the Wilton Park event on New Frontiers, and ImpCon, among others.

Promoting dialogue between evaluators and the private sector is key to building understanding and demand for evaluative products. As discussed earlier, there is presently a rather nascent demand for evaluative evidence, and a lack of clarity about the value evaluators might bring beyond more mainstream advisory services from management consultants and accountancy firms. Evaluators need to not only be open to and able to meet investor and business requirements: they must “do more than evaluate or critique... they also have to become ‘field builders’ to demonstrate their approach and its value.”<sup>7</sup> A variety of new platforms and events are required to promote discourse on differences in language, expectations, challenges, and values.

Within the evaluation sector, partnerships between developing, middle-income, and developed countries will also need to support the building of evaluative capacity across the globe. The move toward country-led evaluation strategies precipitated by the SDG evaluation agenda will require a substantial increase in national technical capacity, commissioning, and evaluation.

### **THE NEED FOR VALUES-DRIVEN EVALUATION**

The applied nature of evaluative inquiry means that findings should have a direct relevance to decisions, whether program decisions, policy changes, strategic changes in direction, or funding allocations. As stated by Patton (2014), “evaluation is something that informs action.” That is, it is a distinctive form of social science inquiry. The conclusions of an evaluation are expected to judge effectiveness, and to place a value on the subject of inquiry; hence they are of keen interest to stakeholders (Barnett and Camfield 2016). As such, the ethical responsibility of the evaluators extends beyond a focus on more traditional research endeavors, in which a “protection of respondents” (i.e., human subjects) predominates—in other words, a “do no harm” principle. Evaluation now has a perhaps heightened requirement, to also contribute to society (a “do good” principle) through deliberations on policy and resource priorities, as well as the associated focus on transparency, accountability, and participation. This requires evaluators to consider more carefully how evidence and knowledge are created and made available and accessible, in a way that facilitates debate among more than just commissioners and immediate stakeholders (Barnett 2015).

In an emerging profession, where commissioners often have a significant, and sometimes unhealthy, stake in evaluation findings, maintaining independence is a continued challenge, especially where commissioners can exert significant control over the scope, methods, approaches, and outcomes of evaluation (Scott 2016). Furthermore, the growth of private sector and blended modalities in the international development field has given rise to heightened concerns over levels of accountability and transparency. Unlike public sector funds, which are generally subject to rigorous evaluation and public scrutiny, many investors are primarily accountable to shareholders rather than the general public (O’Flynn and Barnett 2016).

---

<sup>7</sup>As concluded in Clarke, Barnett, and van den Berg (2015).



The lack of widespread use of rigorous tools for the assessment of impact, combined with the real risk that the evaluative function is being overtaken by management consultants, financial advisors, auditors and so on, presents a fundamental challenge in establishing the validity of private sector impact claims.

Furthermore, there is a distinct lack of clarity concerning how and to what extent citizens' voices can and should be heard in the process of private sector evaluation (O'Flynn and Barnett 2016). Within the field of impact investing, for example, a shift toward capturing impact at the household or individual level could significantly contribute to the empowerment of those individuals affected, either positively or negatively, by social impact initiatives (Jackson 2013; Clarke, Barnett, and van den Berg 2015). However, a values-driven evaluation profession can, and it is argued should, take a more deliberate stance in addressing the trade-off between methodological rigor and ethical principles such as inclusion (Barnett and Camfield 2016).

## CONCLUSION

In short, we argue that in this changing landscape—one of increased interconnectedness, uncertainty, and new actors—the evaluation field will need to both adapt methodologically and form new alliances and partnerships that transcend traditional development assistance. But even beyond this, we argue that evaluation has an ethical and value-based proposition: after all, evaluation is about “value” and “valuing” performance and impact. Evaluation can be about more than simply presenting evidence: it can be used to engage a range of very different interests, support inclusion, raise the voice of the marginalized, and “speak truth to power.” With the necessity of mobilizing private sector resources to reach the SDGs also comes an evaluator's responsibility to perform a new role in supporting transparency and accountability. For example, in the absence of traditional public accountability mechanisms, development evaluation can play a supportive role, alongside metrics and certification, to help hold the burgeoning private sector to account (Jackson 2013). Other professions (accountants, management consultancies, auditors, certification bodies, etc.) are already addressing these emerging needs, but with less focus on the effectiveness, transformational change, participation, transparency, and accountability that a truly values-based evaluation profession could offer. Therefore, not only do evaluators themselves need to adapt to a changing world: they also have a role to play in helping citizens to adapt to globalization through the better use of evidence. This includes advocating for, and being part of, processes that hold public and private sectors to account for their performance and impact—not just in the short term, but also in longer-term consequences, both positive and negative.

## REFERENCES

- Atkinson, A. 2015. *Inequality: What Can Be Done?* Cambridge, MA: Harvard University Press.
- Avantage Ventures. 2011. [“Beyond the Margin: Redirecting Asia's Capitalism.”](#) Avantage Ventures, Hong Kong.

- Bamberger, M. 2016. "Integrating Big Data into the Monitoring and Evaluation of Development Programmes." UN Global Pulse, New York.
- Barder, O., and B. Ramalingam. 2012. "[Complexity, Adaptation, and Results: Views from the Center](#)." *Center for Global Development Blogs*.
- Barnett, C. 2015. "An Evaluator's Perspective: Imbalances of Power in Evaluation Design and Implementation." *Evaluation Connections* December: 8–9.
- Barnett, C., and L. Camfield. 2016. "Ethics in Evaluation." *Journal of Development Effectiveness* 8: 528–34.
- Befani, B., B. Ramalingam, and E. Stern. 2015. "Introduction—Towards Systemic Approaches to Evaluation and Impact." *IDS Bulletin* 46: 1–6.
- Bourguignon, F. 2015. *The Globalization of Inequality*. Princeton: Princeton University Press.
- Brest, P., and K. Born. 2013. "Unpacking the Impact in Impact Investing." *Stanford Social Innovation Review* 1–14.
- Broegaard, E., B. Bull, and J. Kovsted. 2014. "Evaluation Approaches in Situations of Conflict and Fragility." In *Evaluation Methodologies for Aid in Conflict*, O. Winckler Anderson, B. Bull, and M. Kennedy-Chouane, eds., Chapter 2. London: Routledge.
- Clarke, Charlotte, Chris Barnett, and Rob D. van den Berg. 2015. "[New Frontiers for Evaluation in an Era of Market-Oriented Development](#)." Conference summary. Wilton Park Event, July 20–22.
- Duflo, E., and M. Kremer. 2003. "Use of Randomization in the Evaluation of Development Effectiveness." Paper presented at the World Bank Operations Evaluation Department Conference on Evaluation and Development Effectiveness, Washington, DC, July 15–16.
- Flynn, J., J. Young, and C. Barnett. 2015. "Impact Investments: A Literature Review." CDI Practice Paper, Centre for Development Impact.
- Foose, L., and A. Folan. 2016. "[What Can Impact Investing Learn from Microfinance](#)." Stanford Social Innovation Review blog.
- GIIN (Global Impact Investing Network). 2016. *Impact Investing Trends: Evidence of a Growing Industry*. New York: GIIN.
- Greeley, M., H. Lucas, and J. Chai. 2013. "Real Time Monitoring for the Most Vulnerable." *IDS Bulletin* 44 (2).
- Heider, Caroline. 2017. "[Policy Coherence and the Sustainable Development Goals—What Can Evaluation Do About It?](#)" World Bank Independent Evaluation Office blog, January 3.
- Institute for Economics and Peace. 2014. *Global Peace Index 2016*. New York: Institute for Economics and Peace.
- Jackson, E.T. 2013. "Interrogating the Theory of Change: Evaluating Impact Investing Where It Matters Most." *Journal of Sustainable Finance and Investment* 3: 95–110.
- Nakhouda, S., P. Lucci, C. Bennett, J. Bryant, and C. Watson. 2016. "Our Shared Future: Implementing the Sustainable Development, Climate Change and Humanitarian Summit Agendas." ODI Working Paper, Overseas Development Institute, London.
- O'Flynn, P., and C. Barnett. 2016. "Towards Greater Effectiveness and Accountability in Impact Investing." IDS Policy Brief, Institute of Development Studies, Brighton.
- Ofir, Z. 2015. "Throwing Down the Gauntlet: Lessons That Challenge Evaluation in the SDG Era." Paper presented at the Fourth International Conference on National Evaluation Capacities, Bangkok, October 26–30.
- Ofir, Zenda, Thomas Schwandt, Stefano D'Errico, Kassem El-Saddik, and Dorothy Lucks. 2016. "[Five Considerations for National Evaluation Agendas Informed by the SDGs](#)." IIED Briefing, International Institute for Environment and Development, London.

- Oxfam. 2014. *Even It Up: Time to End Extreme Inequality*. London: Oxfam.
- Patton, M.Q. 2014. "Evaluation Flash Cards: Embedding Evaluative Thinking in Organizational Culture." Otto Bremer Foundation, St. Paul, MN.
- Picciotto, R. 2005. "The Evaluation of Policy Coherence for Development." *Evaluation* 11 (3): 311–30.
- . 2012. "Experimentalism and Development Evaluation: Will the Bubble Burst?" *Evaluation* 18: 213–29.
- . 2015a. "Development Evaluation in Transition: New Priorities, New Coalitions, New Instruments." CDI Practice Paper, Institute of Development Studies, Brighton.
- . 2015b. *The 5th Wave: Social Impact Evaluation*. New York: Rockefeller Foundation Evaluation Office.
- Picketty, T. 2013. *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press.
- Rockström, J., W. Steffen, K. Noone, Å. Persson, F.S. Chapin, E.F. Lambin, T.M. Lenton, M. Scheffer, C. Folke, H.J. Schellnhuber, B. Nykvist, C.A. de Wit, T. Hughes, S. van der Leeuw, H. Rodhe, S. Sörlin, P.K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R.W. Corell, V.J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, and J.A. Foley, 2009. "A Safe Operating Space for Humanity." *Nature* 461: 472–75.
- Rowe, A. 2012. "Evaluation of Natural Resource Interventions." *American Journal of Evaluation* 33 (3): 384–94.
- Savedoff, W.D., R. Levine, and N. Birdsall. 2006. "When Will We Ever Learn? Improving Lives through Impact Evaluation." Report of the Evaluation Gap Working Group. Center for Global Development, Washington, DC.
- Schwandt, Thomas, Zenda Ofir, Dorothy Lucks, Kassem El-Saddick, and Stefano D'Errico. 2016. "[Evaluation: A Crucial Ingredient for SDG Success](#)." IIED Briefing, International Institute for Environment and Development, London.
- Scott, C. 2016. "Cultures of Evaluation: Tales from the End of the Line." *Journal of Development Effectiveness* 8: 553–60.
- Social Impact Investment Taskforce. 2014a. "Impact Investment: The Invisible Heart of Markets. Harnessing the Power of Entrepreneurship, Innovation and Capital for Public Good." Report, Social Impact Investment Taskforce, London.
- . 2014b. *International Development*. London: Social Impact Investment Taskforce.
- Stern, E., N. Stame, J. Mayne, K. Forss, R. Davies, and B. Befani. 2012. "Broadening the Range of Designs and Methods for Impact Evaluations." Department for International Development, London.
- Sumner, A. 2012. "Where Do the World's Poor Live? A New Update." IDS Working Paper, Institute of Development Studies, Brighton.
- UN (United Nations). 2015a. "Addis Ababa Action Agenda of the Third International Conference on Financing for Development." Third International Conference on Financing for Development Addis Ababa, July 13–16.
- . 2015b. "Sendai Framework for Disaster Risk Reduction 2015–2030." UN, Geneva.
- . 2015c. "[Transforming Our World: The 2030 Agenda for Sustainable Development](#)." A/RES/70/1. UN General Assembly, New York.
- UNCTAD (United Nations Conference on Trade and Development). 2014. *World Investment Report 2014: Investing in the SDGs: An Action Plan*. New York: UNCTAD.
- UNDP (United Nations Development Programme). 2013. *Humanity Divided: Confronting Inequality in Developing Countries*. New York: UNDP.

- UNHCR (United Nations High Commission for Refugees). 2015. *Global Trends; Forced Displacement in 2015*. Geneva: UNHCR.
- van den Berg, Rob D., and Lee Cando-Noordhuizen. 2017. "Action on Climate Change: What Does It Mean and Where Does It Lead To?" In *Evaluating Climate Change Action for Sustainable Development*, Juha I. Uitto, Jyotsna Puri, and Rob D. van den Berg, eds. New York: Springer International Publishing.
- Weisen, C., and M. Prokop. 2015. "From Millennium Development Goals to Sustainable Development Goals: Challenges and Opportunities for the Evaluation Community." Paper presented at the Fourth International Conference on National Evaluation Capacities, Bangkok, October 26–30.
- White, H. 2009. "Theory-Based Impact Evaluation: Principles and Practice." Working Paper 3, International Initiative for Impact Evaluation, New Delhi.
- World Economic Forum. 2016. *The Global Risks Report 2016*. 11th ed. Geneva: World Economic Forum.

---

## Chapter 19

# Rising to the Challenges of Impact Evaluation - Insights from Piloting a Systemic and Participatory Approach

Adinda Van Hemelrijck

**Abstract.** *This chapter reflects on the use and value of a systemic theory-based and participatory mixed-methods approach for addressing the challenges of impact evaluation in complex development contexts. A Participatory Impact Assessment and Learning Approach (PIALA) was developed and piloted with the International Fund for Agricultural Development in Vietnam in 2013, and in Ghana in 2015, that engages partners and stakeholders in assessing, explaining, and debating systemic impacts on rural poverty. An action research was conducted around the pilots to learn about the processes and mechanisms that make impact evaluations using PIALA rigorous and inclusive but also feasible. The study concluded that inclusiveness and rigor can reinforce each other, even more so at scale, with sufficient capacity. Methodological complementarity and consistency, extensive and robust triangulation, and cross-validation are important attributes. Investing in research capacity may help to reduce costs over time, while enhancing the value of impact evaluation and the uptake of its findings.*

Development today takes place in globalized contexts of growing inequality, uncertainty, and instability, with new rising powers and an infinite number of conflicting issues and interests. The 2030 Agenda for Sustainable Development calls for fundamental systemic changes, and adds demands for inclusiveness and sustainability to those of effectiveness, in order to eradicate poverty and inequality and protect our planet. Interventions, consequently, are becoming ever more complex, with stakes and stakeholders getting more diverse, influences more dense, problems more systemic, and outcomes more unpredictable. This complexity challenges the field of impact evaluation.

Traditional counterfactual-based approaches are generally found to be too costly and difficult to pursue in complex environments, due to high causal density, spillover, time lags, and the unpredictability of events (Befani et al. 2014; Picciotto 2014). They focus too narrowly on specific intervention components, thus “leaving many evaluation questions unanswered” (White 2014, 3). They also do not explain impact or assess its sustainability, given their focus on specific and isolated cause-effect relationships: therefore they cannot tell if, how, or why similar relations would or would not work elsewhere (Picciotto 2014; Ravallion 2012; Woolcock 2013). Finally, engagement of and learning with partners and stakeholders is inhibited by scientific procedures, raising questions about inclusiveness and democratic value (Van Hemelrijck 2013a, 2017a).

Alternative theory-based and complex systems approaches, on the other hand, tend to be time-intensive and to produce evidence that is not comparable across many cases;<sup>1</sup> therefore, they are not suitable for evaluations with larger populations (a larger  $n$ ) that require estimates of impact distribution (Beach and Pedersen 2013). In addition, those studies that allow for participation generally do not set out to rigorously assess causality and to address concerns of bias and rigor (Copestake 2014; White and Phillips 2012). Chambers calls this “a strange omission, perhaps even a blind spot,” and refers to the Participatory Impact Assessment and Learning Approach (PIALA) in this respect as “part of what should be a wave of the future” (Chambers 2017, 108).

PIALA was developed with the International Fund for Agricultural Development (IFAD) between 2012 and 2015 in an attempt to address these challenges. IFAD is a United Nations (UN) agency that provides loans and support to governments for agricultural and rural development programs that aim at reducing rural poverty by changing smallholder production and market systems (IFAD 2016). These are generally medium to large-scale programs that aspire to create sustainable systemic or transformative change, and are implemented by public and private partners in often quite complex political environments. The PIALA initiative sought not to reinvent the wheel, but to develop a model that creatively combines existing designs and methodologies (both quantitative and qualitative) in novel ways to rigorously assess such complex programs, and to bring participation in impact evaluation

---

<sup>1</sup>This is mostly because the cases themselves are not comparable.

to life (Guijt et al. 2013). Inspiration was drawn mostly from the theory-based (in particular, realist) and transformative (including rights-based) traditions (Holland 2013; Mertens 2009; Pawson 2013; Van Hemelrijck 2013a).

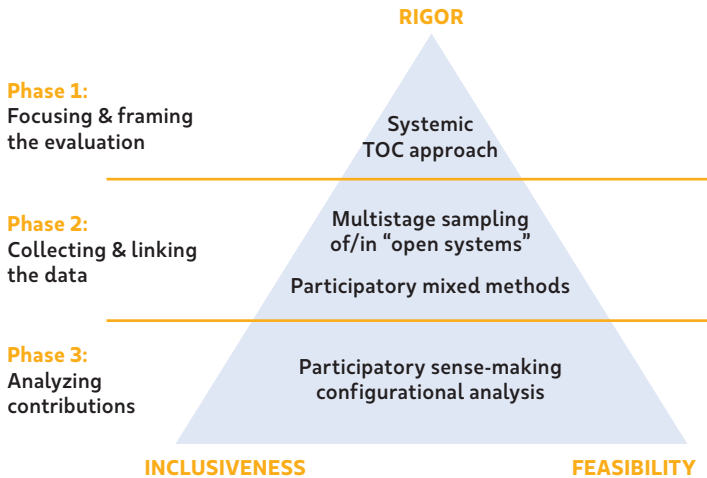
This chapter first describes what PIALA is and briefly presents the two IFAD pilots. It then discusses the main insights from the action research that was conducted around the pilots on how an impact evaluation using PIALA can be rigorous and inclusive. The chapter concludes with some reflections on the value-for-money of the approach, how rigor and inclusiveness may reinforce each other and generate greater value, and the key attributes and conditions for achieving this.

## THE PARTICIPATORY IMPACT ASSESSMENT AND LEARNING APPROACH

PIALA is a theory-based, mixed-methods approach that is essentially participatory. It aims to enable stakeholders to see and learn about impact collectively and systemically, in order to bring about transformative change. It is most suitable for assessing the impact of medium to large-scale projects or programs that are targeting relatively large populations, in contexts where a conventional counterfactual approach is insufficient, difficult, or impossible to pursue. PIALA is not a specific research or evaluation methodology, but an approach that can embed any method and allows for a creative “mixed design” (Stern 2015) combining different evaluation traditions and methodologies, as long as its two overarching design principles—*evaluating systemically* and *enabling meaningful participation*—are maintained (Van Hemelrijck 2016b). These two principles allow for a flexible design, and make it possible for evaluators to adapt PIALA’s five methodological elements to the specific evaluation context and purposes. The five elements follow:

- A **systemic theory of change (TOC)** for visualizing the project’s causal claims, and engaging stakeholders in framing the evaluation and debating the evidence
- **Multistage sampling of/in “open systems”** for enabling systemic inquiry across medium to large-size populations
- Standardized **participatory mixed methods** for collecting, linking, and cross-checking the data in all sampled systems, in a systematic and comparable way
- A two-stage **participatory sense-making** model for engaging stakeholders at local and aggregated levels in debating the emerging evidence
- A **configurational analysis** method for assessing systemic change patterns and drawing conclusions about the distribution and magnitude of their impact across medium to large samples

As shown in figure 19.1, these five elements are designed and put to use in three consecutive phases: framing and focusing the evaluation; collecting and linking the data; and analyzing and debating contributions. To further uphold the desired quality in the design and conduct of an evaluation for

FIGURE 19.1 **PIALA elements and standards**

achieving optimal value within the constraints of available resources, PIALA combines standards of *rigor*, *inclusiveness*, and *feasibility*.

## Methodological Elements

The **systemic TOC approach** forms the backbone for the entire evaluation. It is designed in the first phase of the evaluation process, by means of which the evaluation is focused and framed. It involves a process of reconstructing and visualizing a program’s impact pathways and change hypotheses, and the broader trends and influences, based on a thorough desk review and discussions with key stakeholders. Unlike a classic program/project theory,<sup>2</sup> this approach uses an evaluative lens, assessing the hypotheses by looking backward, from the envisioned impact back to the interactions and mechanisms that presumably have caused or influenced the impact (Funnell and Rogers 2011; van Es, Guijt, and Vogel 2015). Moreover, it views impact from a systemic perspective, resulting from changes in systems of interactions, rather than the direct and isolated relationship between intervention and effect. A systemic TOC approach is most useful for evaluating the changes caused by many different interventions, implementers, contributors, and funders, because it helps to create a shared understanding of complex pathways, and enables different stakeholders to critically engage in parts of the analysis (Van Hemelrijck 2013a).

<sup>2</sup> A program/project theory is constructed from a management perspective, and is focused on strategy and performance looking forward, toward the delivery of planned results.



**Multistage cluster sampling of/in open systems** happens right after the evaluation focus and framing is agreed upon with the stakeholders, as part of the design for the second phase. Its purpose is to ensure sufficient representation of the various populations, in order to enable the comparison and generalization of findings about systemic impact at the medium-to-large scale. If we want to learn about systemic impact, then the system should be the main level of analysis, and thus also the main sample unit (Lain 2016). In the case of government policies and programs, the system is generally too monolithic for a classic counterfactual comparison. However, by focusing on the lowest embedded open system at the local level (e.g., the local supply-chain system) that is entrenched in and affected by the larger system (e.g., the larger commodity value chain, and national policy framework), it is often possible to have a sample that is large enough to cover systemic heterogeneity, and to have large enough subsamples for statistical comparison. Contrasting evidence can then be obtained from the areas where program mechanisms are found absent, dysfunctional, or ineffective, rather than from predetermined control areas that are sampled external to the program (Van Hemelrijck 2017b). Multistage cluster sampling of these local systems, and of populations within these systems, is the most cost-effective method, as it substantially reduces costs and logistics compared to other random sampling strategies (Levy and Barahona 2002).<sup>3</sup>

The appropriate **selection and mixing of methods** to collect qualitative and quantitative data on the different causal claims in the TOC is also part of the design of the second phase. The IFAD pilot in Ghana combined conventional household surveys for statistical poverty analysis; participatory methods for generic change analysis, livelihood analysis, and constituent feedback; SenseMaker for quantitative pattern analysis of perceptions; and key informant semistructured interviews for inquiring the larger system.<sup>4</sup> Methods are selected specific to the causal links in the TOC, and are used on an equal basis. They complement and build on each other analytically to enable the construction of the actual causal paths with the data for each locality or sampled

---

<sup>3</sup> Random sampling is needed for statistical analysis. This depends on the evaluation focus. In an impact evaluation for Oxfam GB in Myanmar, for instance, PIALA's sampling protocol was adapted to fit the specific evaluation focus and requirements, which did not require statistical analysis and thus also not a random sampling (Van Hemelrijck 2017a).

