UNDP
Empowered lives.
Resilient nations.
EVALUATION
FOR AGENDA 2030
PROVIDING EVIDENCE ON
PROGRESS AND SUSTAINABILITY
Rob D. van den Berg
Indran Naidoo
Susan D. Tamondong
Editors
INTERNATIONAL DEVELOPMENT EVALUATION ASSOCIATION
IDEAS
KNOWLEDGE CAPACITY DEVELOPMENT NETWORKING

Evaluation for Agenda 2030

Evaluation for Agenda 2030

PROVIDING EVIDENCE ON PROGRESS AND SUSTAINABILITY

EDITORS

Rob D. van den Berg
King's College London and International Development Evaluation Association (IDEAS), Leidschendam, the Netherlands

Indran Naidoo
Independent Evaluation Office, United Nations Development Programme, New York

Susan D. Tamondong
UN Women Global Evaluation Advisory Council, New York, and International Development Evaluation Association (IDEAS), Manila, Philippines

Empowered lives.
Resilient nations.

© 2017 International Development Evaluation Association (IDEAS)
Exeter, United Kingdom
Email: ideascoordinator@gmail.com

All rights reserved.

Any views expressed in this book are those of the authors. They do not necessarily represent the views of IDEAS, the editors, the authors' institutions, financial sponsors, or reviewers.

This book is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC-ND 4.0), https://creativecommons.org/licenses/by-nc-nd/4.0/. This license permits any noncommercial use, duplication, adaptation, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source. Please cite the work as follows:

Rob D. van den Berg, Indran Naidoo, and Susan D. Tamondong, eds. 2017. *Evaluation for Agenda 2030: Providing Evidence on Progress and Sustainability*. Exeter, UK: IDEAS.

ISBN (paper): 978-1-9999329-0-9
ISBN (electronic): 978-1-9999329-1-6

Cover design: Anne Marie Mascia
Interior design/layout: Nita Congress

Contents

CONTENTS v

FOREWORD ix
Achim Steiner

A READING GUIDE xi
Michael Quinn Patton

ACKNOWLEDGMENTS xix

INTRODUCTION 1
Rob D. van den Berg, Indran Naidoo, and Susan D. Tamondong, editors

Part I Overview and Introduction 7

1. **EVALUATING SUSTAINABLE DEVELOPMENT** 9
Vinod Thomas

2. **NO ONE LEFT BEHIND - A FOCUS ON GENDER AND SOCIAL EQUITY** 23
Marco Segone and Florencia Tateossian

3. **EVALUATION FOR IMPROVING PEOPLE'S LIVES** 35
Mallika R. Samaranayake and Asela Kalugampitiya

Part II Capacities and Capabilities. 49

4. INCORPORATING THE SUSTAINABLE DEVELOPMENT GOALS IN NATIONAL EVALUATION CAPACITY DEVELOPMENT. 51
Indran Naidoo and Ana Rosa Soares

5. PROFESSIONALIZING EVALUATION - A GOLDEN OPPORTUNITY 65
Linda Morra Imas

6. FROM EVALUATION CAPACITY BUILDING TO EVALUATION CAPACITY DEVELOPMENT - A PARADIGM SHIFT 81
Michele Tarsilla

7. MENTORING EMERGING EVALUATORS - SHARING EXPERIENCES FROM THE GLOBAL SOUTH 103
Awuor Ponge, Taiwo Peter Adesoba, Ahmed Tammam, and Tara Devi Gurung

Part III Regional Perspectives 117

8. EVALUATING SUSTAINABLE DEVELOPMENT IN SIDS - LESSONS FROM THE PACIFIC AND THE CARIBBEAN 119
Juha I. Uitto, Jeremy Kohlitz, and David Todd

9. EVALUATION STANDARDS FOR LATIN AMERICA AND THE CARIBBEAN - EXPERIMENTATION AND EVOLUTION 135
Ana Luisa Guzmán and Warren Crowther

10. THE ROLE OF MONITORING AND EVALUATION IN THE MENA REGION, WITH A FOCUS ON THE ARAB UPRISING COUNTRIES. 155
Sonia Ben Jaafar and Awny Amer

11. LEVERAGING MONITORING AND EVALUATION SYSTEMS FOR GOOD GOVERNANCE 169
Rashmi Agrawal, Asela Kalugampitiya, Jigmi Rinzin, and Kabir Hashim

12. INTEGRATING FEMINIST APPROACHES TO EVALUATION - LESSONS LEARNED FROM AN INDIAN EXPERIENCE 187
Rajib Nandi and Rituu B. Nanda

13. EVALUATION COOPERATION IN WEST AFRICA 205
Abdoulaye Gounou

Part IV Safeguards and Resettlement 217

14. ENVIRONMENTAL AND SOCIAL SAFEGUARDS IN INDIA - A CRITICAL ASSESSMENT 219
Shekhar Singh

15. THE IMPERFECT USE OF THE PAST IN RESETTLEMENT ...239
Inga-Lill Aronsson

16. EVALUATING THE BENEFITS AND COSTS OF RESETTLEMENT PROJECTS - A CASE STUDY IN THE PHILIPPINES 257
Marife M. Ballesteros

17. LIVELIHOODS IN DEVELOPMENT DISPLACEMENT - A REALITY CHECK FROM THE EVALUATION RECORD IN ASIA 273
Susanna Price

Part V Toward Sustainability of Impact 291

18. NEW FRONTIERS FOR EVALUATION IN A FAST-CHANGING WORLD 293
Chris Barnett and Rachel Eager

19. RISING TO THE CHALLENGES OF IMPACT EVALUATION - INSIGHTS FROM PILOTING A SYSTEMIC AND PARTICIPATORY APPROACH 311
Adinda Van Hemelrijck

20. EVALUATING MITIGATION PROJECTS THROUGH A THEORY OF NO CHANGE 335
Takaaki Miyaguchi

21. THE WICKED CASES OF EDUCATION AND CLIMATE CHANGE - THE PROMISE AND CHALLENGE OF THEORY-BASED IMPACT EVALUATIONS 347

Emmanuel Jimenez and Jyotsna Puri

22. MEASURING THE IMPACT OF THE EXTRACTIVE INDUSTRY'S DEVELOPMENT PROJECTS 371

Gwendolyn Wellmann

CONTRIBUTORS 389

Foreword

Since 2015, when the Sustainable Development Goals (SDGs) were adopted and the Paris Agreement was concluded, the concept of development has transformed in two major directions. First, the old distinction between "developing" and "developed" countries has disappeared in the face of global threats, particularly of climate change. All countries are now perceived as "developing" toward a sustainable, equitable, and inclusive future. Second, the scope and ambition of development in the coming decades has grown dramatically, and has now integrated the peace and conflict agenda as well as humanitarian efforts in response to natural disasters, including a focus on ecosystem services that are essential for humankind to survive. To tackle these issues, the traditional social and economic development agendas are changing beyond recognition, with innovation, new partnerships, and efforts focused on green societies and economies taking shape. Into this mix of change and ambitious perspectives, evaluation is challenged to provide evidence on what works, where, for whom, and under what circumstances in a labyrinth of new policies, program, and interventions. This book provides inspiration for an emerging new role for evaluation in the global push toward a sustainable, equitable, and inclusive world.

The Independent Evaluation Office of the United Nations Development Programme (UNDP) and the International Development Evaluation Association (IDEAS) are to be applauded for their initiative to start up the discussion on the role of evaluation in the SDGs just a month after these goals were approved. They organized two conferences that took place in parallel, with joint keynote addresses and special sessions: one from the perspective of governments, the other from the perspective of the professional development evaluator. These conferences took place in Bangkok, Thailand, in October 2015. They ended with the Bangkok Declaration on National Evaluation Capacity for the Sustainable Development Goals, which was subsequently included in the Global Evaluation Agenda 2016–2020. While conference proceedings were published in 2016, IDEAS and UNDP's Independent Evaluation Office also approached the most innovative and

forward-thinking contributors to the conference to update their insights for this book, which provides a stimulating array of subjects.

This is of the utmost importance. We need to understand the way forward on the many challenges in the SDGs. If we truly want to "leave no one behind," we need to learn from our mistakes and from efforts to build inclusive and equitable societies that can sustain themselves, not just economically, but in an endurable relationship with nature. While achieving insight is not just an effort of evaluation, it is one of the few systematic and independent assessments of interventions and innovations that we have. In its best incarnation, it includes perspectives from civil society, the private sector, government at all levels, and academia. In Agenda 2030, evaluation has a recognized role to play in understanding the progress that is made toward the SDGs. It is therefore crucial that the discussion started in Bangkok continues, and this book provides fuel to this fire.

This book provides an overview of some of the crucial transformations we see in development, and how evaluation plays an important role in these changes. First and foremost is the shift from donor-centric evaluation to country systems for monitoring and evaluation that can provide evidence for decision making in these countries. The chapters in part I make important points on how evaluation contributes to some of the most significant issues of our time, such as inclusive societies, reducing the carbon footprint of our economies, and gender and equity. Capacities and capabilities for evaluation of progress toward the SDGs are explored in part II. Part III is devoted to regional and national perspectives, underscoring that evaluation is becoming increasingly relevant to countries' development priorities in many regions of the world. Part IV deals with safeguards and resettlement, and demonstrates the negative impacts of development displacement that governments need to be aware of. The book's last part addresses some of the enduring problems of evaluation: how to evaluate impact in complex circumstances; how to evaluate new partnerships and new forms of private sector involvement, such as social and environmental impact investing.

In the meantime, the debate continues, as we see it taken up in new conferences and in deliberations of governments and societies on the potential role evaluation could play in understanding progress toward the SDGs. I hope this book contributes to informing that debate.

Achim Steiner
Administrator, United Nations Development Programme

A Reading Guide

Michael Quinn Patton

Founder and Director, Utilization-Focused Evaluation, Saint Paul, Minnesota, USA

Evaluation is a dynamic, evolving, global profession. This book reflects and illuminates that dynamism, evolution, and global engagement. To place this book in context and help you appreciate its significance, let me preface the new directions presented in these pages with some historical context.

Evaluation as a profession has been international in orientation and membership from its formal association beginnings in the mid-1970s when the Australasian Evaluation Society, the Canadian Evaluation Society, and American Evaluation Association predecessors (the Evaluation Network and the Evaluation Research Society) were all formed. The European Evaluation Society was founded in 1992.

The African Evaluation Association was formed in 1999 in Nairobi. The Latin American Evaluation Network was conceived in Lima in 2003, bringing together the networks of Brazil, Colombia, and Peru. Networks and associations in other regions followed. The International Development Evaluation Association (IDEAS) was inaugurated in 2002 in Beijing to help fill a gap in the international evaluation architecture. This was followed by international conferences, which are now biennial. The conference themes provide an overview of issues in development evaluation, issues still being addressed in the current volume. Watch for these recurring themes, and how they have evolved, as you read this book:

- New Delhi, 2005: "Evaluation for Development—Beyond Aid"
- Johannesburg, 2009: "Getting to Results: Evaluation Capacity Building and Development"
- Jordan, 2011: "Evaluation in Turbulent Times: The Crises of Food, Fuel, and Finances"
- Barbados, 2013: "Evaluation and Inequality: Moving Beyond the Discussion of Poverty"
- Bangkok, 2015: "Evaluating Sustainable Development"
- Mexico, 2017: "Evaluation of the Sustainable Development Goals: Transforming Life through Global and Regional Partnerships"

The 21st century has seen exponential international growth and development of evaluation, highlighted in 2015 as the International Year of Evaluation, as recognized by the United Nations. By 2017, more than a hundred voluntary organizations for professional evaluation had been formed representing thousands of evaluators worldwide.

TEN EVALUATION TRENDS TO LOOK FOR IN THIS VOLUME

The evolution of evaluation reflects the profession's development and adaptation to a rapidly changing world. The practice of evaluation is inextricably linked to changes in the world. At any given moment, evaluation practice worldwide will include traditional ways of doing evaluation that have become established over the last 50 years as well as innovative new directions that are on the leading edge of both development and evaluation. This volume reflects that mix. With that in mind, I offer my top 10 list of things to look for in this book. To the extent that you can identify these developments and challenges, and make the distinctions between traditional evaluation approaches and innovative new directions, you understand the history, dynamism, and future of our profession. As is the case with top 10 lists, I offer a countdown, concluding with the most important challenge—at least as I see it. I would add that my list is necessarily subjective and based upon my own observations and biases, so by offering this list, I invite readers to think about your own observations of evaluation trends and challenges, and look for how the authors in this book acknowledge and engage with those trends and challenges.

10. New evaluands. "Evaluand" refers to the focus of an evaluation, the thing evaluated. Traditional evaluands are projects and programs, which we do an excellent job of evaluating.

We know how to specify SMART goals (specific, measurable, achievable, relevant, and time-bound) and develop performance indicators. We have become skilled at developing logic models and theories of change. We know why and how to distinguish monitoring from evaluation, the different types of evaluation (utilization focused, impact, theory driven, cost-benefit, empowerment, participatory, social justice, etc.), the diverse uses of evaluation (accountability, learning, decision making, enlightenment, etc.), and the importance of working with diverse stakeholders (program staff, policy makers, funders, participants, etc.). We have standards for what constitutes evaluation quality and checklists for what should be included in an evaluation. We know the importance of specifying intended use by intended users. We have a variety of ways of reporting findings. This is by no means a comprehensive or exhaustive list, but, hopefully, it provides a sense that we've learned a lot, know how to do a lot, and merit the designation of being a knowledge-based profession.

But new evaluands beckon. The emergent challenges for evaluation, from my perspective, primarily have to do with new units of analysis and broader areas of focus for evaluation.

Evaluation, we say, "grew up in the projects." As evidenced by what we do well, the profession's origins lie in evaluating projects, and, from my

perspective, we remain in the grip of a self-limiting project mentality. Evaluating community impacts, regional and sectorwide initiatives, cross-sector initiatives, networks and collaborations, global leadership development, innovation, and collective impact pose new conceptual and methodological challenges. In addition, and along parallel tracks, evaluators are being challenged to develop new approaches to evaluating the scaling of innovations, assessing the effects of social media, and using "big data" to examine large and open systems. Ecosystem governance is a leading-edge evaluand and cannot be evaluated as if it is a project or program.

Evaluating principles, such as the 2005 Paris Declaration on International Development Aid, is also different from evaluating projects. Principles-driven programs are different from goals-driven programs. Principles constitute a different kind of evaluand. Principles take on added importance among the new challenges for evaluation because principles are the primary way of navigating complex dynamic systems and engaging in strategic initiatives. Principles undergird efforts at community change and collective impact. Understanding how to evaluate principles, and adapting evaluation concepts, approaches, methods, and processes through principles-focused evaluation will, I believe, provide valuable direction for how to evaluate other new evaluands as we grapple with related emergent challenges (Patton 2018).

So, as you read, watch for new evaluands, innovative units of analysis, and the limitations of forcing complex initiatives into project boxes. With this 10th trend in my countdown list as context, I can be briefer in identifying the remaining things to watch for as you read.

9. Applying complexity understandings. This trend follows from the preceding item but deserves highlighting because of its importance. Projects are closed systems, or at least treated as such in most evaluations, in which boundaries can be established and control can presumably be exercised within those boundaries by both program staff and evaluators. In contrast, complex dynamic interventions, advocacy campaigns, and strategic initiatives are open systems characterized by volatility, uncertainty, and unpredictability—all of which make control problematic. Treating these complicated and complex evaluations like simple projects is inappropriate, ineffective, and insufficient. Indeed, it can do harm by misunderstanding, misconceptualizing, and misrepresenting the very nature of complex change and thereby generating results that are inaccurate and irrelevant. Consider these books on complexity as context for the contributions in this volume. Watch for how complexity is addressed in these pages:

- *Evaluating the Complex: Attribution, Contribution and Beyond*, Kim Forss, Mira Marra, and Robert Schwartz eds. (2011)
- *Evaluation and Turbulent Times: Reflections on a Discipline in Disarray,* Jan-Eric Furubo, Ray C. Rist, Sandra Speer, eds. (2013)
- *Developmental Evaluation: Using Complexity Concepts to Enhance Innovation and Use*, Michael Quinn Patton (2011)
- *Evaluation in the Face of Uncertainty*, Jon Morell (2012)

- *Dealing with Complexity in Development Evaluation: A Practical Approach*, J. Michael Bamberger, Jozef Leonardus Vaessen, Estelle R. Raimondo, eds. (2016)
- *Evaluation Research Methods: Managing the Complexities of Judgment in the Field*, Saville Kushner (2017)

8. Globalization, global interconnectedness, and global systems dynamics. Evaluating global systems dynamics poses a particularly daunting challenge as we learn to view the Earth and the Earth's inhabitants as a holistic, interconnected, and interdependent global system.

Why so much attention to globalization? Consider this recent *New York Times* Business Day report:

> The Fed [the U.S. Federal Reserve] Acts. Workers in Mexico and Merchants in Malaysia Suffer. Rising interest rates in the United States are driving money out of many developing countries, straining governments and pinching consumers around the globe. (Goodman, Bradsher, and Gough 2017)

The agreement on the 17 Sustainable Development Goals (SDGs) and 169 targets has been heralded as a major step forward in global governance toward a sustainable post-2015 development agenda. Certainly, the SDG framework has addressed many of the gaps identified with the Millennium Development Goals (MDGs) by surfacing barriers to sustainability on a systemic level including inequality, consumption patterns, and weak institutional capacity. Nevertheless, the framework remains more sectorally siloed than integrated. Implementation is focused at the nation-state level with few mechanisms to inspire and mobilize innovation within the private sector and civil society to implement alongside national governments.

Poverty, hunger, well-being, education, and ecosystem health are interrelated. The relationships are nonlinear, dynamic, and complex. As you read this book, watch for discussions about, analyses of, and approaches to evaluation of the SDGs that integrate indicators across SDGs and move both down and up scale (down to local levels and up to encompass global dynamics), thereby moving beyond nation-states as the only unit of analysis. By making sense of the interrelationships and interdependencies among and across SDGs, interventions and evaluations will demonstrate sensitivity to and understanding of global systems dynamics.

7. Power, politics, and the realities of evidence. The book's subtitle is "Providing Evidence on Progress and Sustainability." Watch for how the contributors to this book acknowledge and deal with the politicization of evidence.

On April 22, 2017, millions marched for science in 600 cities worldwide. The American Evaluation Association was one of 270 partner organizations that supported the March for Science. The *New York Times* headline on the day of the march read: "Scientists, Feeling Under Siege, March against Trump

Policies" (St. Fleur, 2017). *USA Today* led with: "Marchers for Science Protest 'Alarming' Anti-Science Trends." Anti-science is, ultimately, anti-evaluation evidence. Culturally and politically, the anti-science trends include "alternative facts," "fake news," and a "post-truth" world. In November 2016, the Oxford Dictionaries announced "post-truth" as its Word of the Year:

> **post-truth** *adjective* Relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief. (Oxford Dictionaries 2016)

Casper Grathwohl, president of the Oxford Dictionaries, explained: "Given that usage of the term hasn't shown any signs of slowing down, I wouldn't be surprised if *post-truth* becomes one of the defining words of our time" (Oxford Dictionaries 2016). Science? Just another perspective. Evaluation? Paperwork. Administrivia. Opinion.

The current anti-science political climate calls us to unite with others engaged in defending and supporting science, creating a united front to the larger world. Science is ultimately about evidence, so how evidence is conceptualized and what constitutes evidence matter a great deal for evaluation. Ironically, some evidence skeptics have become sophisticated at undermining the credibility and utility of evaluation by demanding levels of "proof" that are not possible in complex dynamic systems. Evaluators are having to deal with preponderance of evidence, triangulation, use of mixed methods, and rapid feedback as ways of facing the challenges of producing meaningful and useful results in a timely fashion.

The politics of evidence include speaking truth to power, speaking truth to each other, and empowering those at risk of being left behind to speak their truth. A "post-truth" political world undermines the value of systematic evaluation. In short, evidence is not just about data. It's about how people understand what constitutes meaningful and credible evidence. So, watch how the contributors to this volume take on these issues. The future viability of evaluation as a valued evidence-based profession is at stake.

6. Evaluative thinking embedded in evaluation processes. Methods alone do not ensure rigor. A research design by itself does not ensure rigor. High-quality analytical techniques and procedures do not ensure rigor. Rigor resides in, depends on, and is manifest in *rigorous thinking*—about everything, including methods and analysis. This means valuing intellectual rigor. There are no simple formulas or clear-cut rules about how to do a credible, high-quality analysis. The task is to do one's best to make sense of things. An evaluator returns to the data over and over again to see if the constructs, categories, interpretations, and explanations make sense—if they sufficiently reflect the nature of the phenomena studied. Creativity, intellectual rigor, perseverance, insight—these are the intangibles that go beyond the routine application of scientific and research procedures. These are bedrock elements of rigorous evaluative thinking. Rigorous evaluative thinking combines critical thinking, creative thinking, inferential thinking, and practical thinking. Watch for how evaluative thinking is manifest in the contributions of this book.

5. Evaluation as intervention (process use). The mantra of performance management is that *what gets measured gets done*. Process use concerns how the conduct of an evaluation affects what is done quite apart from the findings of the evaluation. Evaluation is no longer simply about producing an end-of-project report. How evaluation is framed, the questions asked, the data collection priorities established, and the processes for engaging major stakeholders constitute interventions in the development process. As you read, watch for how the presence of evaluation affects development interventions.

4. Evaluation understandings shaping intervention designs. One way evaluation becomes an intervention is by helping conceptualize an intervention's theory of change or strategic approach. Evaluation is no longer simply a back-end activity assessing whether goals are attained. Evaluators are now involved in determining how an intervention is conceptualized—for example, how much attention is given to interactions across and among different SDGs. That's an intervention design issue as much as it is an evaluation issue. Watch for how the examples in this book include evaluators playing a significant role in framing development interventions. A good example is chapter 11 on good governance.

3. Failure as learning. A major source of resistance to evaluation is fear of failure. The preceding items in this list all point toward the increased importance of learning from evaluation findings, and a particularly potent form of learning follows from acknowledging and understanding failures. Engineers without Borders has established a stellar approach to learning from failure by issuing an annual failure report. The politics of development make acknowledging failure, and learning from it, particularly challenging, but also essential. Watch for how the authors in this book identify and address failure, and support learning and adaptation.

2. Transformation. Evaluation has traditionally focused on outcomes and impacts. That is no longer sufficient. Climate change changes everything. The urgency of dealing with the implications of climate change has led to a new focus on transformation. Transformation involves multiple and intersecting interventions that lead to major, deep, systemic, and resilient changes at a large scale, across SDGs, and with urgent timelines. Transformation means big changes happening quickly. Time is of the essence. Scenarios supporting the need for transformation include forecasts that by 2050, under current trends, 20 countries will be gone, 60 cities swamped and unviable, and 1.5 billion people displaced.

Evaluating transformation is new territory for evaluators, a new and challenging evaluand. It cannot be reduced to targeted indicators. I would argue that transformation is a sensitizing concept that's only meaningful when applied to a given context. Transformation has to be interpreted contextually and dynamically. Thus, transformation should not be subject to narrow measurement or narrow operationalization because it occurs in nonlinear and often unpredictable ways. The problem is not the measurement of

transformation; the problem is actually engaging with multiple perspectives, multiple kinds of data—qualitative and quantitative, case studies, indicators—and global to local scales in an integrated, systemic way to understand what the global patterns of transformation are. Watch for how the emergent agenda of transformation is addressed in this book. More generally, watch for a sense of urgency and scale that goes well beyond project thinking to global transformation.

1. Focus on intended beneficiaries, especially children. I recently reviewed the five-year strategic plan for evaluation of a major international agency. The plan went into great detail about how evaluations would be conducted, criteria of quality, the nature of reports, the timing of findings, and a commitment to meaningful accountability. Intended beneficiaries of development efforts were essentially invisible. The evaluation appeared people-less and heart-less.

A 2015 UNICEF report entitled *Unless We Act Now* makes the case that children will bear the brunt of climate change. The report documents that over half a billion children live in extremely high flood-occurrence zones; nearly 160 million live in high or extremely high drought-severity zones (UNICEF 2015). Since there is a clear scientific consensus that climate change will increase the frequency of droughts, floods, and severe weather events, children are especially in jeopardy globally. Climate change will not affect all equally. Because of the potentially devastating risks in flood and drought zones, and high poverty and low access to essential services such as water and sanitation in those zones, children and families who are already disadvantaged by poverty are likely to experience the greatest effects of climate change.

Part of the responsibility of evaluation should be to highlight and deepen global understanding of effects on real people. If no one is to be left behind, how evaluations are conducted affects that vision. Watch for how this book illuminates the effects of development efforts on intended beneficiaries in ways that make those people and their lives real to readers.

LOOKING FORWARD

I approached reading this book through the lens of how it manifests major changes in the world and corresponding changes in evaluation practice and theory. I've shared the 10 major developments in the evaluation profession that I think will determine the future relevance and utility of our profession. As I noted earlier, your list of what to watch for will not be the same as mine. So, what will you look for? The diverse authors of this volume offer their own priorities and perspectives. In so doing, they challenge all of us to think about how we understand and will engage with *Evaluation for Agenda 2030.*

REFERENCES

Goodman, Peter S., Keith Bradsher, and Neil Gough. 2017. "Fed Acts. Workers in Mexico and Merchants in Malaysia Suffer." *New York Times* March 16.

Oxford Dictionaries. 2016. "Oxford Dictionaries Word of the Year 2016 is..." Press release. Oxford University Press, Oxford, UK.

Patton, Michael Quinn. 2018. *Principles-Focused Evaluation: The Guide*. New York: Guilford.
St. Fleur, Nicholas. 2017. "Scientists, Feeling under Siege, March against Trump Policies." *New York Times* April 23.
UNICEF (United Nations Children's Fund). 2015. *Unless We Act Now: The Impact of Climate Change on Children*. New York: UNICEF.

Acknowledgments

We would like to express our deep gratitude to the many people involved in bringing this publication to fruition, and especially the authors who have contributed to this book. We are thankful to the individuals and organizations who contributed valuable perspectives, ideas, time, and support throughout the course of this book's development.