<sup>4</sup> Constituent Feedback (also called Constituent Voice) is a methodology developed by Keystone Accountability (<http://www.keystoneaccountability.org>) for collecting quantified feedback and engaging in dialogue with key constituents or beneficiaries, using standardized metrics similar to the customer satisfaction surveys developed in the private sector, and descriptive statistics to produce visual data reports. SenseMaker is a software-based methodology developed by Cognitive Edge (<http://www.sensemaker-suite.com>) that facilitates mass ethnography and provides a way of nearly real-time mapping of social interactions and individual perceptions and motivations to inform adaptive management and policy formulation. It collects large amounts of self-signified micro-stories that capture people's experiences and perceptions of past and future change in ways that enable us to identify emerging patterns of actions and decisions. The software permits statistical analysis at a very large scale.

system, mirroring the envisioned paths in the TOC. They also partly overlap, to permit triangulation. Systematic data collation and quality monitoring makes it possible to cross-check and link the data for building the causal paths during fieldwork in every locality, and to timely identify data gaps or weaknesses that need further inquiry before moving to the next locality. To enable comparison across the sample of systems as the basis for aggregating findings, the methods are more or less standardized. Yet they also remain sufficiently open-ended by including sensing tools such as causal flow mapping that can capture unintended effects and influences, and uncover broader dynamics that are interacting with the program (Van Hemelrijck 2015).

**Participatory sense-making** occurs in the third phase of the process: analyzing and debating contributions. It involves half-day local workshops with 30–50 participants (of whom 60–70 percent are intended beneficiaries) during the fieldwork in each locality, and a one or two-day program-level workshop with 100–130 participants (of whom over 30 percent are intended beneficiaries), shortly after finishing the fieldwork and before turning to the final analysis and reporting. The stakeholders participating in the workshops typically include decision makers, service providers, and intended beneficiaries. They proportionally represent all the different perspectives necessary to cross-validate the evidence and inform the final analysis. They discuss the evidence together and assign value to observed contributions (among other influences) by comparing the actual causal paths revealed by the data with those hypothesized in the TOC. Participatory sense-making in all researched localities and at the aggregated level serves to not only cross-check and strengthen the evidence, but also to create ownership, enable equal voice, and stimulate systemic learning. In essence, it makes an evaluation more democratic (Van Hemelrijck 2016b).

Finally, the **configurational analysis** compares systemic change and impact across the sample of systems to reach conclusions about the distribution and magnitude of impact. Its tools are designed and put to use in the third phase of the evaluation process, after the sense-making. It employs elements of process tracing, contribution scoring, and cross tabulation, and involves four major steps. The first is the aggregated data collation in a standard Excel matrix format, in which all evidence from the field collation matrices as well as secondary sources is synthesized and tabulated alongside the TOC. The next step involves the clustering of the evidence across all the sampled systems to surface patterns or configurations of systemic changes and causal attributes. The third step involves the comparative analysis of similarities and differences in configurations for the specific mechanisms or parts of the system of interest (including cases with and cases without functioning mechanisms).<sup>5</sup> The final step involves integration of the findings for

---

<sup>5</sup> Software such as EvalC3 can be applied to assess the conjuncture of different mechanisms and causal processes. This novel software, developed by Rick Davies, was piloted in an impact evaluation using PIALA for Oxfam GB in Myanmar (Van Hemelrijck 2017a). The software helped to identify sets of causal attributes that are necessary and/or sufficient for specific sets of outcome attributes to occur, and to compare and evaluate the performance of these causal models to find those with the greatest predictive power.

the different parts and mechanisms as the basis for validating (or refuting) the hypotheses in the TOC; zipping up the findings alongside the TOC; and drawing conclusions about the distribution and magnitude of the program's contributions to impact (Van Hemelrijck 2016a, 2016b, 2017b).

## Quality Standards

To enable an evaluation to achieve optimal value with limited budgets while remaining true to the two core principles of systemic thinking and meaningful engagement, PIALA also features a quality framework that includes standards of *rigor*, *inclusiveness*, and *feasibility*.

**Rigor** is defined as the quality of thought put into the methodological design and conduct of the evaluation in a way that enables robust triangulation of different methods and perspectives in order to defeat bias or dominance of a single truth; and to ensure both consistency and responsiveness to local contexts and conditions (Van Hemelrijck 2016b). Whereas mainstream evaluation practice defines rigor as the controlled avoidance of bias through statistical procedure, PIALA builds on the premise that bias cannot be avoided by a single method or procedure, but can be mitigated through systematic triangulation of different methods and perspectives (Camfield, Duvendack, and Palmer-Jones 2014; Carugi 2016; Mertens 2010).

**Inclusiveness** refers to the legitimacy of the ways in which people are engaged in the evaluation, and to the level of impartiality or inclusion of all stakeholder views and perspectives. This has intrinsic empowerment value but also contributes to the robustness and credibility of the evidence and thus to the validity of the findings (Chambers 2015; Pawson 2013). Validity is understood as the extent to which findings are well founded, based on robust evidence, and correspond with the reality of all stakeholders, in particular the populations affected by the project or program being evaluated. By embracing a wide range of stakeholder perspectives and ensuring their equal weight in examining the evaluation questions, the evaluation builds a more accurate systemic picture of impact. Meaningful engagement in constructing, analyzing, and debating this picture, on the other hand, enables equal voice, and contributes to empowerment (Chambers 2017).

**Feasibility** concerns the budget and the capacity needed to meet the expectations of rigor and inclusiveness, and to enhance stakeholders' systemic and collaborative learning (Van Hemelrijck 2016b). The investment in building the capacity of in-country researchers, and in experimentation with novel mixed designs that stretch the limits of conventional evaluation practice, is critical for doing this at a larger scale. Considering feasibility as an explicit and intent-driven (rather than constraint-driven) quality helps to think of this investment in a positive way. Much remains to be learned about how to do it well. Excessive focus on limiting costs starves the evaluation of the oxygen it needs in order to deliver on rigor and inclusiveness and to maximize its value (Chambers 2017).

## TWO IFAD EVALUATIONS THAT PILOTED PIALA

PIALA was piloted in the evaluation of two IFAD-funded programs: the Doing Business with the Rural Poor (DBRP) program conducted in one province in southern Vietnam,<sup>6</sup> and the Roots and Tubers Improvement and Marketing Program (RTIMP), which was conducted countrywide in Ghana.<sup>7</sup> Both were aimed at improving livelihoods and increasing food and income security by enhancing smallholders' capacities to commercialize, and by linking local businesses to markets and industries. DBRP focused on developing diversified short-value chain systems; RTIMP was concerned with developing much longer commodity chains, linked to national and export markets and industries (Guijt et al. 2014; Van Hemelrijck and Kyei-Mensah 2015). Both programs essentially sought to create the mechanisms needed to facilitate rural peoples' access to services, resources, and markets.

Despite some important differences in the context and quality of the two evaluations,<sup>8</sup> both produced quite convincing evidence of program contributions to the improvement of livelihoods as a result of the increased access to services, resources, and markets generated through these mechanisms. The RTIMP evaluation in Ghana, for instance, showed significant improvements in roots and tubers-based livelihoods, with 15 percent of households increasing their incomes above \$2 a day. Very weak or no improvement was found in supply chain areas where the RTIMP mechanisms were dysfunctional or absent. Although positive, the evidence also showed that these improvements were rather limited, fragile, and susceptible to climate and market shocks, particularly for poor and vulnerable households, and in remote and marginalized areas. The improvements in roots and tubers-based livelihoods in Ghana occurred merely between 2009 and 2013, and in about 52 percent of the supply chain areas, or about half of the country. Moreover, no households gained profits above \$4/day from roots and tubers, even though 61 percent

---

<sup>6</sup> The DBRP was implemented from 2008 to 2014 in two provinces (Cao Bang and Ben Tre), with a total budget of \$51 million, including a \$36 million loan from IFAD. The evaluation (Guijt et al. 2014) was conducted in 2013 at a cost of \$90,000 in Ben Tre province only, where the project was implemented in 50 of 164 communes in eight of nine districts.

<sup>7</sup> The RTIMP was implemented from 2007 to 2015 as a national program in 106 of 216 districts, spread across all 10 regions countrywide, with a total budget of \$24 million, of which \$19 million was financed under an IFAD loan. The evaluation (Van Hemelrijck and Kyei-Mensah 2015) was conducted countrywide after project completion in 2015, at a cost of \$233,000, and covered the post-midterm review period from 2010.

<sup>8</sup> Although the evidence suggested strong connections between all observed changes, confidence in causal inference remained relatively weak in the Vietnam evaluation. In this first PIALA pilot, data collation, cross-checking, and quality monitoring was not yet done systematically with the TOC as a backbone structure. Confidence in inference and generalizability was much stronger in the second pilot in Ghana because of its systematic and multilayered triangulation and cross-validation procedure (Van Hemelrijck 2015). This is further discussed in the next section of this chapter.

of the households had invested in roots and tubers businesses. Access to new seeds and farming technologies had initially led to a boom in production across the country, triggering a spillover into processing. Adoption of new processing technologies, though, remained limited in 83 percent of the cases, partly due to limited investment capital. By and large, the finance mechanism put into place by the program proved inaccessible, as it required pre-investment without short-term capital return, posing high risks for smallholders (Van Hemelrijck and Kyei-Mensah 2015). Both in Ghana and in Vietnam, poor and vulnerable households ran considerable risks by engaging in value chains and accessing markets (Guijt et al. 2014; Van Hemelrijck and Kyei-Mensah 2015). These risks were left largely unmitigated due to inadequate market linking and forecast that otherwise could have helped avoiding the observed local market saturation and monopolization; and inadequate poverty targeting that should have made the support services and mechanisms more inclusive and sustainable (Van Hemelrijck 2016a). Recommendations for how to address these issues in similar IFAD-funded programs and projects were made by these two evaluations.

## KEY INSIGHTS FROM THE PILOTING

As mentioned earlier, the PIALA initiative was conceived as an action research to inquire into the conditions, processes, and decisions affecting the rigor and inclusiveness of the two pilots. The action research combined multisited ethnography with cooperative inquiry, and involved extensive reflections with researchers and participants in the two pilot countries, as well as feedback sessions with global experts at IFAD headquarters (Van Hemelrijck 2016b). Insights from the first pilot (in Vietnam) helped to better address the challenges in the second pilot (in Ghana) (Van Hemelrijck 2015). This section summarizes some of the key lessons learned.

### Creating Ownership of the Evaluation

In order to create ownership, key stakeholders need to be sufficiently engaged in the framing and focusing of the evaluation. Ownership implies that the evaluation is wanted, legitimized, and enabled by a shared sense of responsibility for its success. Ownership also enables participation in the analysis, and facilitates learning and greater uptake of evaluation findings and recommendations (Burns and Worsley 2015; Patton 2011). In the case of PIALA, stakeholders are engaged in the framing and focusing of the evaluation through a process of reconstructing and visualizing the TOC (Van Hemelrijck 2015).

In Vietnam, insufficient time and budget was spent on this process, which affected the rigor and inclusiveness of the approach during the entire evaluation. A brief workshop was organized with the program steering committee and managers to discuss program logic and expectations. The process of reconstructing and visualizing the TOC, however, happened *after* the workshop, and independently of the evaluation design. Evaluation questions did not focus on the causal links and assumptions in the TOC, which made it

difficult for the researchers to relate the evidence back to the TOC and arrive at greater precision in causal analysis. Furthermore, limited ownership of the TOC by the stakeholders hindered their critical engagement in sense-making and contribution analysis (Van Hemelrijck 2013b).

Learning from the Vietnam pilot, the TOC process was made a priority and a key deliverable in Ghana. The researchers organized a design workshop to discuss the TOC and design options, and to determine the focus of the evaluation together with key stakeholders. The investment in a more robust and collaborative TOC process bore fruit and laid the foundation for attaining greater quality throughout the entire evaluation, resulting in stronger evidence and ownership of findings (Van Hemelrijck 2015, 2016b).

### Deciding on the Scope and Scale of the Evaluation

*Scale* refers to the size of the sample of the primary sample unit. In the case of PIALA, this was the lowest embedded “open system” that the program sought to change, to generate impact. *Scope* refers to the various components and mechanisms of the system that the evaluation should cover. Generally speaking, the larger the scale, the more relevant the findings for reporting and advocacy will be. Using participatory mixed methods at scale, however, is challenging, and requires sufficient capacity and resources. When research capacity is weak, more resources are needed for training, coaching, and supervision in order to uphold quality (Van Hemelrijck 2015).

Three relevant design options are available for designing an impact evaluation: full scope–limited scale, limited scope–full scale, and full scope–full scale.<sup>9</sup> When choosing a **full scope–limited scale design**, the emphasis is on learning about the project’s total contribution to impact in select cases, under specific conditions. Fieldwork and analysis are less resource-intensive, given the relatively small sample sizes. Yet evaluation findings will not be generalizable for the entire population: therefore they are less useful for influencing policy decisions.<sup>10</sup> With a **limited scope–full scale design**, the purpose is to assess the effects of one or two particular aspects or mechanisms of the project. The TOC is not strictly necessary in order to conduct such a narrow study, but skipping the TOC process may risk missing out on systemic understanding, leading to flawed conclusions. Components are studied in isolation, which does not permit analysis of systemic interactions. For example, a cost-effectiveness study of Farmer Field Forums (FFF) in Ghana recommended a scaling-up, as the adoption of new technologies had proven the success of this mechanism. The PIALA evaluation, however, showed that in a

---

<sup>9</sup>A limited scope–limited scale option is not really relevant for impact evaluation, as it limits the possibility of causal analysis through classic counterfactual comparison, frequency statistics, and/or triangulation and cross-validation of sources and methods.

<sup>10</sup>This does not hold true if the project/program itself is implemented at a limited scale (small *n*), in which case larger within-samples and more stringent triangulation and cross-validation procedures will take up the resources needed to attain the required level of rigor for generalization.

period of weak economic growth, this success in fact contributed to market saturation, which negatively affected livelihoods across the entire country (Van Hemelrijck 2015; Van Hemelrijck and Kyei-Mensah 2015).

In Vietnam, a choice was made for a full scope–limited scale design, but with disparate scales for the different methods. To save resources, participatory methods were employed only in a subsample drawn from the sample of villages where the statistical household surveys were conducted. The assumption was that this would be sufficient to conduct a full scope inquiry of contributions to impacts on rural poverty for the entire program area. However, it generated a disparity in the data sets that caused problems for their subsequent linking. While participatory data on causes and contributions came from only a few villages or cases, survey data on household impact were more widely distributed and not related to the specific cases or villages covered by the participatory methods. This hindered causal inference (Van Hemelrijck 2013b).

In Ghana, by contrast, a conscious choice was made to employ all methods in the same sample and at the same scale. The three design options were discussed with clients and commissioners before any procurement or design work was started, giving them a basic understanding of the cost and value of each.<sup>11</sup> As the future Ghana Agriculture Sector Investment Program (GASIP) was expected to scale up most of the RTIMP mechanisms, the evaluation was found necessary for both reporting and learning. The commissioners therefore chose the most comprehensive design: **full scope–full scale**. This implied six weeks of uninterrupted fieldwork—much longer and far more intensive than the pilot in Vietnam, where fieldwork took only two weeks. The budget was tighter than in Vietnam because of the larger scale and scope, but quality was upheld by a competent research team (Van Hemelrijck 2015; Van Hemelrijck and Kyei-Mensah 2015).

## Deciding on the Counterfactual

Mainstream impact evaluation assumes that comparative analysis of evidence from both treated and nontreated locations is feasible and necessary in order to assess causality and reach generalizable conclusions about impact on rural household poverty. However, in most “real-world” evaluation contexts (Bamberger, Rugh, and Mabri 2012), it is very difficult and costly to arrive at an accurate assignment of locations to specific interventions and identify credible control groups. The challenge occurs, for instance, in cases of unexpected or uncontrolled project expansion and/or spillover, combined with high causal density of other interventions and influences. In such contexts, it is difficult to discern project from nonproject localities, and to find the right matches (Woolcock 2009). In addition, the open systems that form the principal sample unit in PIALA generally do not have clear boundaries such as villages or other administrative units have. Hence the identification and

---

<sup>11</sup> Including Ghana’s Ministry of Food and Agriculture and the IFAD Country Office in Ghana.

matching of control units for these systems and subsampling of various populations from these systems, if even possible, requires fieldwork prior to the evaluation that substantially increases both costs and risks (Chambers 2017).

In the Vietnam pilot, comparative analysis of treated and nontreated units was considered both possible and necessary for assessing household-level impacts, and the village was thought to be the best proxy unit for investigating the short value chains developed by the program. These assumptions were flawed and compromised in terms of analytical rigor, making it difficult to generalize the findings, for three important reasons. First, without a clear definition and identification of the value-chain systems, and thus without having sampled proper proxies based on such a definition, it was difficult to relate the data on changes in capacities, institutions, and livelihoods to the specific value chains and to assess the causal links. Second, the matching of treated and nontreated villages was based on variables that applied to the village as a unit, not to the value chains, again making it difficult to relate the difference revealed by the data to the interventions. Finally, the high heterogeneity in program delivery and incoherence in its value-chain linking efforts, further conflated by the high causal density of other programs and influences in the villages, made it impossible within available budgets to obtain credible control data (Van Hemelrijck 2013b).

Learning from this experience, in Ghana much more work was done to understand and define the principle sample unit. In the evaluation design workshop, the decision was made not to waste resources on identifying and inquiring control groups of households, but instead invest all in the systemic inquiry of the four main commodity supply chains developed by the program (gari, plywood cassava flour, high-quality cassava flour, and fresh export yam). These commodity chains comprised medium to large amounts of supply chains spread over the entire country. The supply chains are loose catchment areas comprising clusters of communities of smallholders supplying the raw produce, and small enterprises or off-takers acting as “supply chain leaders” and manufacturing higher-value products for bigger markets. The supply chains were not entirely homogeneous, as they interacted and overlapped. Hence they often differed in reality from what was sampled on paper. Ensuring that the data collected on these systems remained comparable required much creativity and coordination. Furthermore, no reliable lists of households and beneficiaries were available for the subsampling of farmers, processors, and households within the catchment areas of the sampled supply chains. Identification and matching of control units and sampling of households thus would have required extensive pre-evaluation fieldwork, and the sponsors and other participants in the design workshop voted against this. Instead, a configurational analysis method, which uses heterogeneity in the sample of systems as the basis for counterfactual analysis, was developed. Supply chains with different systemic configurations of treatments and causal attributes were randomly sampled (with probability proportional to size) from the four commodities’ supply chain populations. The samples were large enough to include supply chains with dysfunctional or absent program mechanisms that could serve as a “natural” counterfactual (Van Hemelrijck and Kyei-Mensah 2015).



## Maintaining Independence

In order to avoid positive bias, field mobilization of research participants is best undertaken independently from project management.<sup>12</sup> When research participants suspect that the research is not independent, they are more likely to over- or underreport. On the other hand, they are unlikely to trust outsiders who are not authorized and formally introduced by their leaders. Thus, for the researchers to organize fieldwork at scale and mobilize participants without any help from the program, they need to be good at logistics; know the areas and local customs; and be able to obtain authorization and introduction in ways that do not affect their independence. In contexts where this is not possible, strong facilitation skills are needed to minimize undue influence or interference (Van Hemelrijck 2015; Van Hemelrijck and Kyei-Mensah 2015).

The challenges encountered in Ghana were quite different from those in Vietnam. In both pilots, though, participants trusted the researchers' authorization and independence, which made them feel safe about expressing their views and critically engaging in the group inquiries. In Vietnam, field research cannot be conducted without government permission and interference. Hence, in the DBRP evaluation, local transportation and mobilization was organized by local officials and program staff, which was highly efficient, but challenging in terms of independence. Local leaders and program staff were quite collaborative during fieldwork but omnipresent. The researchers artfully managed to maintain sufficient distance, though, and to safeguard the privacy of the focus groups (Van Hemelrijck 2013b).

In Ghana, the researchers took care of the transportation and mobilization entirely by themselves but without prior notification or engagement of the local officials, allowing for much greater independence. Staff and officials were present only at the discussions to which they were invited. This, however, made them more suspicious of and resistant to the evaluation. Also, the scale of the fieldwork, the remoteness and spread of the communities, the large distances to travel over poor roads, and the difficulty of finding safe and trusted locations for convening people from different communities, made the field inquiries quite onerous. Independence thus came at a substantial effort and cost in Ghana, but compromise in both rigor and inclusiveness was avoided (Van Hemelrijck 2015; Van Hemelrijck and Kyei-Mensah 2015).

## Contextualizing Poverty Analysis

To make it possible to say something about a program's influences on poverty, data on those influences, and on poverty, need to be linkable. Also, poverty has to be defined in ways that are relevant to the specific context and conditions of the villagers in the program area. In both of the IFAD pilots, a Participatory Rural Appraisal (PRA) ranking tool was used for identifying locally relevant

---

<sup>12</sup> According to the Organisation for Economic Co-operation and Development, "independence" implies an evaluation process that is transparent, independent from project management, and free from political influence or organizational pressure (OECD DAC 2010, 25).

characteristics of wealth and well-being, and assessing changes in relative poverty status (Van Hemelrijck 2015). This tool helped create a shared understanding among the participants of their wealth and well-being as the basis for a causal flow-mapping exercise of the changes they had experienced. It also enabled cross-checking and linking of the participatory data with the household survey data on poverty.

The characteristics of wealth and well-being that were obtained from the participatory ranking exercise, however, were not used to design the household survey. For this, the participatory data collection should have happened prior to the evaluation, which would have increased the cost of the evaluation. It is unclear if, and to what extent, this might have generated more rigorous findings on poverty impact, and therefore have justified the extra investment.

In Vietnam, the survey focused only on IFAD's generic poverty indicators, and used the purely income-based, absolute poverty categories of the Vietnamese government. These proved inadequate for assessing changes in poverty status related to the program. Learning from this, greater efforts were made in Ghana to ensure that the household survey questionnaire was sufficiently attuned to the program context and to on-the-ground realities. Poverty characteristics corresponding with IFAD's poverty indicators were selected from the Ghana Living Standards Survey 2009–14 for assessing the households and computing the categories of poverty status (applying a proxy means test and principal components analysis). And here, no major differences were found between the characteristics obtained from the participatory ranking exercise and those used by the household survey (Van Hemelrijck 2015; Van Hemelrijck and Kyei-Mensah 2015).

Arguably, greater rigor could have been obtained in the findings on poverty distribution and impact in Ghana if the questionnaire had asked about household characteristics in much greater detail. Yet more lengthy surveys cost more, and also increase the risk of fatigue and gaming on the part of both respondents and researchers (Chambers 2008; White 2015). Therefore, in both Vietnam and Ghana, the duration of the household survey was kept to a maximum of no longer than 20 minutes. Also the time length of the participatory group discussions were limited to a maximum of two hours (Van Hemelrijck 2015).