We would specifically like to thank Lee Cando-Noordhuizen, who acted as the assistant editor coordinating with everyone, keeping track of all details, and ensuring that the book project was proceeding according to schedule. We value her professionalism and meticulousness.

We appreciate the high-quality layout and design by Nita Congress and, for the cover, Anne Marie Mascia. Janet Hulstrand, Rebecca Gholson, John Seidenberg, and Julia Lee Barclay-Morton copyedited the manuscripts with a fine-tooth comb.

This book has been produced with financial assistance from the U.K. Department for International Development and the Independent Evaluation Office of the United Nations Development Programme. Maureen Woodhouse and Niels Dabelstein provided important financial/administrative support throughout the process.

Note that all dollars cited here in are U.S. dollars.

The Editors

Introduction

Rob D. van den Berg, Indran Naidoo, and
Susan D. Tamondong, editors

Sustainable development is back center stage on the international agenda. After the Earth Summit in Rio de Janeiro, Brazil, in 1992, sustainable development seemed in vogue for a while but lost ground to a more pragmatic perspective in the Millennium Development Goals in 2000, which embodied practical goals and targets that could be met in 15 years' time by the international community and by countries. However, the increasing urgency of climate change and related environmental crises such as biodiversity loss and the growth of chemical and other waste throughout the world caused a recalibration of development processes. This led to the Sustainable Development Goals (SDGs), incorporated into Agenda 2030, which were accepted by the United Nations General Assembly in September 2015. In October 2015, two evaluation conferences took place in parallel in Bangkok, Thailand: one organized by the Independent Evaluation Office of the United Nations Development Programme (UNDP) and the other by the International Development Evaluation Association (IDEAS). These two conferences focused on sustainable development: the UNDP conference on what the newly adopted SDGs meant in terms of development of national capacities, mainly for governments, whether they had the capacity to contribute to understanding progress toward the SDGs, and whether the policies were in place to enable evaluation to play its proper role. The IDEAS conference, on the other hand, focused on the concept of sustainable development and how it could be evaluated, and aimed at bringing best practices and innovation from all over the world to be discussed in Bangkok.

The UNDP conference—the Fourth International Conference on National Evaluation Capacities (NEC)—led to the publication of its proceedings in June 2016. IDEAS does not publish proceedings, but has, since its Global Assembly in Johannesburg, South Africa, in 2009, presented the most challenging and promising perspectives emerging from its conferences in a book. The three books that followed were edited by Ray C. Rist, Marie-Helene Boily, and Frederic Martin. After Ray Rist retired as president of IDEAS, he presented the continuation of the series as a challenge for the new president:

Rob D. van den Berg. A new editorial committee was established by Rob, with Indran Naidoo, former board member of IDEAS and currently director of the Independent Evaluation Office of the UNDP, and Susan D. Tamondong, vice president of IDEAS. The aim was to publish a follow-up book of the Bangkok conference, focusing on some of the themes and on new and promising developments in the field of evaluation of sustainable development.

The present volume thus should be placed in the tradition of the three IDEAS's books on the Global Assemblies in Johannesburg (2009), Amman (2011), and Barbados (2013). It diverges from that tradition, as the book includes many perspectives that were explored with the NEC conference in joint sessions, including perspectives of governments, nongovernmental organizations, the private sector, and academia. Issues discussed included policies and capacities, as well as evaluation methodology and the difficulties for the evaluation profession to find its place between academia and (international) bureaucracies.

Part I of the book comprises further developments of the keynote addresses at the two Bangkok conferences. Vinod Thomas (chapter 1) provides an overview of the main challenges with which evaluators are confronted when evaluating sustainable development. His chapter discusses the various evaluation methods for assessing sustainability: cost-benefit analysis, impact evaluation, green accounting methods, social impact analysis, and safeguard compliance mechanisms. The chapter argues for rigorous frameworks for evaluation, but at the same time underscores the need for innovation and further development of methods. Vinod urges capacity development in countries that need to apply these methods in their evaluations of country-led initiatives.

In chapter 2, Marco Segone and Florencia Tateossian have developed Marco's keynote address from a United Nations' perspective, advocating for equity-focused and gender-responsive evaluations at the national level, enabled by sufficient national evaluation capacity and a monitoring and evaluation (M&E) system that includes all national and international partners. Mallika Samaranayake and Asela Kalugampitiya further develop the former's Bangkok keynote address on the importance of participatory evaluation in chapter 3. In a world where equity concerns are widespread and increasingly seen as behind the rise of populist movements, participatory evaluation can provide evidence to policy makers on what could be done to ensure that "no one (is) left behind," to quote one of the overarching themes of the SDGs.

Part II focuses on capacities and capabilities. Indran Naidoo and Ana Rosa Soares discuss UNDP efforts to support countries in their development of evaluation capacity and national systems for evaluation (chapter 4). They also incorporate lessons from implementation of the Millennium Development Goals to inform how the SDGs should be evaluated, as well as the capacity that countries need to develop and enable their M&E systems. In chapter 5, Linda Morra Imas focuses on the professionalization efforts that were discussed at the Bangkok conference and what has happened since. While much has been accomplished, Linda calls for new efforts to ensure core competencies reflect the SDGs and inform the development of national capacities.

The paradigm shift from capacity building to capacity development is discussed by Michele Tarsilla in chapter 6. After a discussion of contextual and process-related factors that need to inform capacity development, Michele highlights the contribution of IDEAS to the ongoing discourse on evaluation capacity. Lastly, in chapter 7, a group of authors around Awuor Ponge discuss their experiences as "young and emerging evaluators" and the specific barriers and obstacles they face to find their place in the global and national evaluation communities. To tackle these barriers, Ponge, Adesobo Taiwo Peter, Ahmed Tamman, and Tara Devi Gurung advocate for mentoring programs that support young and emerging evaluators.

Regional perspectives are brought to the reader in Part III. Juha I. Uitto, Jeremy Kohlitz, and David Todd highlight the challenges that the small island developing states (SIDS) in the Caribbean and Pacific face to develop national capacities for evaluation of sustainable development (chapter 8). These range from limited human and institutional capacities to low priorities for evaluation in government policies. They argue that these challenges are best addressed by crafting M&E systems that are appropriate for a variety of SIDS contexts, that are country led, and that are supported by external agencies in a coherent manner. In chapter 9, Ana Luisa Guzmán and Warren Crowther tackle the recent development of evaluation standards in Latin America, as proposed by the regional network Latin American and Caribbean Network of Monitoring, Evaluation and Systematization (ReLAC) on the basis of a series of innovative evaluations undertaken in Costa Rica. Their focus is on applying ethics in evaluation and on basic principles such as relevance, viability and trade-offs between evaluation and decision making.

Sonia Ben Jafaar and Awny Amer in chapter 10 discuss regional developments in the Middle East from the perspective of the call for greater accountability of governments to their peoples. The Arab Uprising, also known as the Arab Spring, has initiated a new paradigm on the role of evaluation that promotes national capacities and national M&E systems in which local professionals are valued and supported and contribute through evaluations to the development of their countries and the region. Chapter 11 provides a similar perspective on three South Asian countries: the call for evaluation to contribute to good governance. Rashmi Agrawal, Asela Kalugampitiya, Jigmi Rinzin, and Kabir Hashim reflect on recent initiatives and efforts in Bhutan, India, and Sri Lanka. In India and Sri Lanka, these efforts focus on tackling corruption in public service delivery; in Bhutan, the role of evaluation in promoting good governance and increasing Gross National Happiness is discussed.

Chapter 12 is devoted to feminist approaches and evaluation in India, written by Rituu B. Nanda and Rajib Nandi. Based on a program implemented by the Institute of Social Studies Trust in India, the chapter aims to contribute to a better understanding of how evaluations can support changes in gender-based inequalities and power dynamics. The last chapter of this part, chapter 13, deals with evaluation cooperation in West Africa. Abdoulaye Gounou discusses the capacity and impact evaluation program in West Africa as supported by the International Initiative for Impact Evaluation (3ie), focusing on the countries of the West African Economic Monetary Union. Of great interest is Abdoulaye's discussion of the new program Twende Mbele, which

partners Benin, South Africa, and Uganda in developing appropriate M&E and exchange of experiences.

Part IV aims to discuss the role of evaluation in preventing negative impacts. It starts with a historical overview of environmental and social safeguards in India by Shekhar Singh and highlights that recent political developments have led to reduced priority for these safeguards (chapter 14). Evaluations were essential to bring attention to the detrimental effects of ignoring these safeguards, and Shekhar's chapter develops proposals on how evaluation can continue to play this role in the future. For the next chapters in this part, we turn to resettlement issues—often accompanied by negative impacts of development displacement, which have led to the adoption of social safeguards on resettlement. In chapter 15, Inga-Lill Aronsson provides an anthropological and historical perspective on resettlement, and indicates that heritage and memories tend to be neglected in projects, to the detriment of outcomes. Given this lack of a historical perspective, current resettlement models are insufficient to grasp the longitudinal consequences of resettlement. A consideration of heritage and memory could lead to improvements.

Marife Ballesteros, in chapter 16, focuses on the assessment of benefits and costs of resettlement projects implemented by the Philippine government using a quantitative methodology. The study compares resettlement modes, in-city and off-city, and determines which mode provides greater efficiency and best socioeconomic outcomes. The author discusses areas for improvement that the government can undertake, bearing in mind trade-offs and recommendations for a more efficient resettlement resulting in improved welfare. The last chapter in this part, chapter 17 by Susanna Price, compares international policy perspectives and evaluation outcomes in Asian countries to see how livelihoods are affected, addressed, and evaluated. She argues that livelihood issues are often neglected in laws and regulations concerning resettlement. Furthermore, livelihood risks, livelihood support, and livelihood outcomes in evaluations are rarely seen. The author presents some approaches that may provide a way forward in building the knowledge base on livelihood success and sustainability through evaluation at the country level.

Part V concerns evaluation of impact in its broadest sense and focuses on sustainable development issues. In chapter 18, Chris Barnett and Rachel Eager further develop a contribution of Chris's to the Bangkok conferences in a special session on the new frontiers for evaluation. While new initiatives take shape to achieve sustainable development—and especially new partners from the private sector engage in forms of "impact investing" and social corporate responsibility—the challenge is how evaluation can provide evaluative evidence within these often complex, interconnected, and rapidly changing contexts. They argue for a bolder evaluation agenda, in which evaluators recognize their potential role in contributing to change: to act not just as providers of evidence for others to use, but to proactively engage in an ethical obligation to society, and to stimulate 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, as well as to call others to account.

In chapter 19, Adinda van Hemelrijck reflects on the Participatory Impact Assessment and Learning Approach (PIALA), a systemic theory-based and participatory mixed method for addressing the challenges of impact evaluation in complex development contexts. On the basis of fieldwork in Vietnam and Ghana, Adinda concludes 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 reduce costs over time, while enhancing the value of impact evaluation and the uptake of its findings.

Takaaki Miyaguchi in chapter 20 synthesizes findings from eight different climate change mitigation projects in five different Southeast Asian countries, using the Theory of No Change approach developed by Christine Wörlen. 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. Emmanuel Jimenez and Jo Puri also synthesize findings from various sources to identify gaps in evidence on education and climate change/environment interventions (chapter 21). They discuss the "wicked problems" that evaluators encounter when aiming to bridge the gaps; their analysis points out why the gaps persist and how future evaluations might address them.

Lastly, Gwendolyn Wellman's chapter reports on impact evaluations of the development efforts of a mining company in Ghana (chapter 22). While the company was not primarily interested in measuring the impact of its community/societal development program, the government of Ghana required it to evaluate what it had done. Gwendolyn reports on the process that evolved, and discusses to what extent the evaluation commissioned was able to come up with findings. The chapter concludes with an exploration of "the way forward" for impact evaluations of the development activities of big corporations.