Thus, instead of spending more resources on collecting and analyzing participatory poverty characteristics, or more fine-grained quantitative data on household characteristics to identify poverty categories prior to the evaluation, in Ghana the choice was made to keep the poverty analysis short and instead create room for participation that was more meaningful to the participants. This is what Chambers (2017) calls "appropriate imprecision." The group-based causal flow mapping exercises were found particularly useful by the participants, as it helped them to recall and understand the changes from a systems perspective, and enabled them to engage in collective sense-making with other stakeholders. The assumption is that this contributes to the ability of people to understand and navigate the system within which they are operating, and thus to their empowerment (Burns and Worsley 2015; Merrifield 2002).

## Dealing with Power and Bias

All methods are susceptible to bias, and biases may occur in every phase of the evaluation, from the design to the analysis. Participatory methods, however, are considered more vulnerable than traditional survey-based methods, as they collect perceptions, meanings, and interpretations instead of hard numbers. Yet surveys generating hard numbers are also designed and conducted by human beings with value judgments (Camfield, Duvendack, and Palmer-Jones 2014; Copestake 2014; White and Phillips 2012). Survey questionnaires tend to reflect the assumptions of the designers, while qualitative interviews and participatory inquiries make room for the assumptions of the researched. Both may result in desirability or courtesy bias: the researched tell the researchers what they believe is expected from them (Chambers 2017; White 2015). To overcome such bias, the PIALA evaluations employed participatory mixed methods in a way that enabled extensive and systematic triangulation of different methods and perspectives, and cross-validation of the findings at scale.<sup>13</sup>

Scale, however, can also create bias, as it requires standardization that tends to reduce participation to power-blind procedure (Gaventa and Cornwall 2006; Mosse 2001). In both pilots, attempts were made to avoid this by carefully thinking through how to arrange and facilitate the group processes in order to equal out power imbalances, and by using tools that inherently empower people (Van Hemelrijck 2015). Visual tools were used, such as causal flow and relationship mapping, that enable participants to see *how* data is constructed, and to flag where things are flawed. Appropriate group composition further helped outnumber those who tend to dominate and empower those with less-heard voices, by means of majority or what Chambers (2017, 102) calls “the democracy of the ground.” Of course there is always a danger that more powerful and influential participants will dominate the discussions. Additionally, there is internalized power: the social norms and values that make certain groups believe in and accept their subordination and “voicelessness” (Kabeer 1999). Good facilitators know how to overcome this, and how to “empower through behaviour and attitudes” through careful listening and sharp observation of motives and interactions (Chambers 2017, 122). The researchers in Vietnam and Ghana were trained in this, and

---

<sup>13</sup> *Triangulation* is a principle social science technique that involves the use of more than one type of information or data source, method, and even theory and researcher, for the purpose of crosschecking in order to overcome weaknesses and biases and thus obtain greater credibility of and confidence in findings (cf. <http://www.betterevaluation.org/en/evaluation-options/triangulation>). In PIALA, this goes beyond merely verifying findings: it values different views and perspectives and crosschecks them to build a rich and comprehensive picture of the change processes as the basis for identifying and checking all plausible explanations for causality. *Cross-validation* in the case of PIALA is understood in “realist evaluation” terms as the practice of (dis)confirming findings across multiple independent inquiry cases to strengthen the explanatory power and the confidence in the conclusions about causality and contribution (Pawson 2013).

provided with detailed guidance on facilitation for each of the participatory methods and tools (Van Hemelrijck 2016b).

Rigor then emanates from the combination of good facilitation, and systematic triangulation and cross-validation (Chambers 2017). For the latter, a strict procedure was developed, involving six essential steps:

- 1. Quasi-standardized data collection**, using at least two different methods per type or area of change in the TOC, and each with a minimum of two different sources or groups of people
- 2. Daily reflections on the quality of the processes** by which each of the methods was used, and of the quality of the data generated by these processes
- 3. Triangulation of the data** from the different methods and sources, and identification of data gaps and weaknesses, through systematic data collation alongside the TOC, and scoring of confidence in the emerging evidence, followed by additional data collection, if and where needed, to address the gaps and weaknesses
- 4. Participatory sense-making of the evidence in each locality** together with the local stakeholders, to address remaining gaps and contradictions, and to cross-validate initial findings
- 5. Aggregated collation and crosschecking** of all of the evidence and scorings obtained from all the localities
- 6. Participatory sense-making of the evidence at the aggregated level** together with local and program-level stakeholders, to cross-validate initial findings for the entire program area, and to value program contributions to impact

In Vietnam, this procedure was not yet fully developed, and was found challenging by the researchers. The researchers, most of whom came from a quantitative research background, struggled with triangulation as a way to compile a multiperspective picture, and they were unable to uphold a daily practice of critical reflection on quality and process. Moreover, the tools and guidance for the data collation and quality monitoring proved insufficient (Van Hemelrijck 2013b).

In Ghana, by contrast, the researchers had a mixed background and substantial experience in participatory research. Data collation and quality monitoring were undertaken daily and systematically. A standard set of questions guided the daily reflections on inclusiveness of the processes, and the quality and sufficiency of the data, and a standard table structured around the causal claims and links in the TOC was used for data collation and triangulation (table 19.1). Methods were tightly focused on the causal links, making the triangulation much more straightforward and systematic. A Likert scale rubric was used to score the relative strength of emerging evidence for each of the causal links in the TOC. Since they were better equipped to handle large amounts of data, these researchers were able to finish the data collation, and to identify data gaps and weaknesses in each locality in good

time, and to be well prepared for the sense-making workshops (Van Hemelrijck 2015).

## Deciding on the Scale and Level of Engagement in the Sense-Making

Participation in evaluation is purely extractive if the findings are not returned to the participants and there is no opportunity for them to contest and debate them (Gaventa 2004; Mohan and Hickey 2004). Using PIALA's sense-making model, six village-level workshops with 180 participants, and one provincial-level workshop with 100 participants were organized in Vietnam. In Ghana, there were 23 district workshops with 650 participants, and one national workshop with over 100 participants. The participants in the workshops were sampled purposively from the research participants (Van Hemelrijck 2015).

The outcomes of the participatory sense-making were twofold. First, an additional layer of detail and confirmation of evidence was obtained from the cross-validations in the local and aggregated workshops, adding to the rigor of the evidence, and thus to the validity of the findings. Second, shared ownership was created of both the evidence and the findings, which contributed to the evaluation's inclusiveness and empowerment value. Having participated merely in data collection, people walked into the workshops knowing and owning little: but they left the workshops with a comprehensive picture of the systemic changes and the issues that the evidence had revealed, as well as of stakeholders' various perspectives on these.<sup>14</sup> Critical to the success of the participatory sense-making was both the scale of the workshops, and the way in which they were designed and facilitated. Special competencies are required particularly for doing this at scale. When operating with low capacity and on a shoestring budget, the amount and size of the workshops may need to be trimmed, at the expense of both rigor and inclusiveness (Van Hemelrijck and Guijt 2016).

A truly participatory sense-making process implies the equal and active engagement of all stakeholders. Dynamic environments were created, long presentations by experts were banned, and the various types of evidence were made available in accessible (including visual) formats. Small-group discussions were held, ensuring that people felt "listened to" rather than just "talked at" (Newman 2015). Beneficiaries constituted more than 30 percent of the participants in the provincial and national workshops, and 60–70 percent in the local workshops, giving them sufficient weight in debates with decision makers and service providers.<sup>15</sup>

---

<sup>14</sup> The survey and reflections held at the end of each workshop revealed a high degree of satisfaction among the participants, and of the knowledge and insights gained by them that they found useful for future individual or collective action.

<sup>15</sup> Their group must be larger in the local workshops because they form the primary target group of the project at the local level, while at the aggregated level the primary targets are policy makers and service providers.

TABLE 19.1 **Standard table used in the RTIMP evaluation for within-case data collation and triangulation**

Causal link	Secondary data	Primary QUANT data	Primary QUAL data	Strength of evidence <sup>a</sup>	Strength of causality <sup>b</sup>
<b>Impact Claim – Poverty Reduction</b>					
I2→I1	Insert main findings from the 2010 Ghana Living Standard Survey report, RIMS baseline and the RTIMP M&E	Insert main findings from the household survey			
O3+O2+O1→I2			Insert main findings from the generic change analysis with community members		
<b>Contribution Claim 1 – Enhanced Market Linking</b>					
M1c+M1b+O2+O3→C1b	Insert main findings from the DDA reports, RTIMP Enterprise Record Books, ZOCs progress reports, MoFA and DADU OAs, and the RTIMP M&E	Insert main findings from the livelihood analysis and SenseMaker study with intended beneficiaries (farmers and processors)			
C1a+(M1)→O1 C1b+M1a→C1a			Insert main findings from the KIIs with DDAs, BACs, SCFs, GPCs, food traders and exporters		
		Insert main findings from the Constituent Feedback sessions with intended beneficiaries (farmers and processors)			

(continued)

TABLE 19.1 (continued)

Causal link	Secondary data	Primary QUANT data	Primary QUAL data	Strength of evidence <sup>a</sup>	Strength of causality <sup>b</sup>
<b>Contribution Claim 2 – Enhanced Production of Roots and Tubers</b>					
C2a+C2b→O2	Insert main findings from the RTIMP productivity surveys, progress reports from the SRID, GLDB, DDAs and ZOCs, and the RTIMP M&E	Insert main findings from the livelihood analysis with intended beneficiaries (farmers and processors)	Insert main findings from the KIs with FFF facilitators, extension agents, DDAs, DADU officers, and CSIR, KNUST & UCC research leaders		
M2a+M2b+(M2c)+M1c→C2a M2c→C2b					
<b>Contribution Claim 3 – Enhanced Processing of Roots and Tubers</b>					
M3b→C3a+C3b→O3	Insert main findings from the IFAD/FAO 2014 study on matching grant facilities in Ghana, and the RTIMP and REP M&E and supervision reports	Insert main findings from the livelihood analysis with intended beneficiaries (farmers and processors)	Insert main findings from the KIs with GPCs, BACs and PFI local branches		
M3b+M3c+C1a→C3c					

a. Score 0-6 (cf. rubrics). Justify the score and provide critical notes on remaining data gaps and weaknesses and potential biases.

b. Score 0-6 (cf. rubrics). Justify the score and provide critical notes on the relative influence of RTIMP (through the M-links).

Lessons learned from the Vietnam project helped to improve the sense-making model for Ghana. In Vietnam, discussions took place mostly in mixed-stakeholder groups, and in plenary sessions, which did not give the farmers enough of a chance to collect their thoughts and gain confidence. Learning from this, participants in Ghana first worked in peer groups organized around the part of the TOC that represented their “patch” in the supply chain system—for instance, farmers discussed the production part of the chain. In Vietnam, the reconstruction of the causal flow was done in plenary, which again did not offer sufficient opportunity for farmers to engage in the process. In Ghana, this was done in small mixed groups, organized around geographic areas, with the farmers and processors systematically given the floor first, before all others, to present their views. Plenary discussions took place only on the second day, in a fishbowl set-up, in which beneficiaries constituted the majority of the discussants. This was quite successful and provoked an animated discussion with bankers about the inaccessibility (thus failure) of the microenterprise credit mechanism put into place by the program (Van Hemelrijck 2013b, 2015).

## CONCLUSION

The action research around the two IFAD pilots have demonstrated that a participatory and systemic impact evaluation approach such as PIALA can produce rigorous, valid, and credible evidence that is useful for reporting and learning with partners and stakeholders, in contexts where traditional counterfactual analysis is not feasible. Moreover, the pilots have shown that using similar methods engaging beneficiaries in assessing and debating contributions to impact can contribute to enhancing the impact even ex post. Moreover, using similar methods and processes for collecting, cross-checking, and analyzing data in the two impact evaluations made it possible to compare and identify conclusions, and to formulate recommendations that have wider relevance for investments elsewhere, thus are beyond the individual programs in question.

Compared to the usual cost of theory-based, mixed-methods impact evaluations in countries like Vietnam and Ghana, these pilots were done with shoestring budgets. For example, the estimated budget for a one-year randomized controlled trial study in an IFAD-funded program in Ghana similar to the RTIMP was around \$200,000. But this study only covered one subcomponent of the program, and eight districts in the northern part of the country. The PIALA evaluation of RTIMP, by contrast, cost \$233,000 and covered the entire program, which consisted of three components, each of which had two or three subcomponents. Moreover, it covered 30 districts across the entire country (Van Hemelrijck 2016b).

In every evaluation that aims for greater value with limited resources, trade-offs occur. The PIALA pilots have demonstrated that these trade-offs can be turned into win-wins by carefully considering how rigor and inclusiveness can reinforce each other, and by critically reflecting on the potential loss in value-for-money if one were to be prioritized over the other (Van Hemelrijck and Guijt 2016). Limiting stakeholder engagement in the TOC process



to save time and resources, for instance, leads to a substantial loss of rigor in every phase of the evaluation. Conversely, reducing the scale and level of engagement in sense-making limits the cross-validation and thus confidence in the conclusions, while also reducing the inclusiveness and empowerment value of the evaluation (Van Hemelrijck 2016b). Reducing the sample size of the participatory inquiries, as to reduce the cost, not only limits the scale of participation and thus its impact on voice, empowerment, and ownership of the findings (Burns and Worsley 2015), but also thwarts rigorous causal inference. On the other hand, reducing the scope inhibits conclusion validity by confining the systemic analysis and thus the understanding of complex nonlinear impact trajectories (Woolcock 2009).

The most essential conclusion from the action research around the PIALA pilots is that inclusiveness and rigor can reinforce each other, and that this is even more likely to happen when participatory processes and methods are employed at a larger scale (Van Hemelrijck 2016b). People's participation in impact evaluation can contribute to their understanding of the system in which they are functioning (Burns and Worsley 2015), while also adding to the rigor and credibility of findings, if

- It is both inclusive and meaningful, enabling a robust cross-checking of many different *authentic* voices;
- It avoids the dominance of any single truth, power, or particular viewpoint, thus mitigating bias; and
- It creates space for solid debate and *equal* voice, including the voices of those who are the least powerful and least heard (Van Hemelrijck 2016b).

Scale achieved through rigorous sampling and representative inclusion of all stakeholder perspectives makes it possible to generate knowledge that supersedes isolated anecdotes. Moreover, it also makes it possible to build contrasting evidence from “natural counterfactuals” occurring in the sample, thus reducing doubt in causal inference. Rigor emanates from the thoughtful design and facilitation of the participatory processes at scale, in ways that forestall the dominance of a single truth or viewpoint and enable stakeholders to participate equally and meaningfully in the process. Other essential attributes of rigor are methodological complementarity and consistency, and extensive and robust triangulation and cross validation (Van Hemelrijck 2016b).

Critical to the quality of delivery at scale, clearly, is the capacity of the researchers. In the long run, investing in such capacity helps to reduce costs, while enhancing the value of impact evaluation. For the broader development sector, this implies building capacity through providing guidance and support to new experiments with approaches like PIALA (Van Hemelrijck 2016b). For IFAD and its partners, optimizing value-for-money would imply, for instance, using PIALA as a longitudinal approach integrated with program design. This would create more room for building local research partnerships and capacity for impact evaluation, while bringing quality, continuity, and consistency to IFAD's impact learning agenda at the national and global levels.

## REFERENCES

- Bamberger, Michael, Jim Rugh, and Linda Mabri. 2012. *RealWorld Evaluation*. Sage.
- Beach, Derek, and Rasmus Brun Pedersen. 2013. *Process-Tracing Methods: Foundations and Guidelines*. University of Michigan Press.
- Befani, Barbara, Chris Barnett, and Elliot Stern. 2014. "Introduction—Rethinking Impact Evaluation for Development." *IDS Bulletin* 45 (6): 1–5.
- Burns, Danny, and Stuart Worsley. 2015. *Navigating Complexity in International Development: Facilitating Sustainable Change at Scale*. Practical Action Publishing.
- Camfield, Laura, Duvendack, Maren, and Richard Palmer-Jones. 2014. "Things You Wanted to Know about Bias in Evaluations but Never Dared to Think." *IDS Bulletin* 45 (6): 49–64.
- Chambers, Robert. 2008. *Revolutions in Development Inquiry*. Earthscan.
- . 2015. "Inclusive Rigour for Complexity." *Journal of Development Effectiveness* 7 (3): 327–35.
- . 2017. *Can We Know Better? Reflections for Development*. Practical Action Publishing.
- Carugi, Carlo. 2016. "Experiences with Systematic Triangulation at the Global Environment Facility." *Evaluation and Program Planning* 55: 55–66.
- Copstake, James. 2014. "Credible Impact Evaluation in Complex Contexts: Confirmatory and Exploratory Approaches." *Evaluation* 20 (4).
- Funnell, Sue, and Patricia Rogers. 2011. *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. John Wiley & Sons.
- Gaventa, John. 2004. "Towards Participatory Governance: Assessing the Transformative Possibilities." In *Participation: From Tyranny to Transformation? Exploring New Approaches to Participation in Development*. Zed Books.
- Gaventa, John, and Andrea Cornwall. 2006. "Challenging the Boundaries of the Possible: Participation, Knowledge and Power." *IDS Bulletin* 37 (6).
- Guijt, Irene, Adinda Van Hemelrijck, Jeremy Holland, and Andre Proctor. 2013b. "PIALA Research Strategy: Improved Learning Initiative." Internal document. IFAD, Rome.
- Guijt, Irene, An Nguyen, Lien Bui Ngoc, Huong Dang, and Loi Vu Manh. 2014. "[Report on the Participatory Impact Evaluation of the 'Developing Business with the Rural Poor' Project in Ben Tre, Viet Nam \(2008–2013\): A Pilot Application of a Participatory Impact Assessment and Learning Approach.](#)" IFAD, Rome.
- Holland, Jeremy. 2013. *Who Counts? The Power of Participatory Statistics*. Practical Action.
- IFAD (International Fund for Agricultural Development). 2016. *IFAD Strategic Framework 2016–2025. Enabling Inclusive and Sustainable Rural Transformation*. IFAD, Rome.
- Kabeer, Naila. 1999. "Resources, Agency, Achievements: Reflections on the Measurement of Women's Empowerment." *Development and Change* 30 (3).
- Lain, Jonathan. 2016. "[Rising to the Challenge—Measuring an Expanding Concept of Resilience in Oxfam's Impact Evaluations.](#)" *Policy & Practice Blog*.
- Levy, Sarah, and Carlos Barahona. 2002. "How to Generate Statistics and Influence Policy Using Participatory Methods in Research." Working Paper. Statistical Services Centre, University of Reading.
- Merrifield, Juliet. 2002. "Learning Citizenship." IDS Working Paper 158. Institute of Development Studies, Brighton.
- Mertens, Donna. 2009. *Transformative Research and Evaluation*. Guilford Press.
- . 2010. "Philosophy in Mixed Methods Teaching: The Transformative Paradigm as illustration." *International Journal of Multiple Research Approaches* 4 (1).
- Mohan, Giles, and Samuel Hickey. 2004. "Relocating Participation within a Radical Politics of Development: Critical Modernism and Citizenship." In *Participation: From*

- Tyranny to Transformation: Exploring New Approaches to Participation in Development*, 59–74. Zed Books.
- Mosse, David. 2001. "People's Knowledge, Participation and Patronage: Operations and Representations in Rural Development." In *Participation: the New Tyranny?* Zed Books.
- Newman, Dan. 2015. *From the Front of the Room*. Matter Group.
- OECD DAC (Organisation for Economic Co-operation and Development Development Assistance Committee). 2010. "[Glossary of Key Terms in Evaluation and Results Based Management](#)." OECD, Paris.
- Patton, Michael Quinn. 2011. *Essentials of Utilization-Focused Evaluation*. Sage.
- Pawson, Ray. 2013. *The Science of Evaluation: A Realist Manifesto*. Sage.
- Picciotto, Robert. 2014. "Have Development Evaluators Been Fighting the Last War... and If So, What Is to Be Done?" *IDS Bulletin* 45 (6): 6–16.
- Ravallion, Martin. 2012. "Fighting Poverty One Experiment at a Time: *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. Review Essay." *Journal of Economic Literature* 50 (1): 103–14.
- Stern, Eliot. 2015. *Impact Evaluation: A Guide for Commissioners and Managers*. Bond.
- van Es, Marjan, Irene Guijt, and Isabel Vogel. 2015. [Theory of Change Thinking in Practice. Hivos TOC Guidelines](#). The Hague: Hivos.
- Van Hemelrijck, Adinda. 2013a. "Powerful Beyond Measure? Measuring Complex Systemic Change in Collaborative Settings." In *Sustainable Participation and Culture in Communication: Theory and Praxis*. Intellect Ltd.
- . 2013b. "Improved Learning Initiative for the Design of a Participatory Impact Assessment and Learning Approach (PIALA): Insights and Lessons Learned from the Reflections on the PIALA Piloting in Vietnam." IFAD, Rome.
- . 2015. "Improved Learning Initiative for the Design of a Participatory Impact Assessment and Learning Approach (PIALA): Methodological Reflections Following the Second PIALA Pilot in Ghana." International Fund for Agricultural Development and Bill and Melinda Gates Foundation, Rome.
- . 2016a. "[Stretching Boundaries of Evaluation Practice with PIALA in Ghana](#)." *Evaluation for Africa* blog.
- . 2016b. "Understanding Rigour in Participatory Impact Evaluation for Transformational Development: Insights from Piloting a Participatory Impact Assessment and Learning Approach (PIALA)." PhD work in progress paper. Institute of Development Studies, Brighton.
- . 2017a. "Governance in Myanmar. Evaluation of the 'Building Equitable and Resilient Livelihoods in the Dry Zone' Project." Effectiveness Review Series 2015/2016. Oxfam GB, Oxford, UK.
- . 2017b. "[Walking the Talk with Participatory Impact Assessment Learning Approach \(PIALA\) in Myanmar](#)." *Policy & Practice Blog*.
- Van Hemelrijck, Adinda, and Irene Guijt. 2016. "[Balancing Inclusiveness, Rigour and Feasibility: Insights from Participatory Impact Evaluations in Ghana and Vietnam](#)." CDI Practice Paper 14. Centre for Development Impact, Institute of Development Studies, Brighton.
- Van Hemelrijck, Adinda, and Glown Kyei-Mensah. 2015. "[Final Report on the Participatory Impact Evaluation of the Root & Tuber Improvement & Marketing Program \(RTIMP\): Pilot Application of a Participatory Impact Assessment and Learning Approach \(PIALA\)](#)." IFAD, Rome.
- White, Howard. 2014. "Current Challenges in Impact Evaluation." *European Journal of Development Research* 26.

- White, Howard, and Daniel Phillips. 2012. "[Addressing Attribution of Cause and Effect in Small n Impact Evaluations: Towards an Integrated Framework](#)." 3ie Working Paper 15.
- White, Sarah. 2015. "Qualitative Perspectives on the Impact Evaluation of Girls' Empowerment in Bangladesh." *Journal of Development Effectiveness* 7 (2): 127–45.
- Woolcock, Michael. 2009. "Toward a Plurality of Methods in Project Evaluation: A Contextualised Approach to Understanding Impact Trajectories and Efficacy." *Journal of Development Effectiveness* 1 (1).
- . 2013. "Using Case Studies to Explore the External Validity of Complex Development Interventions." *Evaluation* 19 (3): 229–48.

---

## Chapter 20

# Evaluating Mitigation Projects through a Theory of No Change

Takaaki Miyaguchi

**Abstract.** *Some have argued that, compared to climate change adaptation interventions, evaluating climate change mitigation (CCM) projects is relatively straightforward, due to the fact that there can be a clear, quantifiable goal regarding a reduction of greenhouse gas emissions. However, many donor-funded CCM projects do not seem to focus on output-based contributions, but rather on removing certain preconditions toward such market transformation. A program theory concept known as the Theory of No Change (TONC), put forth by Christine Wörten, provides an evaluation framework that is especially applicable to such CCM project interventions, and can serve as a useful tool in assessing how likely (or not) it is that interventions will achieve a market transformation. With close reference to the TONC evaluation framework, this chapter identifies and analyzes eight different CCM projects, from five different Association of Southeast Asian Nations (ASEAN) countries: Indonesia, Malaysia, the Philippines, Thailand, and Vietnam. Important findings from this analysis include the following: almost all the projects studied addressed barriers of ignorance and lack of expertise for all agent groups (consumers, supply chain, policy makers, and financiers); none of the projects has specifically addressed the barrier of cost effectiveness; and only a few projects specifically focus on harnessing the interest and/or motivation of relevant agent groups.*

The adoption of the Paris Agreement at the 21st Conference of the Parties (COP21), held in Paris in 2015, has marked an important milestone for the international community to unanimously tackle climate change. Although the emergence of the inward-looking Trump administration in the United States has aroused concerns worldwide, the international community needs to remain vigilant and not lose the focus on its fight against climate change. The Earth's climate is indeed changing, and addressing the cause of the problem is of prime importance and significance for all human beings. Climate change mitigation (CCM) is defined as "human intervention to reduce the sources or enhance the sinks of greenhouse gases."<sup>1</sup> In other words, mitigation mainly concerns measures and actions that reduce greenhouse gases (GHGs) by realizing and applying more energy-efficient, or renewable, energy technologies and practices. Although adaptation and loss and damage are both important and interlinked climate change concepts, CCM interventions are the most vital element in solving the root causes of climate change, as well as in transforming markets toward more carbon-neutral products and services.

This chapter discusses, among other issues, the CCM strategy of the United Nations Development Programme (UNDP) for its CCM programs that are funded by the Global Environment Facility (GEF) and implemented in the Association of Southeast Asian Nations (ASEAN) countries, the types of barriers against realizing CCM contributions and market transformation in these countries, and how these barriers correspond to the types of barriers that are introduced in the Theory of No Change (TONC) framework.

The rationale for selecting the ASEAN countries as a test case to apply a TONC program theory framework is that in these countries we have seen robust and steady economic growth over the past decades, with, on average, gross domestic product (GDP) growth of 7.0 percent between 1970 and 1995 (ASEAN Secretariat 2014). The region's recent GDP of 2012 was \$2.3 trillion, which was equivalent to 28 percent of China's GDP, and to 4.3 percent of the world's total GDP, while their population has grown to 617 million in 2012 (ASEAN Secretariat 2016). The importance of realizing green growth, or low-carbon growth, in these vibrant economies should be emphasized.

The kinds of CCM projects analyzed are those utilizing funding through the GEF. The GEF started its first pilot operation in 1991, and, as of 2016, is the largest funder of projects for protecting the environment in the world (GEF 2016). The GEF is a partnership involving a number of so-called implementing agencies, and UNDP and the World Bank are among the largest implementers of environmental projects supported by the GEF. GEF Council approvals of UNDP proposals, for example, have amounted to up to roughly 40 percent of the available funds for commitments up to 2017 (GEF 2016). In this chapter, tendencies and barrier-removal strategies—specifically among CCM projects that are supported by the GEF with UNDP as implementing

---

<sup>1</sup> United Nations Framework Convention on Climate Change, "Glossary of climate change acronyms and terms," [http://unfccc.int/essential\\_background/glossary/items/3666.php](http://unfccc.int/essential_background/glossary/items/3666.php).

agency—are analyzed. The chapter attempts to identify potential gaps within the CCM strategies and the types of barriers being addressed in each of these projects so that we will be able to generate an analytical picture of more effective CCM projects, upon which to build better and more effective strategies in addressing CCM issues not only in the ASEAN region, but also elsewhere in the world.

## THE THEORY OF NO CHANGE

In several communities of practice on evaluating climate change and development, evaluating climate change interventions has been considered to be quite difficult compared to, for example, analyzing interventions in the public health sector, for a number of reasons. For example, the following challenges are posed when evaluating climate change projects: a long time frame, uncertainty about actual climate change patterns and their effects in a given locale, shifting baseline data and changing contexts, measuring nonevents, a lack of universal indicators, contribution versus attribution, and diversity of key definitions and terms (Bours, McGinn, and Pringle 2014). As part of the effort to develop a more useful and effective evaluation framework to be applied in climate change interventions, Uitto and other scholars have emphasized the importance of utilizing a *theory of change* (Uitto 2014; Vaessen and Todd, 2008). A theory of change aims to make sure that the underlying assumptions through which desired changes are triggered and realized are made explicit, by highlighting the contextual conditions that may influence the outcomes or results of the interventions (Funnell and Rogers 2011; Weiss 1998). The theory of change is an important component of program theory, which is “an explicit theory or model of how an intervention, such as a project, a program, a strategy, an initiative, or a policy, contributes to a chain of intermediate results and finally to the intended or observed outcomes” (Funnell and Rogers 2011). The theory of change is thus considered to be a useful approach in evaluating complex international development projects and interventions (Center for Global Development 2006).

However, when it comes to CCM interventions, an otherwise useful theory of change approach does not necessarily tend to help render the relationship between interventions and outcomes explicit. This may be due to the fact that most of the CCM interventions funded and implemented by donors thus far have tended to focus on improving the enabling environment or “preconditions”—that is, they were more input-based than output-based. This tendency is represented by the concept of barrier removal. In many of the donor-funded projects, the focus is on the removal of certain types of barriers, which are believed to be preventing a society from achieving a market transformation and becoming carbon-neutral. In reflecting on these issues, Wörlen has proposed a “Theory of No Change,” with which one is able to “assess whether or not an intervention has been contributing to a more favorable framework for market development for a sustainable energy technology” (Wörlen 2011). Instead of looking at a specific causal relationship between inputs and outputs or outcomes, which is done with the help of a theory of change, TONC looks at whether certain CCM interventions have

met necessary preconditions: that is, whether or not they have the right input framework to be able to remove barriers for a market transformation in a society. If some of the necessary preconditions are not being addressed, then the TONC hints at the possibility of “no change.” That is, it posits that no positive causal change is likely to occur through CCM interventions. In this chapter, the TONC concept is adopted as a test case, to apply to the evaluation of CCM projects of several ASEAN member countries.

## METHODOLOGY

### Identified Projects

To test the TONC concept, a total of eight CCM programs were selected. These projects were implemented by UNDP, were already completed, and were ASEAN member countries. The project titles were as follows; see table 20.1 for a summary of these projects:

- Industrial Energy Efficiency Improvement Project in Malaysia
- Biomass-based Power Generation and Co-generation in the Malaysian Palm Oil Industry in Malaysia
- Removal of Barriers to Biomass Power Generation and Co-generation in Thailand
- Integrated Microhydro Development and Application Program in Indonesia
- Palawan New and Renewable Energy and Livelihood Support Project in the Philippines
- Efficient Lighting Market Transformation Project in the Philippines
- Energy Efficiency Public Lighting Project in Vietnam
- Promoting Energy Conservation in Small and Medium Scale Enterprises in Vietnam

The primary document sources used for this analysis were project documents (ProDocs) and terminal evaluations of the identified projects. The ProDoc is the official, finalized document that lays out important project implementation information as background analysis of the target country, its development objectives, planned activities, schedules, budgeting information, and so on. By evaluating the ProDoc of each of the identified projects, the types of barriers intended to be removed or reduced against market transformation within the countries, as well as targeted sectors and stakeholders through which such transformation is believed to emerge, were analyzed. ProDocs thus provided important information about project activities and interventions that were being implemented through the respective projects. It is also envisaged that, depending on the status of project formulation and implementation, applying an analytical lens (in this case using the TONC concept) to ProDocs can be a good meta-evaluation exercise.

The second type of source for this analysis was terminal ex post evaluations, prepared and submitted by independent evaluators upon operational closure of the respective projects. Based on this source, how the implemented



TABLE 20.1 **Basic project information**

Project title	Country	Duration	GEF funding	Co-financing
			(million \$)	
Industrial Energy Efficiency Improvement Project	Malaysia	1999–2007	7.3	13.5
Biomass-based Power Generation and Co-generation in the Malaysian Palm Oil Industry	Malaysia	2003–07	4.0	10.8
Removal of Barriers to Biomass Power Generation and Co-generation	Thailand	2001–09	6.8	10.2
Integrated Microhydro Development and Application Program	Indonesia	2008–10	2.1	18.5
Palawan New and Renewable Energy and Livelihood Support Project	Philippines	2000–05	0.8	1.8
Efficient Lighting Market Transformation Project	Philippines	2005–10	3.1	12.0
Energy Efficiency Public Lighting Project	Vietnam	2005–11	3.3	12.4
Promoting Energy Conservation in Small and Medium Scale Enterprises	Vietnam	2006–11	5.8	23.4

SOURCE: Project documents (Government of Indonesia and UNDP, 2005; Government of Malaysia and UNDP, 2005; Government of Malaysia and UNDP, 1999; Government of Thailand and UNDP, 2000; Government of Vietnam and UNDP, 2004; Government of the Philippines and UNDP, 2003; Government of the Philippines and UNDP, 2000; Government of the Vietnam and UNDP, 2005).

interventions had addressed various barriers and different stakeholders was systematically analyzed.

## Analytical Framework

In most of the GEF CCM projects, the concept of barrier removal is prevalent. This concept is based on the assumptions that: (1) market transformation toward a more sustainable and energy-efficient society has been prevented because of various types of barriers that exist in many areas, for various agent groups; and (2) transformation of markets occurs when these barriers are reduced or removed by various interventions or activities proposed in the interventions.

The main framework used for analyzing the types of barriers mentioned and targeted by each CCM project is called a TONC, as described above. The argument based on this concept is that when the preset types of barriers are not being addressed by project interventions, it can cause the possibility of generating "no change" toward transformation of market and achievement of CCM objectives, in the case of reduction of GHG emissions.

The theory of change concept presents useful guiding paths toward the achievement of intended results. But the TONC presents the *lack* of such

paths: as such, it is able to provide useful insights into whether or not project interventions are contributing to building a specific enabling environment toward achieving the intended results. When certain interventions are found *not* to be addressing the types of barriers prescribed by TONC, they can be thought of as not being prone to realizing the desired market transformation.

The TONC concept presented by Wörten was only tested for CCM projects that deal with retail products and heating systems, and she has pointed out the need for further research in applying this concept to other fields and cases (Wörten 2011). This meta-evaluation is one such attempt to apply the TONC concept to CCM projects that go beyond these fields.

According to Wörten (2011), the following seven common types of barriers, against which the TONC concept is to be applied, are introduced:

- **Ignorance.** A number of the agent groups may not know the benefits, or even the existence, of specific CCM technologies or products.
- **Lack of expertise.** The different agent groups may lack the expertise to operate and maintain the technologies and the products.
- **Lack of access to technology.** The technologies and the products used may be too expensive, or not readily available in the domestic market, due to insufficient capacity along the supply chain, or a lack of financing.
- **Lack of motivation.** The status quo, traditions, or stereotypes may continue to prevail as a source of resistance toward new technologies and products.
- **Lack of cost effectiveness.** Running the technologies may become, on a total cost of ownership basis, more costly than other traditional energy and technology choices.
- **Lack of affordability.** Such technologies often require large initial investment or upfront costs; a lack of financial support from a host government or commercial banks can therefore represent a barrier.
- **Lack of demand/business model.** It is necessary to generate enough demand in realizing the economies of scale for the supply or the financier's side, as well as to develop innovative and financially robust business models by business owners, especially in small and medium enterprises.

All these types of barriers can be addressed against the four agent groups Wörten introduces: (1) consumers/users; (2) supply chains (such as shops and maintenance technicians); (3) policy makers; and (4) financiers. Table 20.2 shows the TONC barrier types and relevant agent group types.

## Conducted Analysis

This meta-evaluation consisted of two sets of analyses. One was to map the types of barriers and addressed agent groups for each project, by referring to the respective ProDocs. This analysis is designed to see which activities had been planned for implementation and, consequently, which activities did not take place because they had not been planned. The second analysis concerns

TABLE 20.2 **Sectors/stakeholders and TONC barriers**

Agent group	TONC barrier type
Consumer/user	Ignorance
	Lack of expertise
	Lack of access to technology
	Lack of affordability
	Lack of interest/motivation
	Lack of cost effectiveness
Supply chain	Ignorance
	Lack of expertise
	Lack of cost effectiveness
	Lack of business model/no demand
Policy maker	Lack of motivation
	Lack of expertise
	Lack of (fiscal) affordability
Financier	Ignorance
	Lack of expertise
	Lack of cost effectiveness
	Lack of business model

SOURCE: Wörlen 2011.

the respective terminal ex post evaluations in order to collect and analyze the evaluative evidence of the activities that were implemented (or were not).

First, in analyzing the ProDocs of the eight projects, the barriers addressed were categorized according to the seven barrier types described above.<sup>2</sup> Second, by analyzing the ex post evaluations, the author codified the reports by assigning and counting positive (+) or negative (-) evaluative remarks (or mentions). This was not simply based on the assigned ratings within the evaluations (i.e., highly satisfactory, satisfactory, marginally satisfactory, and unsatisfactory). For example, whenever there was a remark that suggested a corrective word, for example “better,” “necessary,” or “needed” action, it was counted as negative. Most of the recommendations offered within the evaluations were also treated as negative (using keywords such

<sup>2</sup>There are two modifications that have been made by the author to the original TONC barrier types, i.e., inclusion of consideration of renewable energy sources such as biomass sources to the barrier of lack of access to technology, and addition of innovative policy models to the barrier of lack of demand/business model. These barriers were further assigned to the equivalent agent groups—i.e., consumer/user, supply chain (e.g., shops, maintenance technician, and those stakeholders engaged in business

as “should” or “recommended”). This was based on the assumption that recommended actions referred to things that should have taken place during the project implementation phase but had not. And of course such directly associated negative terms as “difficult,” “risky,” “poor,” “not properly,” etc., were noted as negative remarks. Similarly, when assigning positive remarks, the author not only referred to the ratings within the evaluation sections, but also paid close attention to the remarks related to the assessed project contributions, for example with such terms as “valuable,” “important,” “vital,” “key,” “success,” and the like.

One must note, however, that since only evaluators’ remarks were codified, this analysis did not assign any value where there were no planned activities. If there was no remark by the evaluator, even though the project may have missed an important barrier entirely, that element was simply left without any attribution, either negative or positive.

The second type of analysis was done without attempting to assign different degrees of positive-ness or negative-ness: that is, the author categorized the remarks, regardless of the strength or weakness of the adjectives into simply either “positive” or “negative.” The author of this chapter is keenly aware of the fact that such way of “quantification” cannot be considered a robust quantitative analysis. What was intended was a sort of qualitative analysis to discover otherwise unheeded patterns by applying a specific set of analytical lens, that is, the TONC framework.

## Findings and Discussion

Table 20.3 summarizes the results of the two types of analyses conducted: (1) the types of barriers addressed by each CCM project, and the involved agent group type against such barriers, plotted as against those prescribed by TONC concept; and (2) the frequency of either positive or negative evaluative remarks identified in each ex post evaluation.

Through this analysis and comparison, several important findings were discovered.

**Ignorance and lack of expertise barriers.** With some exceptions, almost all of the projects studied addressed the barriers of ignorance and lack of expertise for all agent groups where their implemented interventions cover as much as 88 percent of all the cells assigned to this barrier type. This may be due to the fact that UNDP, the implementing agency of these projects, is a development agency that emphasizes and focuses on technical assistance, especially through capacity development, be it through government (policy makers) or industries (depending on the project, but mainly for supply chains). Looking at the high frequency of positive remarks for the implemented interventions addressed to these barriers, one can say that overall, the projects’ intention of increasing a level of awareness and expertise among industries

---

and hard infrastructure), policy makers (e.g., government agencies and line ministries), and financiers (e.g., commercial banks and investors).

TABLE 20.3 Addressed barriers and frequency of evaluative remarks among the identified CCM projects

	Ignorance			Lack of expertise			Lack of access to technology			Lack of affordability			Lack of interest/motivation			Lack of cost effectiveness			Lack of business model/no demand			
	Consumer	Supply chain	Policy maker	Consumer	Supply chain	Financier	Consumer	Supply chain	Financier	Consumer	Supply chain	Financier	Consumer	Supply chain	Financier	Consumer	Supply chain	Financier	Consumer	Supply chain	Financier	
I	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
M	●			●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
M	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
P	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
P	●			●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
T	●		●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
V	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
V	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
	<b>68% (22/32)</b>			<b>88% (28/32)</b>			<b>81% (13/16)</b>			<b>58% (14/24)</b>			<b>38% (12/32)</b>			<b>13% (3/24)</b>			<b>63% (20/32)</b>			

SOURCE: Project documents and terminal ex post evaluations.

NOTE: n.a. = not applicable/relevant. I = Integrated Microhydro Development and Application Program (Indonesia); M = Industrial Energy Efficiency Improvement Project (Malaysia); M = Biomass-based Power Generation and Co-generation in the Malaysian Palm Oil Industry (Malaysia); P = Palawan New and Renewable Energy and Livelihood Support Project (Philippines); P = Efficient Lighting Market Transformation Project (Philippines); T = Removal of Barriers to Biomass Power Generation and Co-generation (Thailand); V = Energy Efficiency Public Lighting Project (Vietnam); V = Promoting Energy Conservation in Small and Medium Scale Enterprises (Vietnam). Percentages represent implemented types of interventions divided by the total number of cells for each barrier type.

LEGEND: ● intervention implemented for agent group. Frequency of evaluative remarks: ■ 1 positive remark; ■ 2 positive remarks; ■ 3 negative remarks; ■ 1 negative remark; ■ 2 negative remarks; ■ 3 negative remarks.

has been well realized. However, an interesting gap identified here is that although the agent group of consumers is covered by all the projects for ignorance barrier (i.e., for awareness raising), only a few projects address the issue of raising expertise or capacity for consumers/users. While it seems relevant to focus on those agent groups who have direct contact or business relations, such as supply chains and financiers, providing capacity development to consumers/users can arguably complement other project activities in raising expertise.

**Cost-effectiveness barriers.** Throughout the eight projects, only Thailand's biomass project seems to have specifically addressed this barrier, which has become visually obvious thanks to the TONC framework, where only 13 percent of this barrier's activities (cells) aimed at different stakeholders was implemented. On top of raising awareness and developing expertise, the issue of a total running cost, or a total cost of ownership basis, is significant when one tries to contribute to long-term, sustainable market transformation within a country. Such issues may have been considered by UNDP as "external risks" to the implemented CCM project interventions, for example, influence from fossil fuel subsidies, or the high cost or price of the technologies themselves. Nonetheless, the cost-effectiveness barrier seems to be an area that should be addressed more consistently in the design of future CCM projects.

**Lack of interest/motivation.** Another interesting gap is seen in the lack of interest/motivation barrier. Although the barrier of ignorance is universally addressed by all of the projects, when it comes to harnessing interest or motivation, there are only a handful that specifically incorporated these elements in project design, and in related implemented activities. Although it is acknowledged that the awareness-raising component in each project does include certain activities for harnessing interest and motivation, there seems to be a lack of emphasis on going beyond the level of simple awareness, and toward motivation and induced actions therefrom. For example, letting commercial banks know about the technologies and their financial feasibility is one thing, but actually inducing their interest and increasing their motivation for creating a proactive behavior/culture can be quite another. Such a potential systemic lack of focus toward this barrier is reflected in the frequency of negative evaluation remarks as well. For example in the Efficient Lighting Market Transformation Project in the Philippines, even after trainings had been conducted to raise awareness and expertise within the private and government banks, at the end of the project, none of them had provided loans for energy-efficient lighting. This surely was reflected negatively in the evaluation. (This activity component was given a rating of unsatisfactory.)

**Policy maker agent group.** When one looks at the interventions implemented by the agent groups, it becomes clear that the agent group of policy makers (e.g., government agencies and line ministries) does not seem to be well covered under such barriers as lack of (fiscal) affordability, interest/motivation, and business/policy model. The issue of fiscal or jurisdictional

affordability of certain policies and regulations, as with the case of the fossil fuel subsidies for the cost-effectiveness barrier, may be considered to be “external” to the range of project activities. For example, a country’s ministry of environment may not have jurisdiction or authority over non-environmental matters, such as financial ones, or those related to taxation/subsidies; however, creating cross-sectoral policies and active interministerial collaboration for establishing a development goal within a country still seems to be an important area for project intervention. Also related, policy makers who are involved with the CCM projects may not be interested or well motivated to devise innovative policy packages, or to collaborate with other ministries. Moreover, there may be an issue of power balance between an implementing agency (such as UNDP) and an executing partner (such as a government and its line ministries). Since government counterparts are one of the most important “clients” of implementing agencies, trying to force them to go outside their comfort zone of siloed political jurisdictions can well be a systemic challenge for the implementing agency. All in all, the barriers pertaining to policy makers can be an important gap that needs to be filled by future CCM design and implementation.

**Lack of demand, business/policy model.** When analyses of project design documents (through ProDocs, as black dots plotted according to the TONC framework) and terminal evaluations are combined, a very strong tendency has been revealed in the barriers addressing the lack of demand, and the business/policy model. In short, for these barriers, much has been implemented but much has also failed. This barrier area has had the largest frequency of negative evaluative remarks, yet with relatively a high percentage (63 percent) of intervention coverage. Activities are indeed happening; however, one can estimate that, due to rigidity of the business and policy models prescribed by the projects, this barrier area overall has been mostly a failure.

## CONCLUSIONS

The theory of change, though itself a useful concept in evaluating programs, often is not an applicable concept in the case of donor-funded CCM project interventions. This is because often the proposed activities do not by themselves generate GHG emission reduction, but instead are aimed at removing certain existing barriers to realizing a market transformation within a country, such as awareness, market, technology, policy, and finance-related barriers. The TONC concept was proposed by Wörlen (2011) as a theoretical framework for identifying important barriers that were not being tackled in a project’s activities. Thus TONC can be a useful tool when evaluating project design (through ProDocs) and implementation (through terminal evaluations), as well as the development impact of such projects.