***

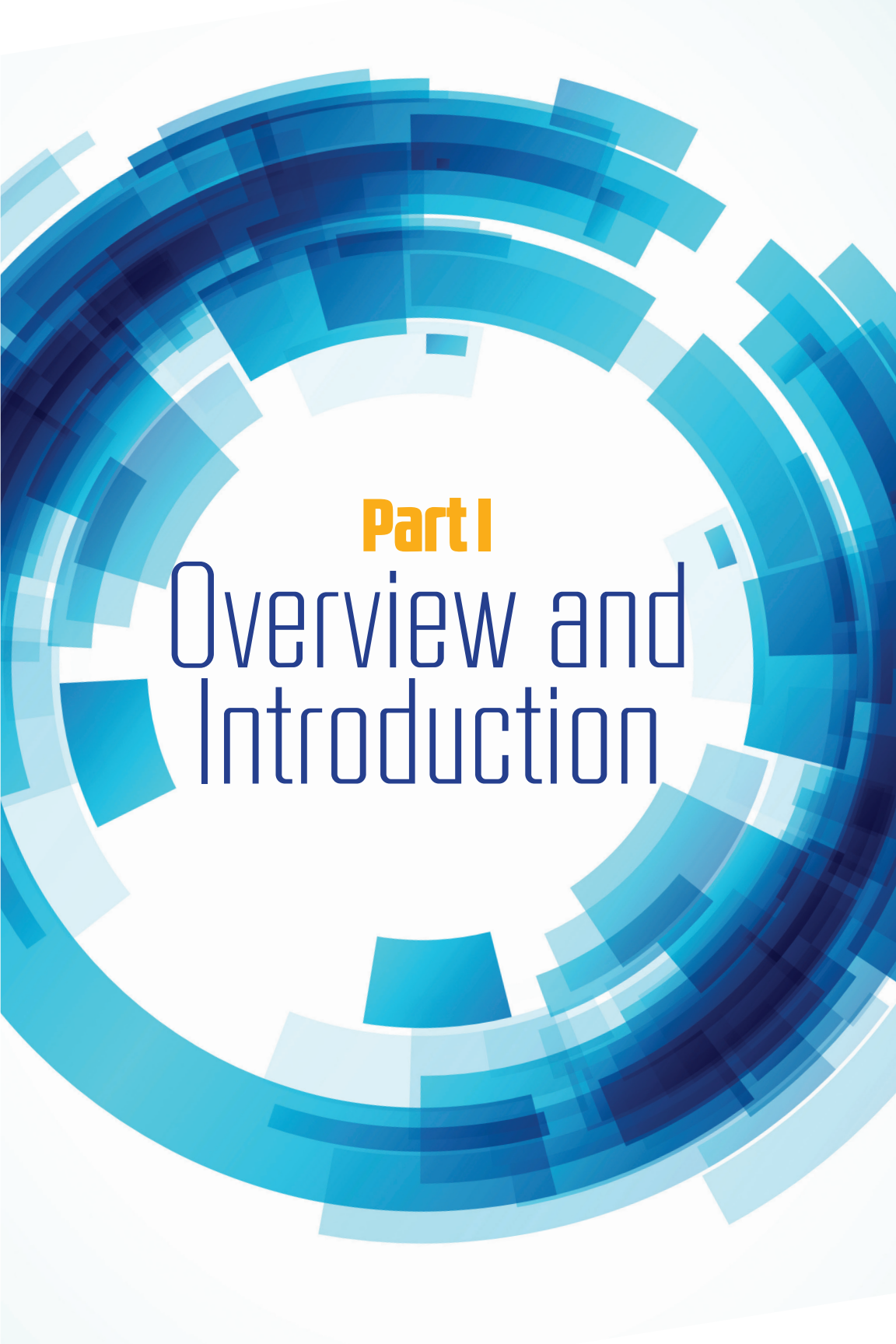
Given the variety and depth of topics in this book, we hope that it provides an overview of some of the important issues in the global evaluation community:

- how to take sustainability into account;
- how to leave "no one behind" when evaluating sustainability;
- what capacities and capabilities are needed to undertake these evaluations;
- how this is taking shape in regions and countries and incorporated into country-led evaluation systems;
- whether the negative consequences and impacts of interventions are sufficiently taken into account and what the role of evaluation can be in highlighting these issues; and

- whether evaluation of impact is sufficiently developed to tackle the growing demand for evidence, including new initiatives and the increasing involvement of the private sector and civil society as partners in development.

Increasingly, evaluation is becoming a truly global issue and concern. While in the past most of the chapters in a book like this would have been written by gray-haired old men from the developed world, this book contains the work of 38 authors, 26 of whom are from the Global South!

Reading this book will certainly lead us to acknowledge that much still needs to be done. But it is better to stand at the beginning of a new road, knowing that this is where we—as an evaluation community—need to go, than to sit down in recognition of where we have failed so far. If one thing may be concluded from this book, it is that the evaluation community is intellectually alive and kicking and aiming to improve further in the coming decades.


Part I
Overview and Introduction

Chapter 1

Evaluating Sustainable Development

Vinod Thomas

Abstract. *The United Nations' new Sustainable Development Goals provide the international mandate and opportunity for countries to focus on socially equitable and environmentally sustainable growth. There is a growing recognition in countries that the quality of growth signified by inclusion and sustainability is vital for how it affects the well-being of people and the planet, and for continuing economic growth itself. But this broader recognition also raises several tough challenges. An important one is managing actual or perceived trade-offs that occur as countries pursue sustainable and inclusive development. One example is food security, for which there is the need to increase areas under cultivation while at the same time to ensure sustainable forest use and conservation. The pressure to develop fossil fuel energy to power growth is another case in point which conflicts with controlling pollution and minimizing damages to human health and climate change. The pursuit of sustainable and inclusive growth also presents challenges to evaluation. It would be fair to say that evaluative priorities and methods have not kept pace with the needs of assessing outcomes in sustainability. Stepped-up evaluative efforts are necessary at several levels, ranging from sound frameworks and methods of analysis to relevant and practical applications, conclusions, and recommendations.*

Vinod Thomas, Visiting Professor, National University of Singapore, vndthomas49@gmail.com.

FRAMEWORK OF SUSTAINABLE DEVELOPMENT

Three dimensions of sustainable development are economic growth, social inclusion, and environmental sustainability,[1] which align with an ecosystems services approach ensuring that environmental services such as clean air and water and nutritious food remain available for future generations. Past initiatives were largely skewed toward attaining high economic growth. Focusing solely on the pace of growth has contributed to increasing inequality, environmental destruction, and climate change—repercussions that threaten economic growth itself. The challenge for evaluation is integrating the social and environmental dimensions while assessing growth.

The basic framework is one that recognizes that for economic growth to be sustainable, we need to value all three forms of capital—physical, human, and natural (Basu et al. 2017; Thomas et al. 2000). Government spending and private investment have long favored the first two forms of capital, with natural resource management getting short shrift. Yet, a country's natural capital—its stock of natural assets—is essential for the pace and quality of growth. Sustainable land use and agricultural practices, and forest and coastal management, need far greater emphasis.

The fact is that raising economic growth remains the principal driver of policy. Earlier thinking was that social inclusion and environmental sustainability are good to have, but that their pursuit presents unacceptable trade-offs to economic growth. Evaluation results, however, have shown that projects with objectives incorporating inclusive growth and the environment have performed well compared to those that have stand-alone objectives (IED 2014, 2015). These results provide support for building inclusion into the design and implementation of projects intended to help raise economic growth. In this and other instances, evaluators need to put more of such contextual evidence into their evaluations.

The argument in fact goes further. In many settings, growth itself seems to depend on inclusion. The intuition is that when all the people are included in the growth process, the possibilities for growth are that much greater. If so, going forward, not just any growth will do; it needs to be growth that is more inclusive. In this case, there would be a premium on generating growth that disproportionately includes the lower income strata in the growth process.

This line of thinking is just as powerful in the case of environmental sustainability. There is growing evidence that sustained growth will not be possible in the future without tackling environmental degradation and climate change. For example, the costs of climate-related disasters in many disaster-prone countries such as Bangladesh, Cuba, Haiti, the Philippines, and Thailand are staggering, and they weigh on economic growth.

In principle, making growth more inclusive and sustainable is assuring development that meets the needs of the present without compromising the ability of future generations to meet their needs. Providing more and better

[1] http://www.un.org/sustainabledevelopment/sustainable-development-goals/.

quality of growth for sustainability requires that high growth be accompanied by quality standards that ensure broad-based benefits through gainful and productive employment and access to opportunities in health, education, and social protection while at the same time ensuring environmental quality that supports future growth.

Nevertheless, sustainable development remains a contentious, complex, and dynamic paradigm. There are many considerations to take into account among which are the many trade-offs and policy issues at stake.

METHODS FOR ANALYSIS

On methods for assessing sustainability, there are five strands to stress. First, cost-benefit analysis, a long-standing economic tool of analysis, can be put to better and wider use to assess sustainability. In particular, the frameworks allow the qualification in the use of market prices to account for externalities such as pollution and congestion. The effect of doing so can be enormous, as in the carbon emissions that aggravate global warming. If, based on such analysis, carbon emissions were to be taxed adequately—in contrast to the absence of such taxes, let alone subsidies that encourage the use of polluting fuels such as coal—the result would change the game.

Furthermore, cost-benefit analysis has not been used often for such purposes. Where it has been used, the results were compelling. For example, environmental impact assessment is mostly based on cost-benefit analysis. It has focused mostly on "do no harm" principles and has been instrumental in preventing investments and projects that would have led to environmental damage.

The use of cost-benefit analysis in some of the multilateral development banks (MDBs) has been on the decline (IEG 2010a). Part of the reason is the greater difficulty in applying the technique in social sectors where lending has been on the rise. But that is not a good enough reason to de-emphasize cost-benefit analysis. The tool is potentially a highly effective means to assess the net gains and losses from interventions.

Second, impact evaluation can help to assess the effects of programs that seek to ensure greater social inclusion and environmental sustainability. The much-cited example is the case of social protection programs, in particular measuring the impacts of conditional cash transfer programs. There are also good examples of the effects of forest protection and natural resource management more generally. While there are many counterfactual evaluations that have contributed such insights, many of the useful efforts have not been experimental.

We also see a strong emergence of new forms of evaluation of impact, such as process tracing, systems mapping, and qualitative comparative analysis. They use methods and tools that are rigorous, but not necessarily experimental. Randomized control trials cannot tackle issues like climate change and sustainability over generations.

Third, green accounting methods in principle are available for better valuation of natural capital (Hamilton 2014). Data are usually a constraint in effectively applying such valuation, but it is clear that when the destruction of

natural capital is not accounted for, it results in inflating long-term growth prospects (Dasgupta 2009). Not accounting for the destruction of natural capital sends the wrong signals for pursuing gross domestic product growth at the expense of de-investing in natural capital, which eventually hurts the growth process itself.

Fourth, social impact analysis brings in especially the harnessing power of participatory process in development planning and implementation and is especially relevant where environmental impact is also taken into account. Even if qualitative in many instances, this work shows a direction that should be encouraged, given its relevance for sustainable development. Social and environmental impact assessment includes the processes of analyzing, monitoring, and managing the intended and unintended social and environmental consequences, both positive and negative, of planned interventions (policies, programs, plans, and projects) and any social and environmental change processes invoked by those interventions. Its primary purpose is to bring about a more sustainable and equitable biophysical and human environment.

Fifth, safeguard compliance mechanisms are essential to ensure that development projects do not cause social and environmental harm. Especially where negative externalities are present, desirable regulations designed to avoid harm would not be followed by private or even public agents without environmental and social safeguards. But safeguards have not historically been considered as part of sustainability criteria of projects, and this must change. A subsequent section discusses safeguards in greater detail.

Rather than thinking of these tools as alternatives, one can have a rigorous framework that mixes methods depending on the issues at hand. It is crucial to list the things that are really important. Incorporation of analysis of counterfactuals can be applied more widely than at present, not only in social areas but also in the agriculture, urban, and infrastructure sectors. These are not necessarily experimental, but can integrate with others' regression analysis, counterfactual systems analysis, and cost-benefit analysis.

We must take cost-benefit analysis more seriously and not hide under the premise of unquantifiable aspects or questions about the reliability of economic, social, and environmental modeling, and a suspicion that sometimes there were biases in the assumptions that were incorporated in these models. It is important to always evaluate both benefits and costs so that the evaluation can objectively guide sustainable development.

Underlying evaluation of impact and cost-benefit analysis is green accounting, which can help enormously in the right valuation, especially of natural capital. Data availability and estimation methods remain a barrier, but they must be overcome with continued efforts and financing for such work, which has high payoffs.

Some applications of these techniques have yielded good results (Thomas and Luo 2011). For example, there have been some striking results in assessing the impact of conditional cash transfer programs in the Philippines. There has been a breakthrough in evaluating sustainable use of natural resources, as for example in Brazil's deforestation and biodiversity loss. Remote sensing in monitoring and evaluation in the management of peatlands in Mongolia has been an interesting experience.

In the urban space, there have been evaluations of the value added when bus rapid transit incorporates a clean development mechanism such as clean energy, as has been done in China. Encouraging commuters to use bus transport in Brazil and Korea is another case in point. The full cost of subsidies for fossil fuels and the value of slashing them have also been assessed quite carefully (IMF 2015; Morgan 2007). More work needs to be done in assessing the full benefits of switching to renewable energy. These issues can be best illustrated with specific areas in mind, a few of which are taken up below.

NATURAL RESOURCE MANAGEMENT

Environmental sustainability is heavily dependent on how natural resources are harnessed and utilized. There are a number of perceived and actual trade-offs when it comes to natural resource management, especially as they involve externalities and the public goods nature of some of these resources.

In some cases there are win-win opportunities, where growth and environmental objectives complement each other. Improving energy efficiency and reducing energy losses are a case in point (Petrie and Thomas 2013). There are other areas where costs are clearly involved, and yet there may be net-win opportunities as these costs are more than offset by societal and/or environmental gains. Switching out of subsidies for fossil fuels (where negative externalities are present) and even providing subsidies for renewables (where positive externalities are present) would be a case in point.

One central issue is the question of the carbon content of economic growth. A byproduct of economic growth is the production of greenhouse gases and the resulting climate change. A general observation is that a 1 percent increase in per capita income induces—on average and with exceptions—a 1 percent increase in greenhouse gas emissions. But some exceptions offer opportunities to promote strategies that both promote growth and limit emissions.

Ending encouragement to use fossil fuels is one avenue. Removing fossil fuel subsidies would increase economic efficiency and reduce greenhouse gas emissions (van den Berg and Cando-Noordhuizen 2017). The largest subsidizers in absolute terms were Egypt, India, Indonesia, the Islamic Republic of Iran, the Russian Federation, Saudi Arabia, and Ukraine—all with more than $10 billion a year in subsidies. Subsidies are 2–7.5 times larger than public spending on health in Bangladesh, Ecuador, Egypt, India, Morocco, Pakistan, the Republic of Yemen, Turkmenistan, and Venezuela.

Using energy-efficient instruments can also help (IED 2014; UNDP 2008). Compact fluorescent lamps draw only 20–30 percent as much power and last much longer. Substituting them for all the incandescent lamps in Sub-Saharan Africa would reduce peak power consumption by 15 gigawatts, roughly 23 percent of the installed capacity.

Together, these findings suggest that a win-win strategy could be built around introducing efficiencies while reducing subsidies and better targeting subsidies to the poor. This would simultaneously reduce the strain on government budgets, free resources to allow extension of energy sources to the poor, and promote more efficient energy use.

Another area of concern is the use of water resources (IEG 2010a; Thomas and Luo 2011). For almost a century, water use has been growing almost twice as fast as population. To meet the demand for water, numerous parts of the world have exceeded sustainable limits of water withdrawal from rivers and groundwater aquifer. The level of water in underground aquifers below Beijing, New Delhi, and many other booming cities is falling rapidly. Major rivers such as the Ganges, the Yangtze, the Nile, and the Jordan are overtaxed and regularly shrink for long periods during the year.

Water shortages already loom in many parts of the world. One-third of the world population, concentrated in developing countries, lives in basins where the water deficit is larger than 50 percent. About 700 million people in 43 countries face water stress, unable to obtain the minimum need of 1,700 cubic meters of water per person per year. And climate change aggravates erratic rainfall patterns, compounding the challenges.

The global water footprint reached 9,000 billion cubic meters a year in 1996–2005. Irrigated agriculture accounts for more than 80 percent of water use in developing countries. Yet, feeding more people and coping with the changing dietary demands from a richer population will require more efficient water use. Without sufficient water, future economic progress could be severely constrained.

But water stress is about more than availability. Rapid economic growth increases not only water use but also pollution. It has changed natural water reservoirs—directly, by draining aquifers, and indirectly, by melting glaciers and the polar ice caps. And overexploitation of groundwater results in salinization, while industrial and agricultural waste pollutes water sources.

The economic benefits of better managing water resources are big, as are the economic costs of inaction. Country examples indicate that proper water management could increase gross domestic product by 5–14 percent. In the Middle East and North Africa, where water shortages are most acute, the cost of environmental degradation from water pollution and excessive withdrawals is estimated at between 2.0 and 7.4 percent of gross domestic product.

It may not be just an issue of better managing scarce resources, but one of changing sources to more sustainable ones. Water scarcity should lead to higher prices for water, which will at a certain point make whole-scale desalinization of seawater economical, but with a transportation problem—how to get this water to remote regions and landlocked countries. Innovations in filters and other desalinization techniques could drive the price of desalinization down, which could bring the tipping point for the market in desalinized water closer.

Yet another aspect is the protection of forests (IEG 2009). An evaluation using forest fires as a proxy found that, on average, protected areas significantly reduce tropical deforestation and associated carbon release, thus reducing carbon emissions while preserving biodiversity. The study examined whether areas subject to strict protection—with essentially no use allowed—fared better than those in which some activity was permitted.

The expectation was that, all things equal, strict protection would have the bigger impact on reducing the incidence of fires, considering differences

in deforestation pressures. But the evaluation found instead that the impact was actually greater when the protected areas allowed sustainable use by local populations than when they did not. This finding is true for Africa, Asia, and Latin America and the Caribbean, when comparing the mean reduction in fire incidence from strict protected areas with that from multiuse protected areas. In Latin America, where indigenous areas can be identified, the impact on fire incidence is extremely large.

Closely related is the protection of the world's precious biodiversity (GEF 2016a). Placing a value on biodiversity loss is not easy, but the high cost of irreversible losses cannot be underestimated. Protecting biodiversity is a critical element for the protection of our planet, and it has been shown to carry with it valuable resources and sources of livelihood, especially for the poor.

NATURAL DISASTERS

Great floods in China and India, superstorms in the Philippines and in the United States, and summer heat waves in Australia and Japan in recent years are manifestations of an alarming trend in the rise of climate-related disasters. The 2010s may well go down as the decade when the trend line of these events headed aggressively upward after a noticeable rise in their intensity and frequency since the 1970s.

Global warming has contributed to warming oceans, more moisture in the air, and higher sea levels, but scientists have been cautious about attributing a flood or storm to climate change. Even so, papers have argued that the intensity of the 2011 Great Flood in Thailand and of Super Typhoon Haiyan in the Philippines are owed in part to changing climate. More recent work has been even more pointed: global warming is shown to have made Japan's unusually hot summer this year 1.5 to 1.7 times more likely.

A consensus, too, is building that climate change has roots in human actions (IPCC 2015). We have known for a long time that weather events turn into disasters for human-made reasons. More people are hurt when they are exposed in harm's way, and when they are vulnerable and unable to cope. But now we also know that the intensity and frequency of the hazards themselves are greater because of human-made global warming.

This understanding profoundly affects how countries engage in disaster risk reduction. Economic growth projections are contingent on addressing climate change. Yet few of the forecasts for global and country growth take into account the impacts of climate change that are already evident, or the massive investment and resources that will need to be mobilized for climate action. Such forecasting is missing from the current estimates for growth, for example, of around 3.0–3.5 percent in 2017 and 5.5–6.0 percent for Asia and the Pacific.

Countries and regions need to build contingency plans into their economic programs. Floods and storms in recent years inflicted sizable economic losses in Australia, China, Indonesia, Republic of Korea, Thailand, and Vietnam. After the financial crisis, governments and multilateral institutions intensified their efforts to anticipate future crises, carrying out stress tests of the

vulnerability and resilience of their banking systems. In the same way, we now need stress tests that can reveal how well countries can withstand the impact of rising natural disasters.

In many respects, such country actions bring both global and local benefits (GEF 2016b). Reducing black carbon emissions that blight so many cities (like Beijing and New Delhi) is a case in point. Phasing out the use of fossil fuels that present the greatest danger to our environment is another. India and Indonesia recently slashed fossil fuel subsidies. Investments in solar photovoltaics in China and Japan and in onshore wind across Europe are pointing the way for increased use of renewable energy.

The five cities most vulnerable to natural hazards are all in Asia: Bangkok, Dhaka, Jakarta, Manila, and Yangon. All of them are overcrowded and in geographically fragile settings. Asia's growth has been characterized by increasing urbanization, making it imperative that climate-friendly urban management become a strategic thrust. And because the poor are hit harder by the effects of climate change than the rest of the population, building resilient communities will be an essential element of poverty reduction strategies.

Climate-related natural disasters are no longer one-off occurrences; rather, they are systemic events that need preventive action. Disaster risk reduction needs to be seen as an investment, going beyond relief and reconstruction to a dual approach of prevention and recovery. Japan invests some 5 percent of its national budget in disaster risk reduction, and this has been shown to reduce human and economic losses when disasters strike.

The main message is that to deliver sustained growth and well-being, we need to value natural capital, recognize the human hand in climate change, and take preventive action against climate-related calamities.

THE DANGER OF CLIMATE CHANGE

Climate change is the greatest known threat to economic growth and well-being and its impacts go far beyond natural disasters (Stern 2006; Uitto, Puri, and van den Berg 2017). To confront this peril, world leaders, especially of the large economies, must commit to much stronger cuts in carbon emissions than currently envisaged at the United Nations conference in Paris. But the challenge is bigger. To bring about lasting change, countries will need to reform the way their economies generate growth.

What makes this difficult is political leaders' differing beliefs about what generates growth. After all, carbon-intensive production created wealth in the past, so many still see a change in this recipe as inimical to expansion. Yet the reality is just the opposite. In the face of mounting disasters linked to human-made global warming, a low-carbon path is the only way to progress.

To appreciate why, note that the current growth path within a quarter century will push carbon concentrations in the atmosphere to the critical 450 parts per million. Beyond this threshold, temperatures will rise above 2 degrees from pre-industrial levels, with catastrophic impact. Just-released data warn that we are already halfway to that dreaded mark. 2016 has surpassed 2015 as the hottest on record. Asia is on the front line of climate-related disasters.

To dodge this dangerous scenario, energy-related emissions alone need to fall by 40–70 percent below 2010 levels by 2050. With energy accounting for two-thirds of emissions today, the required shift from reliance on fossil fuels is huge. Currently, two-thirds to four-fifths of electricity relies on fossil fuels in China, Japan, Russia, and the United States. China and India continue to ramp up coal-fired power stations.

Industrial countries were far and away the main cause of past carbon build-up. But developing countries in Asia is now the origin of some 37 percent of global emissions. Some countries, like Canada, generate relatively low total but high per capita emissions, while others, like India, generate a relatively high total but are low in per capita terms. Among those at the high end in total, Japan is moderate in per capita terms.

Regardless of the historical and current sources, what is clear is that business as usual will sink everyone. We need an economic transformation that is not only in the global interest but also in a country's own interest. Evaluation must take on board this reality, and be especially cognizant of five trends (Office of Evaluation and Oversight 2015; Thomas 2017).

First, renewable energy sources—solar, wind, wave, tidal, geothermal, and biomass—need to expand vastly, supported by research and development and exchange of knowledge. Battery storage, smart grids, and demand measures have to improve. Demand for renewable energy can be augmented by a carbon tax that reduces demand for dirty fossil fuels. Cap and trade schemes can also help, as China plans for 2017. There is a heightened policy debate in Japan about raising the ratio of electricity from renewable and other nonpolluting energy, including the role of nuclear plants.

Second, countries need to move much more quickly out of using polluting fossil fuels. Fossil fuel subsidies in financial terms might amount to some $550 billion globally. But when their negative effects are incorporated, the effective subsidies are much higher (IMF 2015). These subsidies have to be slashed, as India and Indonesia have started to do. The Obama administration's decision to reduce carbon emissions from power plants by 32 percent below 2005 levels by 2030 is positive, but it now needs to be maintained under the Trump administration. Japan is trying to encourage cleaner energy, including via hybrid and electric engines, and is promoting export of cleaner technology.

Third, dealing with local pollution also helps climate mitigation. Abatement of air pollution is urgent in Delhi and Beijing. Some 3 million people die each year from outdoor air pollution. Urban congestion can be lessened with intelligent transport systems, as in Seoul. Corruption and greed damage developed countries too: Volkswagen's cheating on automobile emissions, for example, is a colossal scandal.

Fourth, to withstand destruction from global warming, we need to strengthen roads and embankments, build in safer areas, and invest in rain harvesting, drainage, and early warning, as Japan has been doing. Countries can tap new financing such as the Green Climate Fund, as Fiji just did. It received a $31 million Green Climate Fund grant for a project supported by the Asian Development Bank.

Fifth, we need to protect coastal zones, agricultural land, and forests. In Indonesia, haze from slash-and-burn agriculture to clear areas for palm oil

every year spreads through Southeast Asia, ruining people's health, biodiversity, and economic activities. These fires, on the worst days, emit more carbon than the U.S. economy.

Evaluators have been slow to send this message, but it is now urgent that the discipline come to grips with it. It is only with a swift response to climate change that countries can sustain economic growth and well-being. As a major contributor to the discussions of development effectiveness, evaluation should account for climate effects and provide evidence on social, economic, and environmental costs of delayed action.

ENVIRONMENTAL AND SOCIAL SAFEGUARDS

The value of having environmental and social safeguards is a major area of inquiry. Some studies have pinned the cost of having these safeguards as less than 3 percent of the administrative budgets of projects (IED 2016; IEG 2010b), while their benefits in terms of avoided losses have been far higher.

The demand for safeguards (to manage environmental and social trade-offs) emerged in the 1980s in response to a number of controversial projects funded by the World Bank. Two examples of these projects are the Polonoroeste's BR-364 Amazon highway program in Brazil, which affected indigenous communities, and the Narmada Dam in India, which displaced 90,000 people. These resulted in the crafting of environmental and social policies at the World Bank to ensure a "do no harm" approach in its projects. The regional MDBs followed with similar policies.

MDBs' safeguard policies aim to promote the sustainability of projects by protecting people and the environment from the potential adverse effects of development. For example, the Asian Development Bank's safeguard policy lays down key requirements, including: (1) Identify and assess environmental or social impacts early in the project cycle; (2) develop and implement plans to avoid, minimize, mitigate, or compensate for the potential adverse impacts; and (3) inform and consult affected people during project preparation and implementation. The crucial question for evaluation is how effective the practices are.

The independent evaluation units of the Asian Development Bank and the World Bank (IED 2016; IEG 2010b) document some successes and important gaps. Both discuss strengths in the design of safeguards that must not be diluted and point out weaknesses in implementation and supervision of safeguards, especially for moderately risky projects. They support the eventual use of country systems when they are more equivalent and comparable to MDB systems, signal the need for great caution in switching to them, and recommend continued efforts to strengthen local capacities.

Evaluation must take on board four principles to govern the use of safeguards.