As a test case, a meta-evaluation was undertaken to apply the TONC concept to eight CCM projects that were implemented by UNDP in and with ASEAN member countries. The main objective was to apply the TONC concept to different CCM projects tested by Wörlen, and to find out what

kinds of implications can be drawn through analyzing the projects' barrier removal strategies.

Though obviously limited in its analytical depth, by focusing on the TONC-prescribed types of barriers and key agent groups, it was possible to systematically see the institutional weak spots, and the biases found in project design. The TONC analytical framework has thus proven to be a useful tool for enabling a zoomed-out analysis of CCM project design. It can also help identify and facilitate the necessary actors and agencies before project implementation. When combined with ex post evaluation analysis similar to the one conducted here, the TONC framework can also be a useful tool for summative evaluation about the addressed or unaddressed barriers and interventions of CCM projects. This type of TONC-applied analysis, when accumulated for projects elsewhere, can serve as important reference for future CCM project design and implementation.

## REFERENCES

- ASEAN Secretariat. 2014 (Association of Southeast Asian Nations Secretariat). *ASEAN Economic Community Chartbook 2013*. Jakarta: ASEAN Secretariat.
- . 2016. *ASEAN Statistical Yearbook 2015*. Jakarta: ASEAN Secretariat.
- Bours, Dennis, Colleen McGinn, and Patrick Pringle. 2014. *Monitoring and Evaluation for Climate Change Adaptation and Resilience: A Synthesis of Tools, Frameworks and Approaches, 2nd ed.* Oxford, UK: SEA Change CoP and UKCIP.
- Center for Global Development. 2006. *When Will We Ever Learn? Improving Lives through Impact Evaluation*. Report of the Evaluation Gap Working Group. Washington, DC: Center for Global Development.
- Funnell, Sue C., and Patricia J. Rogers. 2011. *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. Jossey-Bass.
- GEF (Global Environment Facility). 2016. "Global Environment Facility Trust Fund Financial Report Prepared by the Trustee. Summary of Financial Information." GEF, Washington, DC.
- Uitto, Juha I. 2014. "Evaluating Environment and Development: Lessons from International Cooperation." *Evaluation* 20 (1): 44–57.
- Vaessen, Jos, and David Todd. 2008. "Methodological Challenges of Evaluating the Impact of the Global Environment Facility's Biodiversity Program." *Evaluation and Program Planning* 31: 231–40.
- Weiss, Carol. 1998. *Evaluation, 2nd ed.* Avenel, NJ: Prentice-Hall, Inc.
- Wörten, Christine. 2011. *Meta-Evaluation of Climate Mitigation Evaluations*. Arepo Consult.



---

## Chapter 21

# The Wicked Cases of Education and Climate Change - The Promise and Challenge of Theory-Based Impact Evaluations

Emmanuel Jimenez and Jyotsna Puri

**Abstract.** *To make progress in achieving the 17 Sustainable Development Goals of the 2030 Agenda, policy makers need to know what works to move the needle of the 169 targets, and then to act on that knowledge. While more is known now than when the Millennium Development Goals were set, there are still important gaps regarding the measured attributable impact of interventions on outcomes. This paper focuses on two key goals—education and climate change/environment—to illustrate the gaps in what has been learned and what still needs to be learned. It assesses why these gaps persist, and how future evaluations might address them.*

---

Emmanuel Jimenez, International Initiative for Impact Evaluation, [3ie, \[ejimenez@3ieimpact.org\]\(mailto:ejimenez@3ieimpact.org\)](mailto:ejimenez@3ieimpact.org); Jyotsna Puri, Green Climate Fund, [jpuri@gcfund.org](mailto:jpuri@gcfund.org). The views expressed in this chapter are the authors' own and are not the official views of the organizations the authors otherwise represent.

The role of evaluation in helping the world make progress on the 2030 Agenda for Sustainable Development is well argued by other chapters in this volume, as well as by many other thinkers. But it is not enough to monitor indicators: countries also need to know which policies, programs, and other interventions will be effective in moving the 169 indicators of the 17 Sustainable Development Goals (SDGs) to which they have committed.<sup>1</sup> Indeed, in asking for accountable institutions in Goal 16, the SDGs themselves underscore the importance of evaluation.

One key tool in this arsenal is evaluations that address the issue of attribution convincingly. Theory-based impact evaluations do so through methodologies that measure the effectiveness of interventions by posing a counterfactual question: that is, what would have happened without the intervention? In answering this question, they also answer other questions that are important for both donors and implementers to consider: Did the program or policy make a change (and how do we know if it did)? How much was that change? Would the change have occurred anyway, in the absence of the policy? Could it have been done better? Why did the change occur?

Theory-based impact evaluations measure causal change that can be attributed to an intervention, and use a prespecified theory of change to guide their hypotheses and to explain change. Good theory-based impact evaluations usually have the following components: a theory of change; pre-analysis plans; variables that are measured as objectively as possible, using survey data both at the baseline and end line; good pilots and formative work; a good understanding of outcome(s); SMART<sup>2</sup> indicators; good monitoring data and information on implementation fidelity; a good identification strategy; sufficient data size for statistical confidence; and high-quality analyses that mitigate a multitude of possible biases that may creep in over and above the bias of program placement and selection.

In this chapter, we investigate the present state of evidence and argue that theory-based impact evaluations provide a potential partial solution to answering the critical questions that the SDGs are asking. We show that every year many more of these evaluations are being published than ever before. We also discuss the limitations of current methods employed in theory-based impact evaluations, and argue that there are important gaps in the knowledge base in terms of topics and methods that need to be filled if we are to accomplish the goals of the 2030 Agenda in an evidence-based way.

We take a deep dive into two sectors to assess both the opportunities and the limitations for theory-based impact evaluations. Arguably education and environment are sectors that have posed what are termed “wicked” problems for evaluators (see, e.g., Levin et al. 2012). Interventions in these sectors

---

<sup>1</sup> See the Organisation for Economic Co-operation and Development website (<http://www.oecd.org/dac/evaluation/sustainabledevelopmentgoalsandevaluation.htm>) for a description of the 19th Development Assistance Committee meeting, which focused on the implications of the SDGs on evaluation.

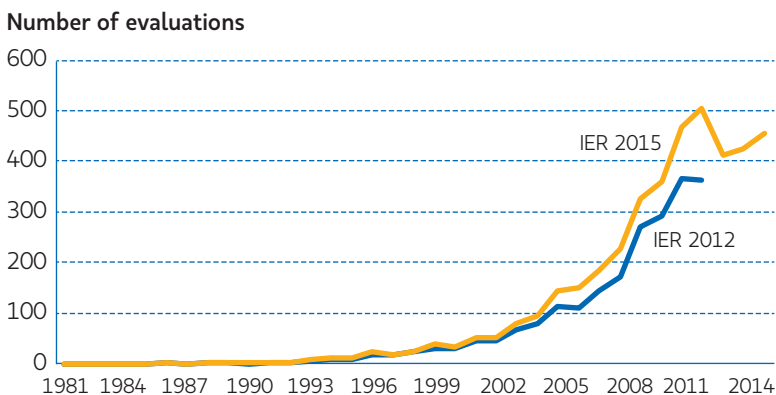
<sup>2</sup> Usually a mnemonic for smart, measurable, achievable, realistic, and time-bound.

resist a single solution because they are applied differently in different contexts. Moreover, the solutions are temporary while they address complex problems that require the use of multiple interventions simultaneously (Schwandt et al. 2016). We discuss how theory-based impact evaluations have tackled these issues and where gaps remain.

## IMPACT EVALUATIONS: TRENDS AND EVIDENCE GAPS

The number of theory-based impact evaluations has risen dramatically in the past 20 years. Figure 21.1 shows just one indicator—the number of theory-based impact evaluations of development interventions that are published per year and that take the counterfactual adequately into account. Figures are derived from the 3ie repository,<sup>3</sup> which was initially analyzed by Cameron, Mishra, and Brown (2016) and are currently being updated. In 1995, there

FIGURE 21.1 **Number of development impact evaluations published each year, 1980–2015**



SOURCE: Miranda, Sabet, and Brown 2016.

NOTE: IER = 3ie Impact Evaluation Repository. Data for 2015 are only for the first three quarters.

were fewer than 50 studies being published per year; by 2015, there were almost 500 and the repository contained more than 4,500 publications. While these figures need to be considered in light of publication lags, they include working papers in the gray literature that have shorter time frames between when the data are collected and when the results become available.

These numbers alone do not tell us anything about the need or demand for evidence. After all, there are untold hundreds of thousands of public and

<sup>3</sup> The 3ie Impact Evaluation Repository is an index of all published impact evaluations of development interventions.

private programs in more than 150 lower- and middle-income countries in the world. But they can be used as possible indicators of where glaring gaps may exist.

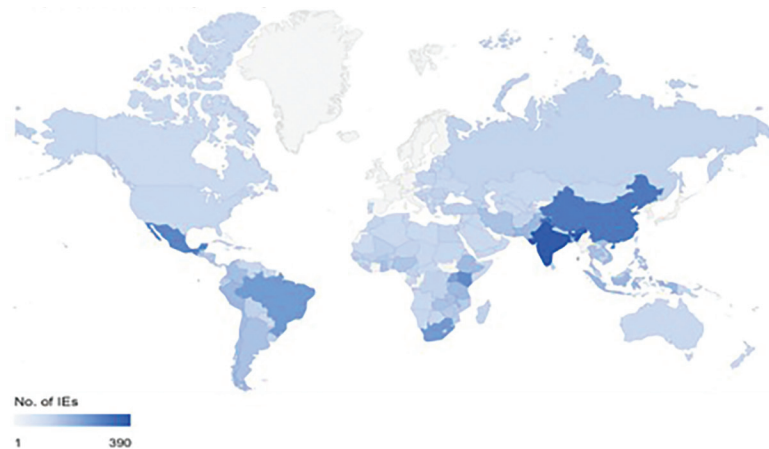
### Geographic Gaps

Even with these promising global trends, the density of evidence from rigorous impact evaluations varies widely across countries. Indeed, this is borne out in figure 21.2, which maps countries where studies included in the repository were conducted.

Figure 21.2 shows that countries in Asia (especially the largest countries, China and India), and parts of Latin America (Brazil, Mexico) and East Africa (particularly Kenya and South Africa) have more theory-based impact evaluations than others. This does not mean that these countries need fewer theory-based impact evaluations in the future. In fact, in terms of size of the economy and population, these countries may continue to need many more evaluations (e.g., the number of evaluations per 10 million people in China is about the same as that in Russia; it is more in India but half that of Brazil). But it does show that there are some regions that lag. There are extremely few (none in many countries) evaluations in West Africa, Middle East and North Africa, and Central Asia, the Pacific countries as well as the poorer countries in Latin America and the Caribbean.

The relatively uncovered regions are sites of fragile and conflict states (FCS), where populations are the most vulnerable. Only about 8 percent of published evaluations were done in FCS countries, and almost half of those were in just two countries—Pakistan and Zimbabwe.

FIGURE 21.2 **Impact evaluations by country**



SOURCE: Miranda, Sabet, and Brown 2016.

These geographic gaps pose a significant challenge for the 2030 Agenda, which is a global action plan: all member countries of the United Nations have committed to it. Yet for many there is a remarkably small evidence base that can attribute improved outcomes to interventions.

## Thematic Gaps

The majority of all published impact evaluations are in four sectors: health, education, social protection, and agriculture. Again, this is not to say that these sectors are saturated and have no further need for more evidence. But it does point to the key sectors that are consuming vast amounts of public and private expenditure, but are not being evaluated. For example, according to Miranda, Sabet, and Brown (2016), there are very few published theory-based impact evaluations in environment and separately in the energy sectors. To put this in perspective, India's public sector budget allocates a significant portion its budget to energy, and the World Bank has devoted 16 percent of its loans to it as well (World Bank 2016).

Arguably one reason for this lack of impact evaluation-related evidence may be the lack of demand in the sector. Indeed in several sectors, the questions examined by impact evaluations have traditionally not been considered important.<sup>4</sup> How much impact does a road make? Do protected areas reduce deforestation? Do climate change programs work to reduce greenhouse gases (GHGs)? Do children learn once they enroll and attend school? These are all examples of questions that have, until recently, not been considered in time-consuming and resource-intensive evaluations.

Another possible reason for this disparity is that it is believed that it is much more difficult to apply popular techniques of theory-based impact evaluations in some sectors, such as national infrastructure investments, or public finance policy, or practices of good governance, than in other sectors, where the interventions are smaller, easier to isolate, and have identifiable possible counterfactual (or comparison) populations. If so, the question is whether rigorous techniques can be developed to address key issues that obviously have huge implications for human welfare. Such efforts would have to take into account several other reasons why such knowledge gaps persist across sectors. For example, there may be disincentives for political economy reasons, for evaluating already scaled-up investments in sectors like transport, where large amounts of capital, both political and monetary, may have already been sunk (Ravallion 2016).

Aside from the density of evidence across broad sectors, there are gaps in thematic areas that are also of programmatic interest. For example, Puri et al. (2014) found that there were fewer than 50 studies of humanitarian assistance, into which the world has pumped over a trillion dollars. Another report found a single impact evaluation study in the governance

---

<sup>4</sup>We use the phrase "impact evaluations" and "theory-based impact evaluations" interchangeably. Indeed, we do not believe good impact evaluations can be undertaken without good theories of change.

and transparency of natural resource management in low- and middle-income countries (Puri 2017). Another concern is that in many impact evaluations, the costs of interventions are not analyzed. These trends present huge challenges for informing a comprehensive, global SDG agenda that encompasses almost all sectors in promoting people, planet, and prosperity.

### **Distributional Gaps**

Increasingly, more questions are being asked not only about the overall effects of interventions, but also of their effects on specific groups such as women and girls, vulnerable ethnic groups, the very poor, and so on. But the number of studies that have done such deep distributional analysis is also relatively low.

This gap holds even for sectors where there are more impact evaluations. For example, one of the findings of a recently completed systematic review by *3ie* of what works in education is that while studies reported on the average effects on all children, "...[few] studies included in the review provided any analysis of sub-populations, including factors such as sex or socio-economic status" (Snilstveit et al. 2016a, 14). A large part of this is driven by the fact that ensuring that impacts are measured with high statistical confidence for underrepresented groups means that the statistical samples need to be much larger. In another study we estimate that in one case, in order to ensure that results were representative for men *and* women, the sample sizes needed to quadruple (because women are traditionally underrepresented in some economic sectors), which meant a concomitant increase in costs (Puri, Rathinam, and Sarkar 2017).

## **IMPACT EVALUATIONS: CHALLENGES OF RELEVANCE AND METHODS**

As the number of impact evaluations have risen, researchers have learned more about their limitations and how to address them.

### **The Challenge of Responding to Questions Important for Policy Making**

Arguably, theory-based impact evaluations answer several questions that are important for policy making: Does the intervention work? How much? For whom? But sometimes they are just not the right instrument to answer the question. Nowhere is the latter point more salient than when researchers try to fit the question to the method. Simply identifying the underlying theories of change is a complicated enough undertaking, and adding the overall requirement of having the measure attributable to change becomes a daunting task (see, e.g., box 21.1).

The other question is to what degree should research be responding to policy, and whether research is important for its own sake. Theory-based impact evaluations tend to lie at the intersection of research and applied work (see, e.g., Puri, Rathinam, and Sarkar 2017). We argue that impact evaluations

**BOX 21.1 The use of impact evaluations in the evaluation of large, complex climate change programs: How can theories of change help?**

Aided by the Food and Agriculture Organization of the United Nations (FAO), the government of Paraguay is implementing a program to alleviate poverty and help reforest a large part of eastern Paraguay and increase the resilience of approximately 62,000 people. The proposal is to implement cross-cutting programming that meets both mitigation objectives (725,000 tons of carbon dioxide mitigated annually) and adaptation objectives (an expected direct increase in resilience and a reduction in poverty for 62,000 people). This three-phase program is spread out over 10 years and supports components such as environmental conditional cash transfers, cook stoves, and agroforestry programs for households. It aims to simultaneously improve the legislative and institutional frameworks mainly of forestry, environmental, and energy regulating entities. The overall objective of the program is to improve the resilience of poor and extremely poor households vulnerable to the impacts of climate change in environmentally sensitive areas of eastern Paraguay.

The implicit theory of change of the program is that (1) once authorities have the requisite funds and approvals, they will be able to set up environmental conditional cash transfer (E-CCT) payment systems that piggyback on existing cash transfer systems that already target poor and vulnerable communities through automated banking systems; (2) households will be targeted successfully; (3) as a consequence of the incentive of E-CCTs, households will start to build and invest in agroforestry systems (for which they will be paid for inputs and provided with technical assistance) they would not otherwise have; (4) households' agroforestry systems will be measured and detectable, which will then trigger payments to them; and (5) forest cover and degradation in eastern Paraguay will be reduced as a consequence and climate change mitigation will occur. There are several assumptions here, including assuming that households will be able to take the surplus produce from the agroforest systems to local and regional biomarkets, and that they will be able to earn incomes from these which will also reduce their income poverty and therefore increase their resilience.

Clearly, all of these statements require either strong previously produced evidence or smaller evaluative tests to understand whether the linkages are working, and whether the overall effects of the program will be achieved.

bridge a very important gap, in that they apply science and rigor to questions that have previously been hand-waved.

## The Challenge of Complexity

Complexity poses a substantial challenge to impact evaluations. Many programs involve a multitude of sectors: for example livelihood programs include interventions in water provision, sanitation, income-generation activities, and health. This usually means that causal pathways are not direct, are cross-linked, and are nonlinear. Separately, it also means that there are a multitude of sectors that every program is aiming to target. Arguably this reduces the incentive for any one sector team to invest in impact evaluations. Moreover, because there are intersectoral links and feedback loops, it becomes harder for impact evaluations to answer the “why” questions once the “how much” questions have been answered. Woolcock (2013) frames this challenge by citing three specific challenges that randomized control trials (RCTs) are unable to deal with. He cites the challenges of “causal density,” “implementation capability,” and “reasoned expectations” as being key features of complex systems that also make it difficult for RCTs to be used for understanding the overall impact of development interventions.

Another aspect of complexity is the measurement of relevant outcomes. For example, test scores may be an important indicator of performance in an education project, and RCTs may indeed be able to measure these well. It may also be possible to create good indicators and to include these in pre-analysis plans. But student stress may be an unintended consequence of these score-enhancing programs. There are two difficulties here: the inability to prespecify all possible consequences in a protocol, and the difficulty in measuring student stress caused by these programs.

Another example of the measurement challenge is climate projects that are aimed at increasing adaptation. A recent survey by the Overseas Development Institute found that there are at least 43 different frameworks for defining and understanding climate adaptation (ODI 2016). Again, because climate change programs also typically incorporate poverty alleviation and equity as a primary objective, these causal chains become very difficult to identify. As Levin et al. (2012) point out, it also becomes more likely that a specific solution to one development challenge creates a new problem for another one. Road building is touted as one such development solution that has clear (negative) implications for forest cover and biodiversity.<sup>5</sup> Additionally, confounding features of programs make it difficult to identify and measure the key change the program is seeking to bring about. Given these challenges, designing impact evaluations becomes even more difficult.

---

<sup>5</sup> Cropper, Puri, and Griffiths (2001) and Puri (2016) discuss the exceptions to this rule.



## The Challenge of External Validity

Limited external validity is another limitation of theory-based impact evaluations. Other authors have raised this concern as an important detractor from impact evaluations (Basu 2013; Pritchett and Sandefur 2014; Woolcock 2009). These are important concerns, and impact evaluations will need to respond to them by using new tools. It is of course true that non-impact evaluations are typically not externally valid either. We argue that in this case theory-based impact evaluations, because they are able to articulate the theories behind overall interventions and also provide statistical estimates with confidence intervals, become easier to aggregate through meta-analysis. Although limited, in these cases it is easier to say something about “net” or “aggregate” effect sizes (see Snilstveit et al. 2016b; Waddington and White 2014).

The knowledge gaps and methodological challenges discussed above pose challenges for evaluating the effects of interventions that will help countries address the 2030 Agenda. But they are not insurmountable. In this and the next section, we discuss these challenges and how evaluators have tried to address them in the “wicked” sectors of education and climate change.

### WICKED SECTOR: EDUCATION

According to UNESCO’s post-2015 Global Education Monitoring report (UNESCO 2015), in order to achieve the ambitious SDG targets for education by 2030, the spending per primary school student in low-income countries needs to be double the current level of spending. The International Education Commission calls for total spending in education to *triple* from its present \$1 trillion. But more funding is not sufficient for addressing the learning crisis: resources need to be directed to programs that work. There are a large number of reports about education, but there are relatively few that address attribution directly. *3ie* recently completed a comprehensive systematic review of the effectiveness of 21 different types of education programs on children’s school enrollment, attendance, drop-out rates, completion, and learning outcomes (Snilstveit et al. 2016a). It included evidence covering more than 16 million children across 52 countries, participating in 216 education programs in 52 low- and middle-income countries. The findings from this study can help inform decisions about effective strategies for achieving the education targets.

The review drew on evidence from 238 impact evaluations and 121 qualitative research studies and process evaluations. Interventions such as cash transfers, structured pedagogy, and computer-assisted learning programs were studied extensively. For other programs, such as school-based health, information to children, teacher interventions, remedial education, and school-day extension, the evidence is more limited. Significant investments are being made for funding interventions in understudied areas such as teacher-related programs. There is an urgent need for generating more evidence to help in informing funding decisions.

The education sector mirrors global variation in the availability of evidence. The greatest number of studies was identified in Latin America and

the Caribbean (87); Sub-Saharan Africa (59); and South Asia (51). Countries where several studies have been conducted include Brazil, Chile, China, India, Kenya, Mexico, South Africa, and Uganda. Evidence is limited or nonexistent for many countries in Sub-Saharan Africa, and for several countries with large populations, such as Bangladesh, Indonesia, and Nigeria.

Aside from inadequate thematic and geographic coverage of some important interventions, the usefulness of impact evaluations for the 2030 Agenda faces another challenge. Many education interventions are methodologically “wicked” to evaluate. Three aspects are particularly important to consider: the logical chain of intervention to results; context; and implementation.

### **The Logical Chain from Interventions to Results**

While some interventions have a relatively simple logical chain from intervention to results, such as the provision of textbooks on learning, or of scholarships on school participation, many others are characterized by causal density. This means that the interventions are “...highly transaction intensive, require considerable discretion by implementing agents, yield powerful pressures for those agents to do something other than implement a solution, and have no known (ex ante) solution” (Woolcock 2013).