- Safeguards must be legally binding, and compliance should be enforceable. Standards that are to be met flexibly during a project's life will not suffice in ensuring protection against spillover damages. Sure, flexibility can speed up project approval, but for risky projects, the resulting damages could just delay project completions.

- International policy should govern safeguards, rather than national systems that by law or in practice are not yet equivalent. Recent years have seen several disasters under national systems; for example, the collapse of a garment factory near Dhaka, Bangladesh; a mining disaster involving a dam burst in Minas Gerais, Brazil; and explosions at a container storage station in Tianjin, China.
- It is not enough to have systems in place; implementation and oversight need strengthening. In particular, downstream supervision of how safeguards are being followed on the ground needs to be bolstered, but without weakening upstream regulation. Monitoring of impacts is essential, not by the investor alone but also by an independent party.
- The efficiency with which processes and procedures are followed can usually be improved a great deal. Greater differentiation in the treatment of high- and low-risk projects can help. Project processing can be speeded through such efficiency improvements, and not through a weakening of the regulation.

Effective safeguards are needed more than ever both for the established lenders and the newcomers. How the international banks apply these defenses will be an indication of their true commitment to the Sustainable Development Goals and the Paris climate accord.

Going forward, MDBs will have to focus more and more on the positive potential of their safeguard policies, i.e., move away from a "do no harm" philosophy to a more proactive stance of harnessing safeguards to "do good" on social and environmental issues. MDBs will also have to support strengthening member countries' capacity in implementing safeguard provisions.

The tensions raised in considering safeguards also provide one of the most powerful illustrations of trade-offs that evaluators must lay out—but which they often sidestep, at least explicitly. Those who have worked on safeguards over the years naturally recognize the uphill battle involved in taking on special interests who would rather not be burdened with adherence to any safeguards. The Trump administration's public statements are a case in point. Proponents of safeguards also have anecdotes and stories, if not quantitative evidence, of how much gain to society sound safeguards bring.

But on the other side, some operational staff in organizations, and certainly special interests, use anecdotes, if not quantitative evidence, of how much safeguards weigh down investment operations. They often minimize the gains that safeguards bring as well.

Clearly there are inherent trade-offs and also different sets of interests driving people's perceptions and even evidence. The role of sound evaluation under these circumstances cannot be overstressed. One way forward is the application of cost-benefit analysis that takes into account both private and social benefits and costs (IED 2016). Data are a constraint in applying such analysis, but where it has been done reasonably well, social benefits of most safeguards (i.e., avoided damages) far outweigh the social costs of having such safeguards (IEG 2010a). That does not mean costs can be reduced further with greater efficiency in executing safeguards. The policy implications

would be to adopt safeguards where net benefits are positive but to continue improving efficiency such that the net benefits are increased further.

CONCLUSION

An overarching implication of all this is the need for an introspective look at evaluation in the context of sustainable development. To remain relevant and effective, evaluation must not favor a risk-averse standpoint of doing what is easy and conventional. Evaluation should encourage innovative thinking and methods that shepherd the dynamics of sustainable development. In addition, systematic identification, analysis, and scaling-up of successful interventions are necessary actions to move forward sustainable development and growth.

Related to this, an equally important activity is developing further evaluation capacity at the country level, as promoted by an increasing number of countries. Implementation is expected to be the key test for the Sustainable Development Goals, and monitoring and evaluation will be an important part given that the development initiatives and related evaluations will be country led.

REFERENCES

Basu, Kaushik, Francois Bourguignon, Justin Yifu Lin, and Joseph E. Stiglitz. 2017. "A New Year's Development Resolution." *Project Syndicate*, Dec. 31, 2016.

Dasgupta, Partha. 2009. "The Welfare Economic Theory of Green National Accounts." *Environmental and Resource Economics* 42: 3–38.

GEF (Global Environment Facility). 2016a. "Biodiversity Mainstreaming in Practice: A Review of GEF Experience." GEF, Washington, DC.

———. 2016b. "The Opportunity of the Commons." GEF, Washington, DC.

Hamilton, Kirk. 2014. "Measuring Sustainability in the UN System of Environmental-Economic Accounting." Grantham Research Institute on Climate Change and the Environment Working Paper No. 154. London School of Economics, London.

IED (Independent Evaluation Department). 2014. *2014 Annual Evaluation Review.* Manila: Asian Development Bank.

———. 2015. *2015 Annual Evaluation Review.* Manila: Asian Development Bank.

———. 2016. *Real-Time Evaluation of ADB's Safeguard Implementation Experience Based on Selected Case Studies.* Manila: Asian Development Bank.

IEG (Independent Evaluation Group). 2009. *Climate Change and the World Bank Group: Phase One: An Evaluation of World Bank Win-Win Energy Policy Reforms.* Washington, DC: World Bank.

———. 2010a. *Cost-Benefit Analysis in World Bank Projects.* Washington, DC: World Bank.

———. 2010b. *Safeguards and Sustainability Policies in a Changing World: An Independent Evaluation of World Bank Group Experience.* Washington, DC: World Bank.

IMF (International Monetary Fund). 2015. "Counting the Cost of Energy Subsidies." IMF Survey. Washington, DC.

IPCC (Intergovernmental Panel on Climate Change) 2014. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Geneva: IPCC.

Morgan, T. 2007. "Energy Subsidies: Their Magnitude, How They Affect Energy Investment and Greenhouse Gas Emissions, and Prospects for Reform." UNFCCC Financial and Technical Support Programme, Bonn.

Office of Evaluation and Oversight. 2015. "Evaluating Climate Change." Washington, DC: Inter-American Development Bank.

Petrie, Peter, and Vinod Thomas. 2013. "Development Imperatives for the Twentieth Century." Economics Working Paper Series 360. Manila: Asian Development Bank.

Stern, N. 2006. *Stern Review on the Economics of Climate Change*. Cambridge, UK: Cambridge University Press.

Thomas, Vinod. 2017. *Climate Change and Natural Disasters*. Rutledge.

Thomas, Vinod, Mansoor Dailami, Ashok Dhareshwar, Daniel Kaufmann, Nalin Kishor, Ramón Lopez, and Yan Wang. 2000. *The Quality of Growth*. New York: Oxford University Press.

Thomas, Vinod, and Xubei Luo. 2011. *Multilateral Banks and the Development Process*. Transaction Publishers.

Uitto, Juha I., Jyotsna Puri, and Rob D. van den Berg, eds. 2017. *Evaluating Climate Change Action for Sustainable Development*. New York: Springer International Publishing.

UNDP (United Nations Development Programme). 2008. *Evaluation of Role and Contribution of UNDP in Environment and Energy*. New York: UNDP.

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.

Chapter 2

No One Left Behind – A Focus on Gender and Social Equity

Marco Segone and Florencia Tateossian

Abstract. *This chapter presents the importance of evaluating the Sustainable Development Goals (SDGs) with a focus on gender and social equity. It analyzes the transformative nature of the new 2030 Agenda, due to its focus on making sure that "no one is left behind." It explains how the SDGs expand and build on the Millennium Development Goals, as well as how heads of states from all over the world have made a commitment to ensure a systematic follow-up and review of the SDGs that is "robust, voluntary, effective, participatory, transparent and integrated" in order to track progress, and argues that to ensure that no one is left behind, the follow-up and review should be informed by country-led evaluations that are equity focused and gender responsive. This should be accompanied by strengthening national evaluation capacities through a systemic approach that looks at the enabling environment and at both institutional and individual capacities, from the supply as well as the demand side. Finally, it argues that strengthening national evaluation capacities to evaluate SDGs in such a way as to ensure that no one is left behind is a common endeavor that requires strong partnerships among various actors, such as national evaluation systems, parliamentarians, voluntary organizations for professional evaluation, and civil society.*

Marco Segone, UNFPA, segone@unfpa.org; Florencia Tateossian, UN Women, florencia.tateossian@unwomen.org.

A TRANSFORMATIVE AGENDA FOR "THE WORLD WE WANT"

We live in a world where a massive concentration of wealth and privilege exists in the hands of a few: the richest 1 percent of the population owns 40 percent of the world's wealth, while the poorest 50 percent of the population owns only 1 percent of the world's wealth. The three richest people in the world own wealth equivalent to the combined gross domestic product (GDP) of the world's 49 poorest countries. It may seem that this is only related to income, but a similar situation exists in the statistics concerning human development as well. Human development indicators from 2015 show that 793 million people are still malnourished (FAO 2015), and that one in three women will be beaten, raped, abused, or mutilated in their lifetimes.[1] These are just a few examples of the many that illustrate the current inequity in the world.

The question is: is this the world we want? Or would we like to live in a world in which inequities have been banished for all humans—everywhere, anytime? Most would agree this is a common goal: so how do we get there?

The good news is that the countries that endorsed the 2030 Sustainable Development Agenda recognize the importance of long-term equitable and sustainable development: more and more countries are implementing social and public policies to try to decrease the gap between those with the most (the best-off) and those with the least (the worst-off) (UN 2015b). In September 2015, leaders from around the world adopted the ambitious 2030 Agenda for Sustainable Development at a historic United Nations (UN) Summit.[2] Agenda 2030 calls for global transformation that focuses on ending poverty, protecting the planet, and ensuring prosperity for all. In January 2016, the 17 Sustainable Development Goals (SDGs) intended to implement this agenda came into force. These new goals—built on the success and the unfinished agenda of the Millennium Development Goals (MDGs)—call on all countries to mobilize efforts to end all forms of poverty, fight inequalities, and tackle climate change, while ensuring that "no one is left behind."

How do the SDGs expand upon and continue the work of the MDGs? There are a number of key differences, both in the process through which the SDGs were identified, as well as the content.

First, the SDGs were identified in a broad and inclusive process. For more than two years, governments, civil society, the private sector, and thought leaders from around the world negotiated and discussed the development of the SDGs. For the first time, 8 million people voted on which of the SDGs were most important to them. This inclusive and participatory process has also encouraged each country to adapt the SDGs to their own national contexts. This will make the level of ownership of the SDGs much stronger.

[1] UN Women, "Facts and figures: Ending violence against women," http://www.unwomen.org/en/what-we-do/ending-violence-against-women/facts-and-figures.

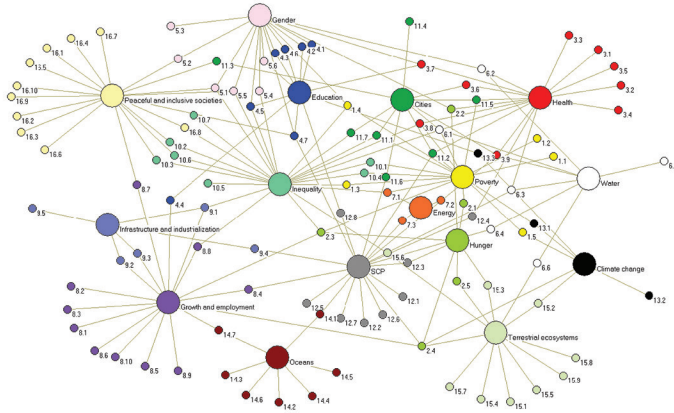
[2] The United Nations Summit for the adoption of the post-2015 development agenda was held in New York, September 25–27, 2015, and convened as a high-level plenary meeting of the General Assembly.

Second, the SDGs are universal. Unlike the MDGs, which had a strong focus on the developing countries (with seven of the eight goals devoted to them), the SDGs are relevant to any country of the world. Rob D. van den Berg, president of IDEAS, has reminded us that "from the perspective of the SDGs, all countries are developing countries."[3]

Third, the SDGs are comprehensive and integrated. While some have noted concern over the large number of goals (17), this also encourages sweeping transformation across a broad number of areas, and encourages the use of partnerships to accomplish these goals. To improve communication and ensure that people understand the ultimate intent of the SDGs and Agenda 2030, the UN has clustered them into "five Ps": people (human development); prosperity (inclusive economic development); planet (environment and climate change); peace (a key component of all development); and partnership (one of the few ways to achieve such sweeping transformation).

As shown in figure 2.1, the SDGs are interrelated and interlinked, which adds to their complexity, but also to their dynamic interaction.

FIGURE 2.1 **The SDGs as a network of targets**



SOURCE: Le Blanc 2015.

Fourth, gender equality and reduced inequalities among and within countries are both stand-alone goals, and they are both mainstreamed to all SDGs. The principle of "no one left behind" is the key principle informing every SDG, and is mainstreamed throughout the entire structure of Agenda 2030.

[3] Rob D. van den Berg, opening speech at 2015 IDEAS Conference, Bangkok.

Fifth—and very important to those in the evaluation community—Agenda 2030 and the SDGs include a follow-up and review mechanism, operating at the national, regional, and global levels. The principles for this mechanism are voluntary and country-owned; open, inclusive, and transparent; support the participation of all people and all stakeholders; are built on existing platforms and processes; avoid duplication; respond to national circumstances; and are rigorous and evidence-based, informed by data that is timely, reliable, and disaggregated. Most important for those in the evaluation community, the follow-up and review mechanism will be informed by country-led evaluations, and calls for strengthening national evaluation capacities.

THE GREATEST OPPORTUNITY AND THE GREATEST CHALLENGE FOR THE GLOBAL EVALUATION COMMUNITY

This is the first time in the history of international development that a follow-up and review mechanism to assess the implementation of a development agenda was adopted unanimously.. This high-level and far-reaching commitment will enable a surge in the demand for country-led evaluation. Key policy makers will request their own national evaluation systems so that they can produce high-quality evaluations to inform the national SDG reviews that countries will be presenting at the UN High Level Political Forum. This is therefore an unprecedented opportunity for the evaluation community.

On the other hand, evaluation of these broad-reaching goals with a central focus on "no one left behind" presents a number of unique challenges:

- How do we evaluate equitable development interventions?
- What are the best questions to use in order to assess whether interventions are relevant, and are having an impact in decreasing inequity, and in achieving results for the worst-off groups?
- What are the methodological, political, social, and financial implications of designing, conducting, managing, and using evaluations that are responsive to issues of social equity and gender equality?
- How can we strengthen the capacities of governments, civil society organizations (CSOs), and parliamentarians to evaluate the effect of interventions on equitable outcomes for marginalized populations?

EVALUATING THE SDGS WITH A "NO ONE LEFT BEHIND" LENS THROUGH EQUITY-FOCUSED AND GENDER-RESPONSIVE EVALUATIONS[4]

The 2030 Agenda made a commitment to ensure a systematic follow-up and review of the SDGs that would be "robust, voluntary, effective, participatory, transparent and integrated," and that would "make a vital contribution to implementation and will help countries to maximize and track progress in

[4] This section is drawn from Bamberger, Segone, and Tateossian (2016).

implementing the 2030 Agenda in order to ensure that no one is left behind" (UN 2015b). Country-led evaluations will be a central element used to inform SDG reviews and, together with strong monitoring data, will help support national policy decision making.

Gender equality and reducing inequalities between and among countries are central to the SDG principle of leaving no one behind. This recognizes the need to go beyond aggregate indicators, which only estimate the proportion of the population who have benefited from a particular intervention. There is evidence that aggregate indicators of progress can conceal the fact that some marginal or vulnerable groups are being left behind. In this context, the goal of the SDGs in reducing inequalities is to

- Identify groups that have been left behind;
- Understand why this has happened; and
- Identify strategies to promote more inclusive approaches that will include these groups.

While strengthening the national statistical system is of paramount importance in order to be able to produce disaggregated data that go beyond national averages, the evaluators will have to explain why certain groups have been left behind, and how this can be corrected. This is why equity-focused and gender-responsive evaluation (EFGRE) is vital.

UN Women, the UN entity for advocating for gender equality and women's empowerment, defines gender-responsive evaluation as having two essential elements: what the evaluation examines, and how it is undertaken. Gender-responsive evaluation assesses the degree to which gender and power relationships—including structural and other causes that give rise to inequalities, discrimination, and unfair power relations—change as a result of an intervention. This process is inclusive, participatory, and respectful of all stakeholders (rights holders and duty bearers). Gender-responsive evaluation promotes accountability regarding the level of commitment to gender equality, human rights, and women's empowerment by providing information on the way in which development programs are affecting women and men differently, and contributing to the achievement of these commitments. It is applicable to all types of development programming, not just gender-specific work (UN Women Independent Evaluation Office 2015). UNICEF, the UN agency for children, defines equity-focused evaluation as a judgment of the relevance, effectiveness, efficiency, impact, and sustainability of policies, programs, and projects that are concerned with achieving equitable development results (Bamberger and Segone 2011). This approach involves using rigorous, systematic, and objective processes in the design, analysis, and interpretation of information in order to answer specific questions, including those of concern to the worst-off groups. It assesses what does work to reduce inequities, and what does not, and it highlights the intended and unintended results for the worst-off groups, as well as the gaps between the best-off, average, and worst-off groups. It provides strategic lessons to guide decision makers and to inform stakeholders (Bamberger and Segone 2011). The UN

Evaluation Group, the professional network of evaluation offices of UN agencies, says in its guidance document that integrating human rights and gender equality in evaluations provides a valuable resource for all stages of the formulation, design, implementation, dissemination, and use of the human rights and gender-responsive-focused evaluations (UNEG 2014).

ROLES OF THE VARIOUS STAKEHOLDERS IN THE SDG FOLLOW-UP AND REVIEW MECHANISM

The key agencies responsible for the implementation of country-led evaluations within each country are the national governments. Since national SDG reviews (see box 2.1) are voluntary, the commitment of governments is critical, particularly as they are the ones who have to decide how to prioritize their limited financial and technical resources among many different development priorities—which are supported by different groups of international and national stakeholders. Given the broad scope of the SDGs, almost all government agencies will potentially be involved, and the national government will play an important coordinating role. One of the challenges is to avoid the "silo mentality" that has been seen in many of the MDG monitoring and evaluation (M&E) activities, where each sector agency works on its own sector-specific studies, with very little coordination between sectors.

At the national level, donor agencies, UN agencies, CSOs, advocacy groups, and foundations can all play important roles in determining the evaluation agenda. But there is always the danger that each donor agency, CSO, and UN agency will conduct their own studies, often with only limited coordination and comparability of data between entities, and with significant duplication. The Inter-Agency Expert Group on SDGs is seeking to avoid these issues by advocating for a global indicator framework for SDGs that would be agreed upon by all member states, with national and subnational indicators used for more localized policy interventions at the country level.

CSOs, including voluntary organizations for professional evaluation (VOPEs), will play an important role in the country-led evaluations at both the national and local levels, and their contribution will be critical in ensuring a truly inclusive consultation and participatory approach. While many governments collect data on local communities and are willing to involve these communities in the data collection process, government agencies are often less willing to involve them in the interpretation of the findings and in the discussion of the policy implications. Civil society, and particularly human rights and feminist groups will have an important role to play in ensuring that the voices of local communities and marginalized groups are heard.

STRENGTHENING NATIONAL EVALUATION CAPACITY FOR SDGS[5]

Using country-led evaluations to inform the SDG follow-up and review mechanisms goes hand in hand with strengthening national evaluation capacities.

[5] This section is drawn from Segone and Rugh (2013).

BOX 2.1 **Principles of SDG follow-up and review mechanisms**

The SDG follow-up and review mechanism will:

a. **Be voluntary and country led**; will take into account different national realities, capacities, and levels of development; and will respect policy space and priorities. Since national ownership is key to achieving sustainable development, the outcome from national-level processes will be the foundation for reviews at the regional and global levels, given that the global review will be primarily based on official national data sources.
b. **Track progress in implementing the universal goals and targets**, including the means of implementation in all countries in a manner that respects their universal, integrated, and interrelated nature as well as the three dimensions of sustainable development.
c. **Maintain a longer-term orientation; identify achievements, challenges, gaps, and critical success factors; and support countries in making informed policy choices.** This will help mobilize the necessary means of implementation and partnerships; support the identification of solutions and best practices; and promote the coordination and effectiveness of the international development system.
d. **Be open, inclusive, participatory, and transparent for all people**; and will support reporting by all relevant stakeholders.
e. **Be people centered and gender sensitive; will respect human rights; and will have a particular focus on the poorest, most vulnerable, and those furthest behind.**
f. **Will build on existing platforms and processes where these exist; will avoid duplication; and will respond to national circumstances, capacities, needs, and priorities.** These will evolve over time, taking into account emerging issues and the development of new methodologies, and will minimize the reporting burden on national administrations.
g. **Will be rigorous and evidence-based, and will be informed by country-led evaluations and data that are high-quality, accessible, timely, reliable, and disaggregated by income, sex, age, race, ethnicity, migration status, disability, and geographic location and other characteristics that are relevant in national contexts.**
h. **Will require enhanced capacity-building support for developing countries, including the strengthening of national data systems and evaluation programs**, particularly in African countries, least developed countries, small island developing states, landlocked developing countries, and middle-income countries.
i. **Will benefit from the active support of the UN system and other multilateral institutions.**

SOURCE: UN 2015, paragraph 74; emphasis added.

Country-led evaluations are essential in order to bring evidence together with monitoring data to inform the review of the SDGs.

The United Nations General Assembly resolution on national evaluation capacity building (UN 2015a) has set the stage for an understanding of the importance of evaluation capacity development.

This resolution emphasizes the importance of building capacities for the evaluation of development activities at the country level. It calls for interaction and cooperation among all relevant partners, including those of the UN system, and both national and international stakeholders, to coordinate efforts to further strengthen member state capacities for evaluation. Most importantly, the resolution emphasizes that national ownership and national priorities form a strong base for building national capacities to manage and oversee evaluations. Through this resolution, the member states agree that evaluation is an important component of development processes, and recognize evaluation as a country-level tool that can help strengthen and support development results toward the achievement of the SDGs.

National evaluation capacity development is a complex field in which different stakeholders have different roles to play, based on their respective value added. This complexity encourages the use of a systemic approach to national evaluation capacity development, while fully recognizing that each country has its own unique context and realities. This makes it necessary to not only look at actors at different levels and across sectors, but also, crucially, at the network of relationships or connections between them in each country. Such a viewpoint illustrates the fact that weaknesses in capacity at any level, or with any key actor, will affect the capacity of the whole system to deal with a problem in order to achieve a goal. Therefore, a country-specific systemic approach to national evaluation capacity development is needed, particularly when addressing evaluation capacities for country-led evaluations of the SDGs that are equity focused and gender responsive.

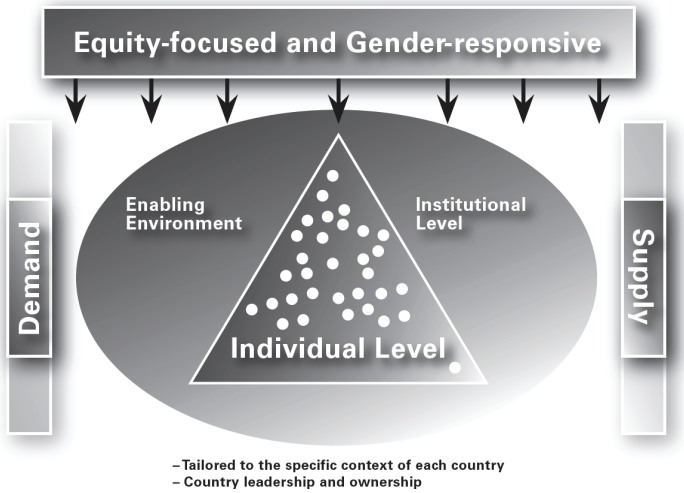
Individual and Institutional Evaluation Capacities Enabled by a Supportive Environment

In the past, evaluation capacity development focused on strengthening the capacities of the knowledge and skills of individuals. However, it is by now clear that capacity development should be based on a systemic approach that takes into account three major levels (individual, institutional, and external enabling environment); and two components (demand and supply[6]), and that both should be tailored to the specific context of each country (figure 2.2).

The enabling environment for evaluation is determined by a culture of learning and accountability, by which we mean the degree to which information is sought about past performance; and the extent to which there is a drive to

[6] "Supply" refers to the capability of professional evaluators to provide sound and trustworthy evaluative evidence. "Demand" refers to the capability by policy makers and senior managers to request sound and trustworthy evaluative evidence, with the aim of using it in strategic decision-making processes.

FIGURE 2.2 **A systemic and integrated approach to national evaluation capacities development**



SOURCE: Segone and Rugh (2013), 5.

continuously improve, and to be responsible or accountable for actions taken, resources spent, and results achieved. Such a culture is embedded in tacit norms of behavior, and an understanding of what can and should—or should not—be done; and in many cases, by behaviors being role-modeled by leaders.

An enabling environment is also supported or created through governance structures that demand independent evaluation, be it through parliaments or governing bodies, and that is further enhanced through VOPEs that set standards and strive toward greater professionalism in evaluation. Therefore, VOPEs should be supported, so that they can foster indigenous demand for and supply of evaluation, including by the setting of national evaluation standards and norms. There are also examples of governments soliciting the advice and involvement of VOPEs, not only in formulating evaluation policies and systems, but also in the implementation of evaluations that are consistent with those policies.

The institutional framework for evaluation ensures that a system that can implement and safeguard the independence, credibility, and utility of evaluation within an organization exists. Such an institutional framework has the following characteristics:

- Includes a system of peer review, or assurance that the evaluation function is set up to safeguard and implement the principles of independence, credibility, and utility
- Establishes safeguards to protect individual evaluators, evaluation managers, and the heads of evaluation functions

- Puts in place a multidisciplinary evaluation team that can ensure the credibility of evaluation by understanding multiple dimensions of evaluation subjects and combining the necessary technical competence
- Secures the independent funding of evaluations at an adequate level, to ensure that the necessary evaluations are carried out, and that budget holders do not exercise inappropriate influence or control over what is evaluated and how
- Combines measures for impartial or purposeful selection of evaluation subjects to ensure impartiality on the one hand, and increased utility on the other, by making deliberate choices linked to decision-making processes
- Sets out a system to plan, undertake, and report evaluation findings in an independent, credible, and useful way (to increase objectivity in the planning and conduct of evaluations, systems that increase the rigor, transparency, and predictability of evaluation processes and products are needed)
- Institutes measures that increase the usefulness of evaluations, including the sharing of findings and lessons learned that can be applied to other subjects

An evaluation environment is essential to support country-led evaluations of the SDGs. The UN resolution on capacity building for evaluation at the country level, and the strong commitment of evaluation to support the follow-up and review of the SDGs, are key drivers to enhance evidence-based policy making to achieve the SDGs.