It is thus not surprising that interventions that have a direct and simple link to the desired outcome—short results chains—are more effective. For example, cash transfer programs were the most effective intervention to boost school attendance. Where the outcomes of any one intervention are conditioned by the effectiveness of other interventions that may be beyond the scope of the program, the results tend to be more mixed. For example, in contrast to their effect on school participation, cash transfers have very little effect on learning outcomes as measured by mathematics or reading scores. This may not be surprising, given that most of the programs were conditioned on school participation and attendance, not on test performance. But the fact that learning outcomes were not significantly affected may also be a reflection of the low quality of schools that children were incentivized attend. In Colombia, for example, school vouchers (which are effectively conditional cash transfers) had no effect on learning outcomes if they were limited to government schools, but had a positive effect in those areas where the recipients were able to use them for entry into private schools, most of which were perceived to be of higher quality (Barrera-Osorio et al. 2011).

Another example is the need to address the incentives of the most important actor in affecting students in classrooms—the teacher—in almost any intervention. Some of the teachers who were delivering the Reading to Learn program in Kenya chose not to accept the class materials because they considered them difficult to master. This may have been one of the reasons that the program did not improve children’s performance in written and oral literacy exams. Similarly, the evidence on computer-assisted learning programs suggests that while the implementation of training for teachers is an issue, program designs need to also consider teacher workloads, as well as their attitudes and motivation for making radical

changes in the way they teach. School-based health programs also require teachers to participate in program delivery. Hence, programs need to consider whether this is increasing the workload for teachers and disrupting the regular class routine.

### **Baseline Conditions and Local Context**

Many successful interventions have tailored their design well to the existing human and social capital of specific contexts. This is particularly important for interventions aimed at children and households, and those aimed at improving governance.

School feeding programs, for example, have had the largest effect in areas characterized by high levels of food insecurity, malnutrition, and low school attendance. The effects have been much smaller in better-off areas where enrollment was already high, and malnutrition less of an issue. The school feeding program in Guyana, for instance, was implemented at a time when there was a documented increase in food insecurity for poor families. Not surprisingly, the program had large positive effects on school participation and learning. However, in Chile, the effects of the program on school participation were found to be small or nonexistent. In this case, the program was implemented at a time when extreme malnutrition had been eliminated, and enrollment rates were already high.

The baseline level of social capital has been found to be more important in interventions aimed at improving the system of governance of schools. School-based management and community-based monitoring had the best take-up in settings with high levels of social capital and a tradition of local participation. In the Philippines, where the effects of school-based management were consistently positive, qualitative evidence suggests that parents and communities were willing and able to make basic decisions about schooling when given the opportunity to do so. In contrast, results in most other contexts were disappointing. Evidence from Niger and Gambia pointed to low social and human capital as an important constraint for school-based management programs. Programs that rely on parental engagement for successful implementation may be better targeted in contexts where there is sufficient social and human capital to be able to hold other stakeholders accountable. For instance, where school committees are educated, or have experience in another community organization, parental monitoring of teacher attendance is likely to increase in response to the grant. Where these conditions are not met, programs may have a higher chance of success if there is a strong capacity-building component that is focused on facilitating community involvement. More generally, when parental engagement is a key part of the theory of change of a program it is important to assess the local capacity to engage in the way assumed by the theory of change. Programs could then be designed to account for any deficit in social and human capital.

It is therefore imperative that decision makers obtain accurate baseline information at the design stage of the program. This is required in order to tailor new programs to target the main constraints and achieve better outcomes.

## Capacity to Implement

The success, or more often the failure, of a program has often been attributed to the way the program has been implemented. Issues related to implementation have frequently been reported for a range of programs. The effect sizes of some programs were much smaller due to implementation problems. For example, in Kenya's and Uganda's Reading to Learn, as well as in Mali's Read, Learn, Lead programs, school materials and other tools were not delivered in a timely manner, which may partly explain why they had no effects, or very small ones, compared to the overall average for structured pedagogy on learning outcomes (Snilstveit et al. 2015). Similarly, the distribution of textbooks to students was found to be lower than intended in the case of a few programs that were providing school materials.

Several computer-assisted learning programs have faced issues such as insufficient, damaged, and dysfunctional equipment, lack of Internet access, and software not being compatible with hardware. Insufficient training of teachers is another issue that has been brought up as a challenge for several programs, including computer-assisted learning. Implementation issues, particularly with respect to the transfer of funds affected the success of several school-based management programs. Grants were not disbursed as intended, and significant delays were reported for several programs. Finally, unforeseen circumstances such as epidemics and conflicts have also delayed the implementation of education programs.

In most cases, these issues have cropped up due to the lack of capacity for implementation at various levels of the supply chain. In some cases, the inability to ensure a sustained and timely supply of resources has affected the effectiveness of programs. The difficulty in implementation is also often seen in programs that include a range of activities, and that have ambitious goals and long causal chains. This leaves a lot of room for implementation failure. In contexts where there is limited capacity to implement it may be necessary to give up on some of the objectives in the interest of making the program capable of implementation.

## Summary Implications

All of these complications point to the need for better-designed impact evaluations: those that study multiple options (or arms) to test different combinations of interventions would be greatly beneficial in addressing the causal complexity of some education interventions. But the examples above also point to the need for mixed methods in evaluation. Rigorous case studies, such as in Woolcock (2013), as well as incremental approaches to learning as in Andrews, Pritchett, and Woolcock (2012) would be one way to approach this. Finally, rigorous estimates of effects must be accompanied by equally rigorous studies of implementation.

## WICKED SECTOR: CLIMATE CHANGE

In this section we discuss the overall strengths and limitations of using theory-based impact evaluations in climate change programs and policies. It is clear that the international policy arena has parsed climate change into

several components, perhaps recognizing their overwhelmingly large reach and scope. So, for example, there are different conferences of parties (COPs) for climate and for forestry. Organizations and funding are also largely segregated into three different areas or sectors—mitigation, adaptation, and forestry. The United Nations Framework Convention on Climate Change recognizes these areas. Therefore, we define climate change activities and sectors as all those that help to reduce or stabilize GHG emissions, and that help to increase adaptation to climate change and its resulting uncertainties and weather extremes.

In the mitigation category, a host of types of policies and programs are included—these include policies and programs that increase access to and the use of low-emission energy and power generation; programs that increase access to and the use of low-emission transport; energy-efficient buildings, cities, and industries; and programs and policies that aim to increase sustainable land use and forest management, including reducing emissions from deforestation and forest degradation, or REDD+, programs. In the adaptation category, the range of policies and programs includes those that increase resilience and enhance livelihoods of vulnerable people, communities, and regions; that increase resilience of health and well-being; that increase food and water security; that increase the resilience of infrastructure and built environment to climate change threats; and that increase the resilience of ecosystems.

### Challenges of Evaluating Climate Change Action

Evaluations of programs and policies that deal with climate change encounter some of the challenges laid out in the section on education, and others as well. First there is the challenge of distal impacts. Climate change mitigation takes time (and scale): assessing overall contribution to climate change mitigation requires long time horizons. With theory-based impact evaluations, some of this is dealt with by underlying theories, mapping outcomes, and assessing efficacy and program success (see, e.g., box 21.1). The overall question related to understanding and measuring change, however, remains a challenge. Indeed, an evidence gap map examining the effects of land use policies on mitigation (Snilstveit et al. 2016b) found that, although there were 221 studies that rigorously looked at the impacts of land use policies and interventions on outcomes such as tree cover, livelihoods, and health, there were *no* evaluative studies that linked these, in an attributable way, in developing countries, to GHG emissions. This not only speaks to the difficulty of waiting for long periods of time for these impacts to show: it also underscores the difficulty of measuring GHG emissions. The other difficulty in these programs is that in order for there to be a *measurable* effect on even GHG emissions, programs have to account for “leakages,” that is, the likelihood that mitigation programs in one area may lead to the displacement or movement of emission activities to other areas. Impact evaluations therefore have to cover large areas, in order to ensure that there is a net effect on GHGs. This public good nature of climate change action imposes large transaction costs, but it also means that impact evaluations that aim to measure attributable change have to focus on large-scale action, and this may not always be possible (box 21.2).

**BOX 21.2 A large-scale mitigation program: An example of solar home systems**

Bangladesh's solar home systems (SHS) program—supported by the World Bank, GIZ, KfW, the European Union, the Inter-American Development Bank, and the multidonor Global Partnership for Output-based Aid trust fund—aims to provide energy for poor and vulnerable households. To gauge its uptake and effects, an evaluation was undertaken 10 years after its inception. The findings revealed a complex set of factors at play. To begin with, by 2013, only 10–12 percent of off-grid households had access to SHS off-grid devices, and diffusion rates were low: on average, a maximum of one-third of eligible households had adopted SHSs. The households that adopted the devices were, on average, much richer (80 percent higher incomes than non-adopting households) and better educated, with high percentages of non-agricultural income and a higher level of household assets. More than 78 percent of the adoption had only occurred during the last three years of the program. Despite their SHS adoption, most households continued to depend on traditional sources of energy. While there was some evidence of a substitution effect with SHS replacing kerosene, SHS households overall consumed more energy compared to non-SHS households—indicating that the income effect was stronger than the substitution effect. An important factor influencing adoption is the cost (including interest cost) and maintenance of SHS devices. But over a quarter of those taking out loans for the SHS devices—which are sold on credit, with loans provided for three years with a flat interest rate of 6 percent—were late in their repayments. Clearly for the program to conclude that it has been effective in achieving its long-term goal, given that the magnitude of change in overall emissions will be important, evaluation will need to measure the income effect and substitution effect over time.

SOURCE: Adapted from Asaduzzaman et al. 2013.

The second challenge for climate change evaluations is that most climate change projects have multiple objectives. This means that other than feedback loops and backward and forward links, most climate change projects are not just planned and implemented to maximize impact on climate change, but to simultaneously affect social, economic, health, and agricultural objectives. This means that the strength of links in a theory of change are frequently not the same, that they intersect and impacts mostly depend on the efficacy of several links being realized. This makes impact evaluations—which assume that a single intervention will lead to the overall impact, all other things being held constant—difficult to plan, implement, and realize in this space.

Third, climate and the environment are inherently public goods. This means that the overall impact of interventions is not individually determined by the successful implementation of one project over one discrete area.

Rather, in a twist on the problem of tragedy of commons, it is characterized by the problem of large numbers with small payoffs. This scale problem has two implications for impact evaluations. First, it means that impacts do not show unless there are a large number of agents. Second, they do not show unless a large number of agents are successfully undertaking these actions. Therefore, impact evaluations of climate change programs and policies in most cases have to concentrate on measuring attributable change at the outcome level. Furthermore, in most cases, although small programs may themselves be successful, we still may not see any changes in overall climate change-related impacts: this is true for both mitigation and adaptation programs. In some cases this means that small climate-related impacts have to accumulate until we are able to confidently detect and witness a change. As Bamberger, Rao, and Woolcock (2016) and Woolcock (2009) explain it, the impact function for climate change programs and policies may be nonlinear, or they may be horizontal straight lines before we see any impact. Examining and understanding the role that scale plays in identifying and measuring impact with statistical confidence while planning impact evaluations is therefore very critical.

Credible and high-quality impact evaluations are critically dependent on defining the appropriate questions and “system boundaries,” that is, defining the type and nature of the interventions that will be examined through impact evaluations. This in turn implies that impact evaluations are only able to examine a shortlist of interventions that have been defined using a variety of designs, such as factorial designs and pipeline designs. They are, by themselves, unable to compare interventions that have not been shortlisted. Arguably though, theory-based impact evaluations can deal with these through doing good initial formative work and specifying unintended consequences, and by undertaking rigorous qualitative work along with data collection that can help inform areas that previous theories may have been blind to (see, e.g., Rao and Woolcock 2002). We believe this is important to understand and concede, primarily because theory-based impact evaluations also have the advantage of lending themselves to systematic reviews with statistical meta-analyses that help us understand aggregate average effects, but also help us view the distribution of these effects; identify and analyze outliers; and examine other effects, such as “dose-response” pathways. In the next section, we will review some learning from impact evaluations in this area, and discuss new areas that theory-based impact evaluations should focus on, given the challenges and gaps.

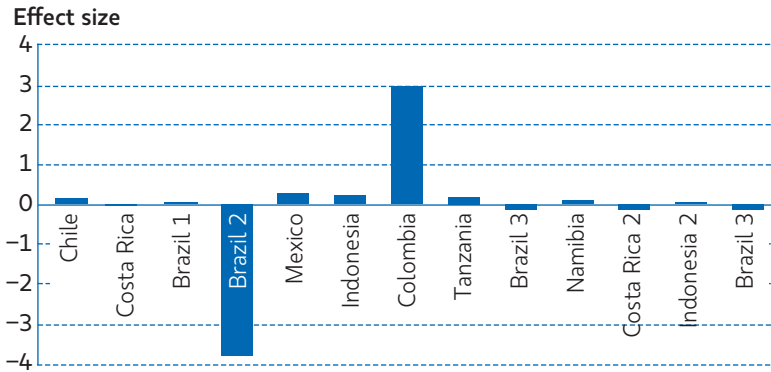
### **What Are We Learning?: Recent Evidence from Theory-Based Impact Evaluations and Systematic Reviews**

Theory-based impact evaluations have helped us understand the amount of change that environmental programs are bringing about. In figures 21.3 and 21.4, we show an illustrative summary of the magnitude of impact that several impact evaluations are able to measure in forestry programs.<sup>6</sup>

---

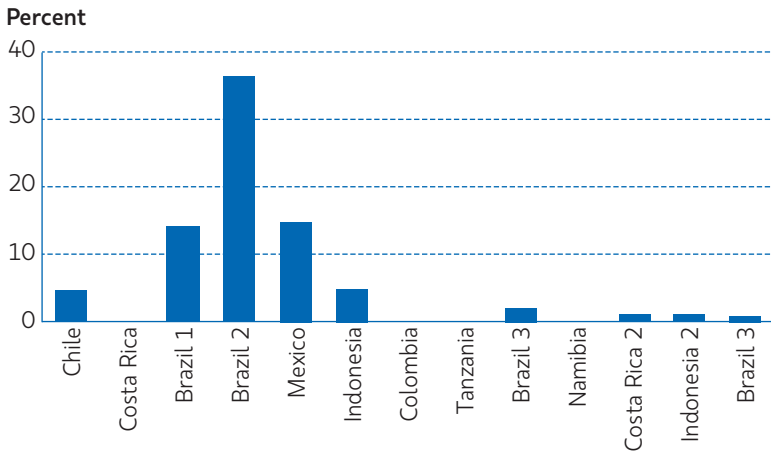
<sup>6</sup> See PLOS Collections (2016) for a summary of statistics from different forestry programs.

FIGURE 21.3 Standardized effect sizes from forestry programs



SOURCE: Adapted from Börner et al. 2016.

FIGURE 21.4 Estimate reduction in forest loss as a consequence of forestry programs



SOURCE: Adapted from Börner et al. 2016.

**Bias.** The important thing to note is that impact evaluations help deal with the problem of endogeneity and placement bias. Cropper, Puri, and Griffiths (2001) and Chomitz and Gray (1996) account for the attributes of plots where protected areas tend to be sited or located. Since these are areas that have low agricultural productivity and are likely to be remote, it is clear that any naïve estimate that does not consider this selection bias is likely to have extremely biased results.

**Targeting.** Impact evaluations can also help us understand the effectiveness of targeting: Are programs really reaching the populations that they



need to, and are these the populations that programs most need to target? Alix-Garcia, Shapiro, and Sims (2012) found that the countrywide Payment for Ecosystem Services program (PSA) in Mexico, with a budget of more than \$5 million, was quite successful in targeting households that were eligible for the program. In contrast, Azofeifa et al. (2007) found that in Costa Rica, the PSA program did not target those locations that were most likely to change land use. As a consequence there were very small changes in forest cover caused by the PSA program.

**Subgroup effects.** Impact evaluations can also help to address questions of equity and heterogeneous impacts. Somanathan, Prabhakar, and Mehta (2009) have shown that after accounting for potential selection and placement bias, community-managed forests performed better in raising crown cover by 12–16 percent when compared to unmanaged commons, but only for forests of broadleaf trees, not pine trees. Understanding the effects on subgroups, however, requires that sample sizes be selected in such a way that they are representative for the subgroups of interest.

**Comparing different kinds of programs.** Many studies have examined programs that engage communities and compare their effectiveness with the status quo, such as government managed systems, or unmanaged systems. For example, Tachibana and Adhikari (2009) showed that in Nepal, community co-managed forests recover much more quickly than forests where communities are solely managing their forests. And Cropper, Puri and Griffiths (2001) found that protected areas are less effective in protecting forests than wildlife sanctuaries by themselves, perhaps because the latter have more resources devoted to them.

**Are we doing the right things?** In our view, one of the key questions that impact evaluations should answer is, “Are the right things being done?” A study by Wynes and Nicholas (2017) found that of the top four mitigation actions that individuals can undertake to reduce GHGs, only two are discussed in high school textbooks. The other two actions are completely ignored. If we are to focus policy and action on the most effective actions, evaluations need to start asking the question of whether the right strategies are being pursued programmatically, rather than evaluating only the implementation of policies. While traditional evaluations have mostly been unsuccessful in this area, impact evaluations can help us respond to this overall question.

**Trade-offs.** A relatively but clearly an important question in climate change is examining any potential trade-offs between economic outcomes on one side and environmental outcomes on the other. This is particularly important in the case of programs that aim to reduce the consequences of development on the environment. A good example is provided by the Alix-Garcia et al. (2013) study of Oportunidades, a conditional cash transfer program, and its consequences for forests. The authors of this study found that forests were detrimentally affected as a consequence of a cash transfer program, and

that the theory-based impact evaluation helped to measure the magnitude of this effect. This is important, because program managers and policy makers can then measure the magnitude of this effect and make policy decisions accordingly.

## **CONCLUSIONS: ARE THERE COMMON OPPORTUNITIES TO ADDRESS THE GAPS?**

Theory-based impact evaluations have been used across the development and humanitarian sectors to inform the effectiveness of programs. This includes investigating the best ways to deliver humanitarian assistance (see, e.g., Doocy and Tappis 2016; Puri et al. 2017); examining the effectiveness of self-help groups in empowering women through microfinance (see, e.g., Brody et al. 2016), community-driven development (King, Samii, and Snilstveit 2010), sanitation programs (Buck et al. 2017), farmer field schools (Waddington and White 2014), agricultural insurance (Barooh, Kaushish, and Puri 2017), reducing poverty (Banerjee et al. 2015), and day care programs (Leroy, Gadsden, and Guijarro 2012). In so doing, they have helped to reduce ambiguity in our knowledge of the effectiveness of development programs.

However, many challenges remain. First, theory-based impact evaluations have not succeeded in meeting the methodological challenges discussed in this chapter. Additionally, it is clear that theory-based impact evaluations have not really leveraged the data revolution. The methods traditionally employed in theory-based impact evaluations have largely remained the same, predicated on the assumption that data are scarce and infrequent. Advances are being made with machine learning that use frequent, high density, spatially disaggregated data to understand especially the heterogeneity of impacts, but they are making their way only very slowly into theory-based impact evaluations.

Second, theory-based impact evaluations have largely shied away from meso-level or macro investigations. Causal identification through the use of controls or comparison groups remains a challenge here. Some studies are using innovative methods such as synthetic controls and machine learning (see, e.g., Acemoglu et al. 2013; Sills et al. 2015). Others are still venturing into meso-level investigations using regional controls (see, e.g., Bos et al. 2017). These applications, however, remain infrequent.

Third, theory-based impact evaluations have stayed mostly quiet on systems thinking and on understanding what changes institutions. Methodologies have been limited in this space. This is important because most agencies, especially environmental agencies such as the Global Environment Facility, the Climate Investment Fund, and the Green Climate Fund are aiming to achieve “transformational change.” An important characteristic of transformational change is being able to detect and measure systems change. This is an important policy imperative not just for environment-related organizations, but for the development sector as a whole, and it will be useful for the evaluation community to engage closely with applied academics to explore methodological options that help to identify and measure causal contribution in this area.

As recent papers have demonstrated (PLOS Collections 2016), the size of causal change differs dramatically depending on the spatial resolution of data: the less the resolution, the greater the imprecision, but the higher the resolution, the greater the heterogeneity in impacts across intervention sites. Arguably, therefore, it may be even more important to measure the cost-effectiveness of programs and projects (PLOS Collections 2016). Unfortunately this is not something that a lot of evaluations do. The absence of data on costs of implementation is usually cited as a reason for this absence of analysis. But we believe it as important to understand overall effect size as it is to measure cost effectiveness. Unfortunately, as we have shown in one of our papers, there are very few studies that examine cost effectiveness (Puri et al. 2016).

We also believe that evaluations that contribute to implementation science by examining *how* programs may be implemented are far more important than measuring the overall effects of programs. Although other sectors, such as nutrition and health, have long held this as an important area of exploration (see, e.g., Menon et al. 2014), impact evaluation techniques, especially those related to causal identification, have not found widespread use. Comparing different delivery mechanisms and how effective they are in realizing results is especially important. An example of this can be seen with Doocy and Tappis (2016), where the authors compared the effectiveness of cash transfers versus food transfers, versus in-kind transfers in humanitarian contexts. Within this class of research we also recommend using impact evaluations to examine “last mile” questions. Most programs assume that good implementation leads to good results. However, as has been most recently explored by the behavioral insights literature, good implementation is a necessary, but not a sufficient condition for success in development programs. These last mile problems have been examined in the context of the adoption of new technologies (e.g., Burwen and Levine 2012) or for new instruments (Barooah, Kaushish, and Puri 2017). Most programs fail because they presume that good monetary incentives are in themselves sufficient to ensure results. However, the literature on behavioral insights has now shown us that these assumptions are unrealistic.

Despite all of these challenges, we remain sanguine. Theory-based impact evaluations have been able to answer many difficult questions. They have helped policy makers and evaluators understand and measure overall results, and deal with a variety of biases while understanding the impact of development assistance. They have arguably helped to turn the tide in international assistance by providing comparisons of the effectiveness of different programs, that for long periods of time had been accepted as being successful and useful. Theory-based impact evaluations have provided us with a method for comparing strategies as well understanding their relative impact, while developing a systematic way to aggregate effects and understand average impact. Clearly the field is still in its infancy, though, and new, customized methodological advances will be required if we are to answer the questions that are relevant to the policy community.

## ACKNOWLEDGMENTS

The authors would like to acknowledge, without implicating them, the comments received from Rob D. van den Berg and the research assistance provided by Swarnakshi Luhach.

## REFERENCES

- Acemoglu, D., S. Johnson, A. Kermani, J. Kwak, and T. Mitton. 2013. "The Value of Connections in Turbulent Times: Evidence from the United States." *Journal of Financial Economics* 121 (2): 368–91.
- Alix-Garcia, Jennifer M., Elizabeth N. Shapiro, and Katharine R.E. Sims. 2012. "Forest Conservation and Slippage: Evidence from Mexico's National Payments for Ecosystem Services Program." *Land Economics* 88 (4): 613–38. doi:10.3368/le.88.4.613.
- Alix-Garcia, Jennifer, Craig McIntosh, Katharine R.E. Sims, and Jarrod R. Welch. 2013. "The Ecological Footprint of Poverty Alleviation: Evidence from Mexico's Oportunidades Program." *Review of Economics and Statistics* 95 (2): 417–35. doi:10.1162/REST\_a\_00349.
- Andrews, Matthew R., Lant Pritchett, and Michael Woolcock. 2012. "Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA)." HKS Faculty Research Working Paper Series RWP12-036, John F. Kennedy School of Government, Harvard University.
- Asaduzzaman, M., Mohammad Yunus, A.K. Enamul Haque, A.K.M. Abdul Malek Azad, Sharmind Neelormi, and Md. Amir Hossain. 2013. "Power from the Sun: An Evaluation of Institutional Effectiveness and Impact of Solar Home Systems in Bangladesh." Bangladesh Institute of Development Studies, Dhaka.
- Azofeifa, Arturo Sanchez, Alexander Pfaff, Juan Andres Robalino, and Judson P. Boomhower. 2007. "Costa Rica's Payment for Environmental Services Program: Intention, Implementation, and Impact." *Conservation Biology* 21 (5): 1165–73. doi:10.1111/j.1523-1739.2007.00751.x.
- Bamberger, Michael, Vijayendra Rao, and Michael Woolcock. 2016. "Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development." Policy Research Working Paper No. 5245, World Bank, Washington, DC.
- Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348 (6236). doi:10.1126/science.1260799.
- Barooah, Bidisha, Bharat Kaushish, and Jyotsna Puri. 2017. "Understanding Financial Risks for Smallholder Farmers in Low- and Middle-Income Countries: What Do We Know and Not Know?" 3ie Scoping paper 9, International Initiative for Impact Evaluation, New Delhi.
- Barrera-Osorio, Felipe, Marianne Bertrand, Leigh L. Linden, and Francisco Perez-Calle. 2011. "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia." *American Economic Journal: Applied Economics* 3 (2): 167–95. doi:10.1257/app.3.2.167.
- Basu, Kaushik. 2013. "The Method of Randomization and the Role of Reasoned Intuition." Policy Research Working Paper No. 6722, World Bank, Washington, DC.
- Börner, Jan, Kathy Baylis, Esteve Corbera, Driss Ezzine-de-Blas, Paul J. Ferraro, Jordi Rosés, Renaud Lapeyre, U. Martin Persson, and Sven Wunder. 2016. "Emerging Evidence on the Effectiveness of Tropical Forest Conservation." *PLOS ONE*. doi:10.1371/journal.pone.0159152.

- Bos, Astrid B., Amy E. Duchelle, Arild Angelsen, Valerio Avitabile, V. de Sy, Martin Herold, Shijo Joseph, et al. 2017. "Comparing Methods for Assessing the Effectiveness of Subnational REDD+ Initiatives." *IOP Science: Environmental Research Letters* 12 (7). doi:10.1088/1748-9326/aa7032.
- Brody, Carinne, Thomas de Hoop, Martina Vojtkova, Ruby Warnock, Megan Dunbar, Padmini Murthy, and Shari Dworkin. 2016. "Economic Self-Help Group Programs for Improving Women's Empowerment." *3ie Systematic Review* 23, International Initiative for Impact Evaluation, London.
- Buck, Emmy De, Hans Van Remoortel, Karin Hannes, Thashlin Govender, Selvan Naidoo, Bert Avau, Axel Vande Veegaete, et al. 2017. "Promoting Handwashing and Sanitation Behaviour Change in Low- and Middle-Income Countries: A Mixed-Method Systematic Review." *Systematic Review* 36, International Initiative for Impact Evaluation, London.
- Burwen, Jason, and David Levine. 2012. "A Rapid Assessment Randomized-Controlled Trial of Improved Cookstoves." *Journal of the International Energy Initiative* 16 (3).
- Cameron, Drew, Anjini Mishra, and Annette Brown. 2016. "The Growth of Impact Evaluation for International Development: How Much Have We Learned?" *Journal of Development Effectiveness* 8 (1): 1–21. doi:10.1080/19439342.2015.1034156.
- Chomitz, Kenneth M., and David A. Gray. 1996. "Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize." *World Bank Economic Review* 10 (3): 487–512. doi:10.1093/wber/10.3.487.
- Cropper, Maureen, Jyotsna Puri, and Charles Griffiths. 2001. "Predicting the Location of Deforestation: The Role of Roads and Protected Areas in North Thailand." *Land Economics* 77 (2): 172–86. doi:10.2307/3147088.
- Doocy, Shannon, and Hannah Tappis. 2016. "Cash-Based Approaches in Humanitarian Emergencies." *3ie Systematic Review Report* 28, International Initiative for Impact Evaluation, London.
- King, Elisabeth, Cyrus Samii, and Birte Snilstveit. 2010. "Interventions to Promote Social Cohesion in Sub-Saharan Africa." *Synthetic Review* 002, International Initiative for Impact Evaluation.
- Leroy, J.L., P. Gadsden, and M. Guijarro. 2012. "The Impact of Daycare Programs on Child Health, Nutrition and Development in Developing Countries." *3ie Systematic Review* 7, International Initiative for Impact Evaluation, London.
- Levin, Kelly, Benjamin Cashore, Steven Bernstein, and Graeme Auld. 2012. "Overcoming the tragedy of super wicked problems: constraining our future selves to ameliorate global climate change." *Policy Sciences* (Springer US) 45 (2): 123-152. doi:10.1007/s11077-012-9151-0.
- Menon, Purnima, Namukolo M. Covic, Paige B. Harrigan, Susan E. Horton, Nabeeha M. Kazi, Sascha Lamstein, Lynnette Neufeld, Erica Oakley, and David Pelletier. 2014. "Strengthening Implementation and Utilization of Nutrition Interventions through Research: A Framework and Research Agenda." *Annals of the New York Academy of Sciences*. doi:10.1111/nyas.12447.
- Miranda, Jorge, Shayda Sabet, and Annette Brown. 2016. "[Is Impact Evaluation Still on the Rise?](#)" *International Initiative for Impact Evaluation* blog.
- ODI (Overseas Development Institute). 2016. "Analysis of Resilience Measurement Framework and Approaches." Brief.
- PLOS Collections. 2016. "[Measuring Forest Conservation Effectiveness](#)." PLOS.
- Pritchett, Lant, and Justin Sandefur. 2014. "Context Matters for Size: Why External Validity Claims and Development Practice Do Not Mix." *Journal of Globalization and Development* 4 (2): 161–97. doi:10.1515/jgd-2014-0004.

- Puri, Jyotsna. 2017. "Development Priorities: What Are We Learning about What We Know?" Presentation at the African Evaluation Association Meetings, Kampala, March 27–31.
- Puri, Jyotsna, Anastasia Aladysheva, Vegard Iversen, Yashodhan Ghorpade, and Tilman Brück. 2014. "What Methods May Be Used in Impact Evaluations of Humanitarian Assistance?" 3ie Working Paper 22, International Initiative for Impact Evaluation, New Delhi.
- . 2017. "Can Rigorous Impact Evaluations Improve Humanitarian Assistance?" *Journal of Development Effectiveness* 9 (4): 519–42. doi:10.1080/19439342.2017.1388267.
- Puri, Jyotsna, Francis Rathinam, and Ritwik Sarkar. 2017. "Real World Impact Evaluations—What Are We Learning." Draft Paper, International Initiative for Impact Evaluation, New Delhi.
- Puri, Jyotsna, Megha Nath, Raag Bhatia, and Louise Glew. 2016. "Examining the Evidence Base for Forest Conservation Interventions." Evidence Gap Map Report, International Initiative for Impact Evaluation, New Delhi.
- Rao, Vijayendra, and Michael Woolcock. 2002. "Integrating Qualitative and Quantitative Approaches in Program Evaluation." In *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*, Francois Bourguignon and Luiz A Pereira da Silva, eds., Chapter 8. New York: Oxford University Press.
- Ravallion, Martin. 2016. "How Can We Better Address the Gaps in Our Knowledge about Development Effectiveness?" In *The Oxford Handbook of Professional Economic Ethics*, George DeMartino and Deirdre McCloskey, eds. doi:10.1093/oxfordhb/9780199766635.013.025.
- Schwandt, Thomas, Zenda Ofir, Dorothy Lucks, Kassem El-Saddick, and Stefano D'Errico. 2016. "Evaluation: A Crucial Ingredient for SDG Success." IIED Briefing, International Institute for Environment and Development, London.
- Sills, Erin O., Diego Herrera, Justin A. Kirkpatrick, Amintas Brandão Jr, Rebecca Dickson, Simon Hall, Subhrendu Pattanayak, et al. 2015. "Estimating the Impacts of Local Policy Innovation: The Synthetic Control Method Applied to Tropical Deforestation." *PLOS ONE* 10 (7). doi:10.1371/journal.pone.0132590.
- Snilstveit, Birte, Jennifer Stevenson, Daniel Phillips, Martina Vojtkova, Emma Gallagher, Tanja Schmidt, Hannah Jobse, Maisie Geelen, Maria Grazia Pastorello, and John Eyers. 2015. "Interventions for Improving Learning Outcomes and Access to Education in Low- and Middle-Income Countries." 3ie Systematic Review 24, International Initiative for Impact Evaluation, London.
- Snilstveit, Birte, Jennifer Stevenson, Radhika Menon, Daniel Phillips, Emma Gallagher, Maisie Geelen, Hannah Jobse, Tanja Schmidt, and Emmanuel Jimenez. 2016a. "The Impact of Education Programmes on Learning and School Participation in Low- and Middle-Income Countries." 3ie Systematic Review Summary, International Initiative for Impact Evaluation, London.
- Snilstveit, Birte, Jennifer Stevenson, Paul Fenton Villar, John Eyers, Celia Harvey, Steven Panfil, Jyotsna Puri, and Madeleine McKinnon. 2016b. "Land-Use Change and Forestry Programmes: Evidence on the Effects on Greenhouse Gas Emissions and Food Security." International Initiative for Impact Evaluation, London.
- Somanathan, E., R. Prabhakar, and Bhupendra Singh Mehta. 2009. "Decentralization for Cost-Effective Conservation." *Proceedings of the National Academy of Science of the United States of America* 106 (11): 4143–47. doi:10.1073/pnas.0810049106.

- Tachibana, Towa, and Sunit Adhikari. 2009. "Does Community-Based Management Improve Natural Resource Condition? Evidence from the Forests in Nepal." *Land Economics* 85 (1): 107–31. doi:10.3368/le.85.1.107.
- UNESCO (United Nations Educational, Scientific and Cultural Organization). 2015. *Education for All 2000–2015: Achievements and Challenges*. Paris: UNESCO.
- Waddington, Hugh, and Howard White. 2014. "Farmer Field Schools: From Agricultural Extension to Adult Education." Systematic Review Summary, International Initiative for Impact Evaluation.
- Woolcock, Michael. 2009. "Toward a Plurality of Methods in Project Evaluation: A Contextualised Approach to Understanding Impact Trajectories and Efficacy." *Journal of Development Effectiveness* 1 (1): 1–14. doi:10.1080/19439340902727719.
- . 2013. "Using Case Studies to Explore the External Validity of Complex Development Interventions." *Evaluation* 19 (3): 229–48. doi:10.1177/1356389013495210.
- World Bank. 2016. *World Bank Annual Report 2016*. Washington, DC: World Bank. doi:10.1596/978-1-4648-0852-4.
- Wynes, Seth, and Kimberly A. Nicholas. 2017. "The Climate Mitigation Gap: Education and Government Recommendations Miss the Most Effective Individual Actions." *IOP Science: Environmental Research Letters* 12 (7). doi:10.1088/1748-9326/aa7541.





---

## Chapter 22

# Measuring the Impact of the Extractive Industry's Development Projects

Gwendolyn Wellmann

**Abstract.** *In the past decade and a half, increasing pressure has been put on big corporations and, in particular, on the extractive industries—mining, oil and gas, and forestry—to go beyond philanthropy, and to make social contributions that contribute directly to society's development, particularly in ways that help to achieve international development goals. At the same time, an increase in social unrest directly linked to mining operations has led to an increased realization that contributing to the development of affected communities is an important risk-avoidance strategy. While many community development projects have been implemented over several years, impact assessments were not seen as necessary; and it has only been where there has been increased pressure from governments, such as Ghana's, that impact assessments of these projects have been done. This chapter discusses the impact of evaluation methods used by a mining company in Ghana to measure the impact of their community/societal development program. The chapter concludes with an exploration of the way forward for impact evaluations of the development activities of big corporations.*

The term “extractive industry” refers to any industry that extracts resources from the Earth: it mainly refers to mining, oil and gas, and to a lesser degree, forestry. The extractive industry is enormous and occupies a significant space in the economies of many resource-rich countries. This sector accounts for at least 20 percent of total exports, and at least 20 percent of government revenue, in 29 low-income and lower-middle-income countries (Smith 2012). In eight of these countries, the sector accounts for more than 90 percent of total exports, and 60 percent of total government revenue. Three of the world’s largest companies are extractive companies.

Although the sector is not necessarily more complex than other industrial or economic sectors, it carries with it significant and diverse economic, societal, and environmental implications and challenges. Over the past five decades at least, the economic and environmental implications have been tackled, and their mitigation has been legislated for the most part. The social implications have taken somewhat longer to raise hackles, perhaps because they are less visible than environmental degradation: it is only since 2002 that the extractive industry (primarily mining) has begun to mitigate some of these effects. Social implications refer to the socioeconomic circumstances and health of populations living in the vicinity of the mines. Reserves are often found in remote areas with limited economic activity and major social needs, and the industry has long-term horizons, with reserves depleted over several decades, which means that the mines or oil fields, and their cumulative social impacts, will be there for just as long.

The first section of this chapter reports on the use of monitoring and evaluation (M&E) in the extractive industry’s community development projects in West and East Africa, as observed during the period 2002–12: the second section deals with the assessment of the impact of community development initiatives implemented by a mining company operating in Ghana, West Africa. The third section explores the way forward for impact evaluations of the development activities of big corporations.

## **M&E IN EXTRACTIVE INDUSTRY DEVELOPMENT PROJECTS**

The community development projects referred to in this section were implemented by mining, and oil and gas, companies in West and East Africa. These development projects were implemented by 10 operational mines and 3 exploration projects, which were owned by 5 multinational companies. The author worked for these companies in the role of independent contractor, and as such was involved in the development, implementation, and/or evaluation of the community development projects. As is the norm, nondisclosure agreements were signed with the companies before work could commence, and these agreements prevent the naming of the companies and/or the relevant projects discussed in this chapter. The development projects discussed were spread over four different countries, and were implemented during the period 2002–12.

## 2002–03

In the initial period (2002–03), the companies gave much more attention to environmental issues and impact on the environment, primarily because pressure from environmental advocates had started in the 1970s, and environmental issues were included in the mining codes of various countries in the decades that followed. Relationships with directly affected communities were not high on the priority list of most companies, and government officials and local traditional chiefs were the only persons considered to be local stakeholders. Community development was not on the priority list. Any infrastructure project that the company initiated was directly linked to the needs of the extractive project. These projects mostly involved the construction of roads, and the provision of electricity and potable water.

Toward the end of 2003, a few projects experienced social unrest as traditional environmentally focused nongovernmental organizations (NGOs) started to highlight the social impact of the projects, and stakeholders became more aware of their rights. The reputable extractive companies recognized that in order to be responsible corporate citizens, they had to address socioeconomic development issues at their operations. It was during this time that development initiatives within communities started to shift away from pure infrastructure to also include community capacity building, and local economic (livelihood) projects. At this time, none of the projects included any M&E. Infrastructure projects were monitored by engineers, and capacity-building and livelihood projects were monitored against a budgeted amount. There was no tracking of the number of beneficiaries, nor of the effectiveness or the impact of the project. Once the money was spent, the project was considered to have been implemented successfully by both the staff and the management of the companies.

## 2004–07

As governments in resource-rich countries matured, improved access to education and the Internet resulted in better-informed stakeholders, and as the International Finance Corporation's approach evolved to include promoting environmentally and socially sustainable growth in developing countries,<sup>1</sup> more countries started to legislate the mandatory implementation of community development projects by the extractive companies operating within their borders (IEG 2011). In 2005, the International Council on Mining and Metals (ICMM) released its *Community Development Toolkit*, to be used as a guide to implementing community investment by its member companies

---

<sup>1</sup> The International Finance Corporation (IFC) is the world's largest multilateral development bank, providing financial support and technical advice to private firms in developing countries. Although only one of the companies discussed here is an IFC client, all the companies adhere to IFC guidelines (referred to as IPs—performance standards for best practice).

globally.<sup>2</sup> This toolkit did contain a section on M&E, but it was very basic; and even though the ICMM released an updated toolkit in 2012, the M&E section was not updated, and the assessment or evaluation of the impact of community development projects was not covered.

By 2007, most mining companies had moved away from seeing community development projects as philanthropy, and had begun to see them more as an essential part of their risk-avoidance activities. Companies started to spend hundreds of thousands of dollars on identifying social risks and implementing community development projects, but barely a thousand dollars to measure the impact of their development projects. Similarly, including community development projects as a means for avoiding social unrest did not change the methods of resource extraction activities, which remain governed by cost effectiveness and return on investment (ROI) considerations.

In Ghana, three of the companies decided to outsource community development projects to professionals, and appointed international NGOs to implement them.<sup>3</sup> These community development projects focused on the education, health, and local economic development sectors, and were similar across all three companies. The NGOs introduced M&E, but it was limited to preset indicators developed by the NGOs without input from either the affected community members or the company. The tools used were standardized M&E procedures and tools.

## 2008–12

After the initial four-year contract ended, the companies decided not to outsource this function anymore, but rather to employ development professionals. As the external evaluations conducted on the programs implemented during 2004–07 indicated that there was a need for more engagement with community members when choosing and designing development projects, this became a priority. Most of the development professionals employed were trained in community liaison and external engagement: thus there remained a dearth of M&E skills across the board. The companies required data about only these indicators: number of beneficiaries; amount expended; number of social complaints directly related to the company's operations; and number of social unrest incidents, irrespective of cause. This was the full extent of any M&E.

In 2012, the government of Ghana, as part of the Ghana Environmental Protection Agency's (EPA's) Akoben Programme,<sup>4</sup> for the first time ever demanded a report on the impact of community development initiatives on affected communities. The Akoben Programme used a rating system

---

<sup>2</sup> All the mining companies referred to in this chapter belonged to the ICMM between 2002 and 2014.

<sup>3</sup> One international NGO was contracted by several gold mining companies to provide this service.

<sup>4</sup> <http://www.epaghanaakoben.org/>.

(figure 22.1) to assess the environmental standing of an operating mine or mining project (e.g., the development of, or expansion of, a mine).<sup>5</sup> The results were published in all the leading newspapers, and the competition for praise (and/or shame) among the companies was strong.

FIGURE 22.1 **Akobon Programme rating system**

Rating level	Performance	Implications
Red	Poor	Serious risks
Orange	Unsatisfactory	Not in compliance
Blue	Good	In compliance
Green	Very good	Applies best practices
Gold	Excellent	Committed to social performance

## IMPACT ASSESSMENT IN GHANA

### Mining in Ghana

The second-largest gold deposit in Africa is located in Ghana, and the historical importance of gold mining in the economic development of the country is considerable and well documented.<sup>6</sup> Large-scale industrial gold mining in Ghana dates back to the last quarter of the 19th century. It was restructured and modernized under the post-1983 Economic Recovery Programme, after a period of decline under government control in the nationalist era in the two decades since the early 1960s (Hilson 2002, 2004). Some of these changes included a revised mining code (the Minerals and Mining Law [PNDC 153] of 1986), and resulted in the sector seeing sustained increases in foreign investment, output, and export volumes. Investment increased substantially between 2006 and 2009, facilitated by a further revised mining code, which was consolidated in the 2006 Minerals and Mining Act 703. Under this law, all minerals are owned by the state, and the holder of the mining lease must pay a royalty to the state of not less than 3 percent and not more than 6 percent of their gross revenues. In addition to paying royalties, mining companies also contribute to taxes, employment, contracting, and investing in community development.

The gold mining sector contributes significant amounts to the global economy through their production activities and expenditure on goods and services, but the socioeconomic impacts of this sector are not well understood. The direct economic contribution of the gold mining industry to the world economy during 2013 was over \$171.6 billion, which is almost seven

<sup>5</sup> Each mine or mining project was assessed individually, irrespective of how many mining projects a company owned.

<sup>6</sup> See Agbesinyale (2003); Akabzaa, Seyire, and Afriyie (2007); Hilson (2002, 2004); and Kesse (1985).

times its contribution in 2000.<sup>7</sup> The biggest in-country expenditures are for suppliers and employee wages.

The gold mining sector in Ghana contributes a significant amount of funding toward supporting development, and it is useful at this point to highlight both the sector's contribution and that of international aid to the country between 2000 and 2012. Official international aid received by Ghana increased by 202 percent during the period 2000–12, rising from \$598.2 million to \$1,807.9 million (Stamp 2015). In contrast, direct gross value added amounts during the same period increased by 1,174 percent: from \$273 million in 2000 to \$3,476.4 million in 2012 (Stamp 2015).<sup>8</sup>

## Two Mines and Their Community Investment

The mines that are the subject of this chapter, and which belong to one company, are located in the Birimian and Tarkwaian gold belts, which characterize the western half of Ghana.<sup>9</sup> Although nowadays companies have large community relations departments and sophisticated manuals for stakeholder engagement, community development, and impact mitigation, in Ghana in 2002, on one of Africa's biggest and most productive mines, none of this existed. It can be correctly assumed that this was the case at most mines in West Africa, if not in the world.

However, as discussed above, local and international events, and especially increased social risks, left the companies no other alternative than to start addressing social issues. The initial corporate social responsibility projects were primarily community-level infrastructure projects: hospitals, communal toilets, schools, and roads. There was very little understanding that these buildings meant nothing if people were not using them and benefiting from them. There was confusion as to why the recipients were not grateful to the companies for providing them with these buildings, and specifically in Ghana, some meetings were held among mining companies to address this shared problem. It became clear that miners are best at mining, not at development work, and that they needed people with expertise in development. At this time, any monitoring of any social project was being done by the engineers, and was related to the building of the buildings or roads, and the amount of money being expended: there was no measurement dealing with the number

---

<sup>7</sup>The contribution can be assessed by calculating the gross value added (GVA), which is a calculation that estimates the contribution of industrial activity to a nation's gross domestic product (GDP). It is important to note that GVA does not refer to production, but rather economic value, nor does it refer to profit.

<sup>8</sup>Direct gross value added (GVA) estimates the economic value of the gold mining industry's production to the Ghanaian economy. Indirect GVA estimates the value of economic production resulting of the industry's expenditures on raw materials, good and services.

<sup>9</sup>While the company has provided permission to use the data presented in this study, permission is granted on the basis that the company and the mines will remain anonymous.

of beneficiaries, and there was definitely no thought given to social impact, whether positive or negative.

The company that is the subject of this chapter began by implementing infrastructure projects in the communities that had been most affected by its mining activities since 1999. Initially, it adopted an ad hoc approach by simply responding to requests made by the communities. This approach changed when in 2002 a trust fund, which two years later became a foundation, was set up in the names of the affected communities to ensure that sufficient funds were set aside annually for the development of these communities. One of the first tasks of the foundation was to develop a five-year community development plan to ensure a coherent approach to the provision of infrastructure to the affected communities.

The vision of the five-year plan was to be a high-impact, results-focused, sustainable, and integrated community development program that focused on economic growth, quality of life improvement, and empowerment through infrastructure development and capacity building. Its immediate goal was to improve the quality of life for 30,000 people in the 16 primary stakeholder communities by 2010. An international NGO was contracted to develop the five-year plan and to implement it.

Each of the individual projects that formed part of the five-year program had its own M&E logical framework (logframe). The M&E involved preset indicators developed without input from either the affected communities (the beneficiaries) or the client (the company). The NGO was using standardized procedures and tools. With the passing years, however, and with the realization that the clients were primarily interested in being able to report on the amount of U.S. dollars being spent, and how many people were being directly assisted, those two indicators became the main measurements used by the implementer.

The evaluations involved external experts coming in to measure performance against preset indicators, using standardized procedures and tools: there was no focus on the impacts (whether intended or unintended, positive or negative) of any of the projects. The evaluation reports of six community development programs implemented by various mining companies in the country indicated that the M&E of the projects, as reflected in the logframe, appeared to be an afterthought. Although the logframes were well executed, with the objectives, objectively verifiable indicators, means of verification, and assumptions well laid out, one got the sense that this was done merely in order to tick a box, and that the logframe was never again looked at until it was studied by the external evaluator. Three of the five-year program's project planning documents contained no logframes, and there was only a brief paragraph referring to M&E.

## The Impact Assessment

Two years after the completion of the implementation of the community development program, this company, like many others in Ghana, was obligated to report on the impact of its community investment projects as part of the Ghana EPA's Akoben Programme. It therefore wanted to measure

not only the impact of the five-year program, but also the company's earlier infrastructure development projects. It was at this time that the lack of any baseline data was discovered. While each of the projects had an M&E plan, there had been none for the overall program, and no baseline data had been collected before the program was initiated.

As there were only limited other resources that could be used to create a baseline against which the program's impact could be assessed, the impact assessment had to adopt a "before" and "after" methodology that was designed to quantitatively and qualitatively determine the program's outputs and impacts on the communities.

Four impact assessment criteria were used:

- Individual project relevance and appropriateness
- Status of individual project implementation
- Changes in the community's access to education, water and sanitation and health care
- Individual project impact and its sustainability

Structured questionnaires were used to collect data from a sample of 990 households randomly selected across 16 communities from a total number of 11,677 households. In an attempt to also gather qualitative data by using the most significant change approach, other data collection techniques included focus group discussions and key informant interviews (Davies and Dart 2005; Serrat 2009). Data gathered were validated with available information from the local government.

## Results

The results of the assessment were as follows:

- **Appropriateness and relevance of projects.** All of the different types of infrastructure projects implemented in the communities were found to reflect the felt needs of the beneficiary communities. They also reflected the policy objectives of the Tarkwa Nsuaem Municipality, the Prestea Huni Valley District Assembly, and the central government, as well as the United Nations' Millennium Development Goals (MDGs).
- **Status of implementation.** All of the projects had been completed at the time of the assessment.
- **Changes in community access.** All of the communities' access to education, health, and water and sanitation services were significantly improved.
- **Impacts and sustainability.** There was a general consensus among stakeholders interviewed within the communities as well as at the district level that the company's interventions had led to significant improvements in the provision of quality infrastructure to the communities.



Significant improvements are reflected in increased access to basic education, health care, water and sanitation services, road transportation, and other socioeconomic facilities in the beneficiary communities. This increased access has had a positive impact on the living conditions of residents in terms of improved enrollment in schools, standard of educational achievement among pupils, a reduction in morbidity, an enhanced image of the communities, and increased productivity. These results from the household survey were validated using supplementary information supplied by the government agencies responsible for health, education, and water and sanitation. For instance, the Bompieso Junior High School recorded a pass rate of 86 percent of the students in 2002, which had improved to a 100 percent pass rate in 2011. Similarly, the 2002 pass rate at the Damang Junior High School was only 24.5 percent, which had improved significantly in 2011, with a 95 percent pass rate.

Participant perceptions of changes in access to education are shown in table 22.1; their perceptions of the impacts of educational infrastructure are shown in table 22.2.

In terms of health care, the awareness and practice of family planning was very low in the communities prior to the introduction of the health interventions. Records from the health directorates confirmed an increase in acceptance of family planning. The Prestea Huni Valley Health Directorate reported that family planning acceptors increased by 6 percent between 2008 and 2010, and the Tarkwa Nsuaem Health Directorate reported an increase in family planning in the community of New Atuabo from 1007 patients in 2009 to 1225 in 2010. The research respondents indicated that the quality of health care has improved after the company's health interventions, and that as a result community members are healthier.

Table 22.3 indicates the perception of respondents as to how the interventions have translated into improving health care facilities in the communities, and the overall effect on the communities' well-being.

It is expected that the provision of health care facilities will to some degree impact on health education. The assessment therefore examined the level of knowledge of participants in relation to health care. Results indicate that whereas 27.7 percent of the respondents reported some basic knowledge about health care prior to the company's intervention, the situation has improved significantly, to 80.4 percent after the intervention. The percentage of respondents who reported that health education was bad in their communities decreased from 72.3 percent prior to intervention, to 19.6 percent post-intervention.

It is, however, important to note that the increase in health education cannot be attributed solely to the company's intervention, since Ghana Health Services had also been involved in health programs in the area at the same time.

Access to health care facilities, medications, and health education is expected to have a positive impact on the incidence of diseases in an area. The director of health services at Prestea Huni Valley reported that disease incidence had been reduced by approximately 65 percent in the district, while

TABLE 22.1 Perception of improvements before project implementation and after project completion

Indicator	% of respondents reporting (before)					% of respondents reporting (after)				
	Very good	Good	Bad	Very bad	Total	Very good	Good	Bad	Very bad	Total
Classroom condition	1.0	25.0	64.5	9.5	100.0	34.6	56.5	1.5	7.4	100.0
Maintenance	0.6	28.2	63.4	7.8	100.0	23.9	69.0	6.3	0.8	100.0
Building quality	0.5	26.2	61.8	11.5	100.0	39.5	58.3	2.0	0.2	100.0
Teacher numbers	1.1	35.6	55.5	7.8	100.0	25.2	66.5	7.9	0.4	100.0
Distance	Very far	Far	Close	Very close	Total	Very far	Far	Close	Very close	Total
	6.4	21.9	63.2	8.5	100.0	1.3	7.4	74.1	17.2	100.0

SOURCE: Field Survey, 2012.

TABLE 22.2 Impacts of educational infrastructure

Indicator	% of respondents reporting (before)					% of respondents reporting (after)				
	Very good	Good	Bad	Very bad	Total	Very good	Good	Bad	Very bad	Total
Enrollment	1.1	32.0	59.3	7.6	100.0	32.7	53.8	12.2	1.3	100.0
Performance	1.8	34.1	57.3	6.8	100.0	32.2	64.8	2.7	0.3	100.0
Aesthetics	1.0	30.4	60.5	8.1	100.0	31.7	65.3	3.0	0.0	100.0
Interest	1.2	34.0	56.1	8.7	100.0	30.2	64.9	4.7	0.2	100.0
Community image	1.2	35.3	57.8	5.7	100.0	29.6	64.9	5.0	0.5	100.0

SOURCE: Field Survey, 2012.

health care managers at the community level all indicated that there had been no outbreaks of epidemics in the communities in recent years.

There is a positive correlation between morbidity and mortality: so with a decline in morbidity, mortality was expected to fall, and records from Prestea Huni Valley Health Directorate confirmed that they did. Records from the directorate indicate that maternal mortality in the district had dropped from 57 per 100,000 live births in 2000 to 43.3 per 100,000 live births in 2009.

Many institutions contribute toward the development of health care delivery, so it was important to determine the perception of the respondents in terms of which institutions were responsible for the improvement in health care delivery in the area. According to respondents, the improvement in health conditions could be associated with many institutions: the municipalities (the assemblies); the company; other mining companies; the community; the central government; and private health care providers. However, the greatest recognition was given to the company. As many as 414 of 990 respondents believed that the improvement in health conditions in their community is as a result of the company's interventions, followed by the Assembly, the central government, private health care providers, and other mining companies, in that order.

One of the key indicators of access to water is proximity (or distance) to a safe, potable water source. Results of the study show that the company-funded water projects had led to significant improvements in this respect. During the household survey, respondents were asked to describe the distance they covered to reach a water source both before and after the company constructed a water facility in their community (figure 22.2).

In addition to distance, household respondents were also asked to assess the adequacy and reliability of the water supply in their community before and after the intervention (table 22.5).

Until the company provided them with potable water sources, most of the communities relied on water collected from streams and rivers, rendering them susceptible to water-borne diseases (figure 22.3).

Approximately 49 percent of respondents said that prior to the company's intervention, the distance to their toilet facility was either "close" or "very close," while the remaining 51 percent described it as either "far" or "very far." After the intervention 88 percent of respondents indicated that their toilet facility was either "close" or "very close" (figure 22.4).

Respondents' assessment of the level of access, adequacy, and reliability of toilet facilities in their community before and after the company's intervention produced results similar to those described above (table 22.6).

## WHAT IS THE WAY FORWARD?

Despite the lack of baseline data, and complicating factors such as the implementation of similar projects by other companies, aid agencies, and the government in the same communities at the same time, one could draw the conclusion that the development projects implemented by the company did

TABLE 22.3 Health indicators

Indicator	% of respondents reporting (before)					% of respondents reporting (after)				
	Very good	Good	Bad	Very bad	Total	Very good	Good	Bad	Very bad	Total
Health education	0.7	27.0	61.1	11.2	100.0	15.7	64.7	17.4	2.2	100.0
Family planning	2.0	27.3	61.3	9.4	100.0	17.0	57.9	22.9	2.2	100.0
Disease incidence	1.8	24.5	64.1	9.6	100.0	14.8	65.4	17.6	2.2	100.0

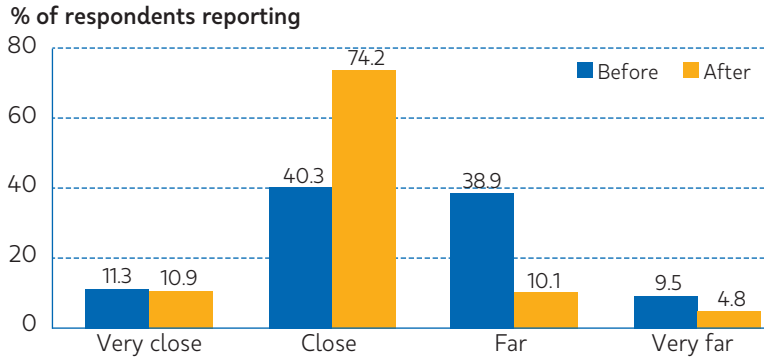
SOURCE: Field Survey, 2012.

TABLE 22.4 Entity responsible for overall improvement of the health conditions

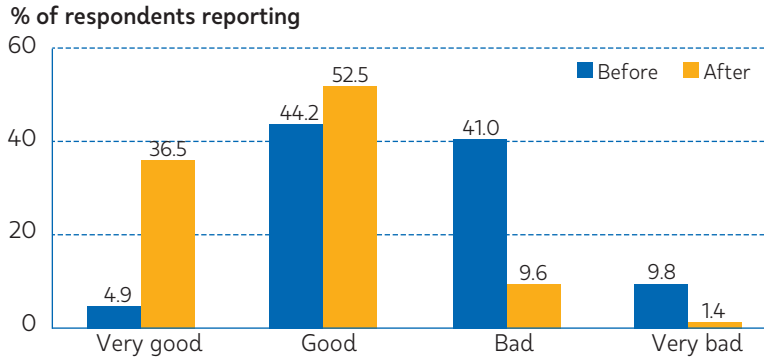
Condition	Company		Community		Municipal assembly		Another mining company		Private health care provider		Central government	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Improvement	414	70.2	288	69.2	272	63.1	26	70.3	96	64.9	265	66.3
Deterioration	176	29.8	128	30.8	159	36.9	11	29.7	52	35.1	135	33.7
Total	590	100.0	416	100.0	431	100.0	37	100.0	148	100.0	400	100.0

SOURCE: Field Survey, 2012.

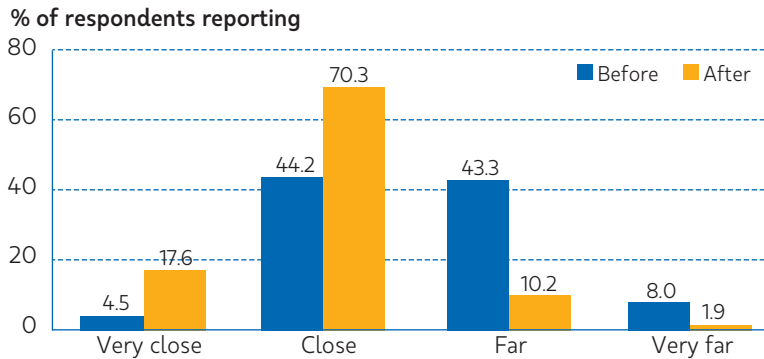
**FIGURE 22.2 Percentage of respondents reporting various distances to water source before and after company intervention**



**FIGURE 22.3 Percentage of respondents reporting various levels of water quality before and after company intervention**



**FIGURE 22.4 Percentage of respondents reporting various distances to toilet facility before and after company intervention**



**TABLE 22.5 Household respondents' assessment of access, adequacy, and reliability of water before and after company intervention**

Indicator	% of respondents reporting (before)					% of respondents reporting (after)				
	Very good	Good	Bad	Very bad	Total	Very good	Good	Bad	Very bad	Total
Access	6.0	49.3	37.1	7.6	100.0	35.5	55.0	8.6	0.9	100.0
Adequacy	6.3	50.2	38.8	4.7	100.0	25.8	49.5	21.0	3.7	100.0
Reliability	6.8	47.4	38.9	6.8	100.0	23.6	51.6	20.4	4.4	100.0

SOURCE: Field Survey, 2012.

**TABLE 22.6 Household respondents' assessment of access, adequacy, and reliability of toilet facilities before and after company intervention**

Indicator	% of respondents reporting (before)					% of respondents reporting (after)				
	Very good	Good	Bad	Very bad	Total	Very good	Good	Bad	Very bad	Total
Access	3.0	36.3	51.8	8.9	100.0	23.7	68.0	7.9	0.4	100.0
Adequacy	2.8	29.4	57.0	10.8	100.0	18.2	68.7	11.6	1.5	100.0
Reliability	2.5	29.2	57.9	10.4	100.0	17.8	71.0	9.6	1.6	100.0

SOURCE: Field Survey, 2012.

have a positive impact, and did contribute to the country's MDGs. Nevertheless, the situation does raise some concerns.

The challenge of a lack of baseline social and economic data against which the effectiveness of socioeconomic development initiatives can be measured, while not unique to the extractive industry, is one that needs to be addressed.

A critical issue for the extractive industry remains the need to secure social license to operate.<sup>10</sup> This often results in heavy investment in improving socioeconomic conditions of affected and/or host communities. It is also becoming the norm for companies to make targeted investments that focus on the same social and economic challenges that national governments are also seeking to address. More and more extractive companies are becoming aware of the unprecedented focus on the role of business in attaining the Sustainable Development Goals (SDGs), and there is also more and more pressure on companies to report on their impact on society, whether good or bad, through reporting platforms such as the Global Reporting Initiative.<sup>11</sup> At the same time, despite a history of sometimes antagonistic relationships, some governments have started to engage with extractive companies as potential partners in development, and this dialogue has opened up new possibilities for these companies to play an active development role in developing countries. Many of these countries are also heavily reliant on official aid from donor countries, although the rate of growth in the economic value of the extractive product is significantly higher than the aid received. In recent years, the Ghanaian government has opened a dialogue with the extractive industries operating within the country to see how development could be leveraged. While this initially took the form of requiring the companies to report only on their spending on aspects of the country's development priorities, in recent years there has been a demand that the company also report on the impact of their community development investments. This has in turn opened the question of how this can be measured, especially in a milieu where development initiatives come from several different mining companies operating in a small area (e.g., the Tarkwa Nsuaem Municipality<sup>12</sup>), and often initiate projects in the same communities, while government departments are

---

<sup>10</sup> "Social license to operate" generally refers to a local community's ongoing acceptance and approval of a company's project and/or the company's continued presence in the area. It is now internationally recognized as a prerequisite to development of any project. Conflict between extractive companies and local communities can result in operations being disrupted by protests, damage to property and other violent incidents. Franks et al. (2014) found that conflicts between mining companies and communities can cost the company around \$20 million per week as a result of production delays, and Henisz, Dorobantu, and Nartey (2011) found that between 1993 and 2008, the estimated value of two-thirds of the gold controlled by 26 gold mining companies, owned by 19 publicly traded companies, was related to the companies' management of external relationships with affected communities and host governments.

<sup>11</sup> Global Reporting Initiative: <https://www.globalreporting.org>.

<sup>12</sup> Both gold and manganese are mined in this municipality.

also continuing to do their normal development work. Although it is not the case in the Tarkwa Nsuaem Municipality, there are also often aid agencies and NGOs implementing similar development projects. How then does one measure the impact of a single company's community investments?

In the evaluation discussed in this chapter, the evaluators tried to assess the opinion of the respondents as to whom (i.e., a specific company or government agency) any specific improvement in their access to health care, education, and water and sanitation could be attributed. In all instances, the majority of the respondents attributed the changes to the company that was conducting the research, but in "shared" communities (i.e., communities where more than one mining company was affecting the community), the results were a mixed bag of different companies. Very few respondents indicated that government initiatives were responsible for their improved quality of life.

As we venture into the future with much more emphasis on the role of business in development, and business's contribution to the SDGs, and with more pressure than ever for companies to report on their community impact to shareholders, stakeholders, and the public in general, more precise measuring and evaluating tools will be required. A good start would be a reliable and accurate baseline: if that is not possible, the natural experiment study methodology could work. While companies cannot take responsibility for a lack of national or regional data, they can work in partnership with communities and other stakeholders, such as local universities, to support systematic data collection to either build or update existing data sets, and the analysis thereof.

Companies should also endeavor to employ not only development professionals, but also M&E professionals. These professionals will have a difficult task, as the M&E currently being implemented in the extractive industry's development projects only monitors the implementation of the project, not its impact. Evaluation has asked only one question: "Has the project been implemented successfully?" It is impossible to effectively measure impact without asking the right questions, and without taking the increasing unpredictability of results due to social volatility and climate change into account, and developing a more flexible and dynamic approach.

The main question, however, is whether evaluation practices are equipped to take on the issues that the extractive industry will encounter, such as weaknesses in governance, extreme poverty, inequality, economic disparities, and social exclusion. How does one accurately measure human well-being (or improved well-being) as an outcome of a community development investment? How many companies, even those that employ development professionals, would be aware of McGregor and Sumner's three-dimensional model, based on Sen's concept of development, as the freedom to realize human capabilities (McGregor and Sumner 2010; Sen 1999, 2009)? How many evaluators would know how to apply it?

In terms of independent evaluations, which are still being paid for by the companies, how does one go beyond client-controlled guidelines that do not really allow for much independence, nor allow divergence from evaluating projects or programs against predetermined goals, goals which might have been formulated in an era of less awareness of the social justice issues



pertaining to the industry and its operations? Similarly, how does one do a completely independent evaluation in what can sometimes be a hostile environment?

Picciotto calls very strongly for progressive and adaptive evaluation, which is “based on values and geared to public interest and combines the vision of democratic, committed, morally engaged evaluation with an emphasis on results that serve the public interest” (Picciotto 2016, 274). The extractive industry’s development initiatives must be evaluated using these concepts in order to report in a realistic way the companies’ enormous impact on communities, and their contributions to the global and host country SDGs.

## REFERENCES

- Agbesinyale, Patrick. 2003. “Ghana’s Gold Rush and Regional Development: The Case of the Wassa West District.” *Journal of Ghana Science Association* 7 (1): 38–49.
- Akabzaa, T.M., J.S. Seyire, and K. Afriyie. 2007. *The Glittering Façade: Effects of Mining Activities on Obuasi and Surrounding Communities*. Accra: Third World Network.
- Davies, Rick, and Jess Dart. 2005. “[The ‘Most Significant Change’ \(MSC\) Technique: A Guide to Its Use](#)”
- Franks, D.M., R. Davis, A.J. Bebbington, S. Alia, D. Kempa, and Martin Scurrah. 2014. “Conflict Translates Environmental and Social Risk into Business Costs.” *PNAS* 111 (21): 7576–81.
- Henisz, Witold J., Sinziana Dorobantu, and Lite Nartey. 2011. “[Spinning Gold: The Financial Returns to External Stakeholder Engagement](#).” The Wharton School, University of Pennsylvania.
- Hilson, Gavin. 2002. “Harvesting Mineral Riches: 1000 Years of Gold Mining in Ghana.” *Resources Policy* 28 (1–2): 13–26.
- . 2004. “Structural Adjustment in Ghana: Assessing the Impacts of Mining-Sector Reform.” *Africa Today* 5 (2): 53–72.
- ICMM (International Council on Mining and Metals). 2005. *Community Development Toolkit*. London: ICMM.
- IEG (Independent Evaluation Group). 2011. *Assessing IFC’s Poverty Focus and Results*. Washington, DC: World Bank.
- Kesse, Godfried, O. 1985. *The Mineral and Rock of Ghana*. A.A. Balkeme.
- McGregor, Allister, and Andy Sumner. 2010. “Beyond Business as Usual: What Might 3-D Well-Being Contribute to MDG Momentum?” *IDS Bulletin* 41 (1): 1–4.
- Picciotto, Robert. 2016. “Bringing Inequality Back in from the Cold: Toward a Progressive Evaluation Model.” In *Poverty, Inequality, and Evaluation: Changing Perspectives*, Ray C. Rist, Frederic P. Martin, and Ana Maria Fernandez, eds., 253–77. Washington, DC: World Bank.
- Sen, Amartya. 1999. *Development as Freedom*. Oxford, UK: Oxford University Press.
- . 2009. *The Idea of Justice*. Cambridge, MA: Harvard University Press.
- Serrat, Olivier. 2009. “[The Most Significant Change Technique](#).” Knowledge Solutions 25, Asian Development Bank, Manila.
- Smith, James L. 2012. “[Issues in Extractive Resource Taxation: A Review of Research Methods and Models](#).” IMF Working Paper WP/12/287. International Monetary Fund, Washington, DC.
- Stamp, Maxwell. 2015. *The Social and Economic Impacts of Gold Mining*. London: World Gold Council.