At the individual level, a capacity development strategy should strengthen the ability of senior management to strategically plan evaluations and to identify the key evaluation questions, to manage evaluations for their independence and credibility, and to effectively make use of evaluation results. Mackay underlines the importance of identifying and supporting leaders or natural champions who have the ability to influence, inspire, and motivate others to design and implement effective evaluation systems (Mackay 2007). Leadership is not necessarily synonymous with a position of authority; it can also be informal, and can be exercised at many levels. Therefore, the evaluation capacity development strategy should, especially in the initial stages, identify and support as appropriate, national and local leaders in public administration and intergovernmental monitoring, as well as in evaluation groups and national VOPEs. It should also be linked to the national processes that focus on the country-level review of the SDGs. By giving national M&E departments or agencies responsibility for SDG follow-up and review, evaluation can become a key source of support for these national reviews. On the supply side, a capacity development strategy should enhance behavioral independence—independence of mind and integrity; knowledge of and respect for evaluation standards; and agreed-upon evaluation processes and products—as well as professional competencies through formal education, specialized training, professional conferences and meetings,

on-the-job training such as joint country-led evaluations, and communities of practice and networking, for example VOPEs.

Fostering Demand for and Supply of Evaluation

A distinction should be made between the capacity of policy makers and advisors to use evidence, and the capacity of evaluation professionals to provide sound evidence. While it may be unrealistic for policy makers and advisors to become competent experts in evaluation, it is both reasonable and necessary for such professionals to be able to understand and use the evidence produced by evaluation systems in their policies and practices. Integrating evidence into practice is a central feature of policy-making processes, and in this case, for integrating it into the follow-up and review mechanisms of the SDGs. An increasingly necessary skill for professional policy makers and advisors is to know about the different kinds of evidence that are available; how to gain access to them; and how to critically appraise evidence. Without such knowledge and understanding it is difficult to see how a strong demand for evidence can be established and, hence, how to enhance its practical application. However, it is also important to take into consideration that the national SDG review process is a political process, informed by evidence. The use of evidence in national SDG reviews depends on the combination of the capacity to provide quality and trustworthy evidence on the one hand, and the willingness and capacity of policy makers to use that evidence on the other. The extent to which evidence is used by policy makers depends, in turn, on the policy environment. To strengthen an enabling policy environment, policy makers may need to provide incentives to encourage policy makers and advisors to use the available evidence. These can include mechanisms to increase the "pull" for evidence, for example, requiring spending bids to be supported by an analysis of the existing evidence base, as well as to facilitate the use of evidence, such as integrating analytical staff at all stages of the policy implementation. CSOs, including VOPEs, should play a major role in advocating for the use of evidence in policy implementation. Think tanks, with the help of mass media, can also make evidence available to citizens, and citizens can demand that policy makers make more use of it.

CONCLUSION

The way forward is complex, but one thing is clear: there is no single ministry or organization that can do it alone. Evaluating the SDGs to ensure that no one is left behind is a common endeavor that requires strong partnerships among a variety of actors.

EvalPartners—a global partnership for evaluation capacity development that brings together approximately 60 organizations, including regional VOPEs, UN agencies, multilateral banks, academies, CSOs, and governments from around the world—was launched in March 2012, with the purpose of strengthening evaluation capacity. In 2013, at the Third International Conference on National Evaluation Capacities in Brazil, EvalPartners declared 2015 the International Year of Evaluation (EvalYear). EvalYear was a global,

bottom-up movement acknowledged by the UN General Assembly. It advocated for and promoted demand and use of evaluation in evidence-based policy making and positioned evaluation in the policy arena. It also kicked off the launch of several more initiatives, including The Global Parliamentarian Forum for Evaluation, three new networks (EvalGender+, EvalYouth, and EvalSDGs), and the Global Evaluation Agenda 2016–2020. EvalPartners, together with EvalYear and the new networks, are part of an enabling environment to enhance the use of evaluation to inform SDGs follow-up and review mechanisms.

Because evaluation can be a powerful agent of change, it will be up to evaluators, policy makers, and leaders around the world to make Agenda 2030 a reality. Therefore, we encourage everyone—evaluators, commissioners of evaluation, policy makers, and parliamentarians, among others—to be ambassadors of evaluation within their departments, organizations, and countries. This is essential in order to make the ultimate goal of evaluation a reality, and to help enact the change from the world we have to the world we want.

REFERENCES

Bamberger, M., and M. Segone. 2011. "How to Design and Manage Equity-focused Evaluations." UNICEF, New York.

Bamberger, M., M. Segone, and F. Tateossian. 2016. "Evaluating SDGs with a "No One Left behind Lens" through Equity-Focused and Gender-Responsive Evaluations." UN Women, New York.

FAO (Food and Agriculture Organization of the United Nations). 2015. *The State of Food Insecurity in the World*. Rome: FAO.

Le Blanc, David. 2015. "Towards Integration at Last? The Sustainable Development Goals as a Network of Targets." DESA Working Paper No. 141. United Nations Department of Economic and Social Affairs, New York.

Mackay, K. 2007. *How to Build M&E Systems to Support Better Government*. Washington, DC: Independent Evaluation Group of the World Bank.

Segone, M., and J. Rugh, eds. 2013. "Evaluation and Civil Society: Stakeholders' Perspectives on National Evaluation Capacity Development." Evaluation Working Paper. New York: UNICEF.

UN (United Nations). 2015a. "Building Capacity for the Evaluation of Development Activities at the Country Level." A/RES/69/237. UN General Assembly, New York.

———. 2015b. "Transforming Our World: The 2030 Agenda for Sustainable Development." A/RES/70/1. UN General Assembly, New York.

UNEG (United Nations Evaluation Group). 2014. "Integrating Human Rights and G ender-Equality in Evaluations." UNEG Guidance Document. UNEG, New York.

UN Women Independent Evaluation Office. 2015. "How to Manage Gender-Responsive Evaluation: Evaluation Handbook." UN Women, New York.

Chapter 3

Evaluation for Improving People's Lives

Mallika R. Samaranayake and Asela Kalugampitiya

Abstract. *This chapter focuses on the evolutionary process of the acceptance of the use of evidence for policy formulation and decision making. An effective evaluation system and enabling environment are required for achieving such objectives. The complementarity between the demand for evaluation findings and the capacity to meet that demand is of paramount significance. A strong enabling environment for evaluation encourages the generation of impartial, technical, strategic, and citizen-based information to ensure that national planning and budgeting reflect the needs of the people. Evaluation provides a means to enhance participation of civil society groups, as well as an opportunity for stakeholders to interact with members of parliament and improve consultation and representation. The engagement of parliamentarians with evaluation is becoming increasingly important in the context of the Sustainable Development Goals. Parliamentarians are responsible for passing policies and laws, and for prioritizing and endorsing budget allocations, all of which requires evidence as a basis for national decision making. Parliamentarians are also well positioned to question disparities in society and approve the resources needed to overcome them, again requiring access to sound and comprehensive evidence to aid decision making and better serve those who are left behind. A growing movement of parliamentarians involved in this process can help lead the way.*

Mallika R. Samaranayake, Community of Evaluators–South Asia, mallikasamare@gmail.com; Asela Kalugampitiya, EvalPartners, asela.kalugampitiya@gmail.com.

Over the years, development cooperation has operated on the principle of promoting social justice through the transfer of resources to the poor. The global system of social justice provides long-term consideration of how to do this while a country is going through a process of development, even if there are no expectations that this will accelerate economic development. National governments and donor agencies have a long way to go in developing the ability to make transparent and evidence-based policies and decisions so that investments become more effective. This requires greater recognition of the various dimensions of poverty reduction, based on evaluation findings.

Poverty is related to other problems of underdevelopment. It has become a major global issue that affects more than half the world's population. In 2015, the World Bank revised the international poverty line: the new threshold is $1.90 per day. This is a very low margin: the people who live under this threshold are considered to live in extreme poverty. The World Bank's mission is a "world free of poverty." As such, the role of evaluation cannot be underestimated in understanding the causal factors of poverty and identifying appropriate interventions in addressing such issues.

EvalPartners, in collaboration with other stakeholders, developed and launched the Global Evaluation Agenda 2016–2020, according to which national evaluation policies and systems play an important role at the country level. This agenda builds on support from the United Nations (UN) General Assembly Resolution GA/RES/69/237 (UN 2015a) and the more than 90 events that were organized around the globe in celebration of EvalYear 2015.

EvalAgenda 2020 (EvalPartners 2016) highlights the importance of strengthening an enabling environment for evaluation by developing institutional capacities, including voluntary organizations for professional evaluation (VOPEs); individual capacities for evaluation, focusing on evaluators, commissioners, and users of evaluation; and the need for developing links among stakeholders. Such an approach ensures that "no one is left behind" in the evaluation process, and that interventions are designed for sustainable development.

In the above context, evaluation is not a "stand-alone." The complementarity between the demand for evaluation findings and their use, and the capacity and ability to supply the respective services, is of paramount importance.

This chapter focuses on the historical factors and evolutionary process of acceptance of evaluation as a means of providing much-needed evidence for policy formulation and decision making. As such, utilization-focused evaluation, as well as equity and gender-focused evaluation, have come to be accepted progressively. Achievement of such objectives is envisaged through an effective evaluation system supported by an enabling environment, and institutional and individual capacity development so that in time "improving people's lives" with "no one left behind" will become a reality.

MILLENNIUM DEVELOPMENT GOALS

The Millennium Development Goals (MDGs) are the world's time-bound and quantified targets for addressing extreme poverty in its many

dimensions—poverty, hunger, disease, lack of adequate shelter, and exclusion—while promoting gender equality, education, and environmental sustainability. They are also basic human rights—the rights of each person on the planet to health, education, shelter, and security.[1] In September 2000, the leaders of 189 countries gathered at UN headquarters and signed the historic Millennium Declaration, committing to achieve a set of eight measurable goals that ranged from the reduction of extreme poverty and hunger by half, to promoting gender equality and reducing child mortality by the target year of 2015. Despite the progress made in reducing poverty, the number of people living in extreme poverty globally remains unacceptably high, and given global growth forecasts, poverty reduction may not be fast enough to reach the target of ending extreme poverty by 2030. According to the World Bank, in 2013 10.7 percent of the world's population lived on less than $1.90 a day, compared to 12.4 percent in 2012, and 35 percent in 1990. This means that in 2013, 767 million people lived on less than $1.90 a day, compared to 881 million in 2012, and 1.85 billion in 1990.[2] While the trend is certainly positive, as the Bank notes, "The effort to end extreme poverty is far from over, and there are many challenges remaining."

SUSTAINABLE DEVELOPMENT GOALS

The UN Conference on Sustainable Development held in Rio de Janeiro in June 2012 (Rio+20) stimulated a process for developing a new set of Sustainable Development Goals (SDGs). These goals were formulated in a highly participatory process; and they were subsequently approved by the General Assembly of the United Nations in September 2015. This is the new global plan for improving people's lives: it is comprised of 17 goals and 169 targets aimed at resolving major socioeconomic issues. It will cover the next 15 years. The SDGs have replaced the MDGs, which expired in 2015. While the MDGs focused primarily on poverty and health, the SDGs also cover the environment, human rights, and gender equality, among other new goals.

In this context, many organizations will integrate the SDGs into the development programs that will be carried out with the goal of improving peoples' lives around the globe from 2016 onward. Measuring and evaluating these programs will help donors, implementing agencies, beneficiaries, and other stakeholders to identify trends, measure changes, and capture knowledge in order to improve the performance of programs and increase transparency. A fundamental principle underpinning this process is giving a voice to the people themselves. This highlights the need for participatory evaluation processes, methods, and tools as part of the capacity building of evaluators, institutions, and other stakeholders.

[1] United Nations, http://www.un.org/millenniumgoals/bkgd.shtml.

[2] World Bank, "Poverty," http://www.worldbank.org/en/topic/poverty/overview.

THE NEED FOR PARTNERSHIPS IN PROMOTING EVALUATION

The emergence of global partnerships, reflected by the Fourth International Conference on National Evaluation Capacities (NEC) and the International Development Evaluation Association (IDEAS) Global Assembly in 2015, which were conducted in parallel, can be considered a great achievement and a contributory factor in achieving the Global Evaluation Agenda. The NEC Conference addressed the issue of integrating evaluation principles with human development practices. The IDEAS conference focused on evaluating sustainable development by enhancing evaluation capacities, both institutional and individual. These two initiatives complemented each other in reaching toward the final goal of improving people's lives.

In the context of the SDGs, it is necessary to assess how such goals are being realized in the developed world as well, since the SDGs are being adopted by *all* countries, both developed and developing. With this paradigm shift toward supporting the SDGs, it is prudent to examine how evaluation will be able to provide evidence of such support in achieving the SDGs.

As emphasized by the Paris Declaration on Aid Effectiveness and the follow-up declarations of Accra, Busan, and—more recently—Nairobi (in the Global Partnership for Effective Development Cooperation),[3] the need for international cooperation to support development in partner countries in line with priorities that reflect those of the people is now well established: this includes the role of evaluation, and is a major breakthrough. It represents a paradigm shift from donor-driven evaluations to country-owned joint evaluations.

Conceptual Framework

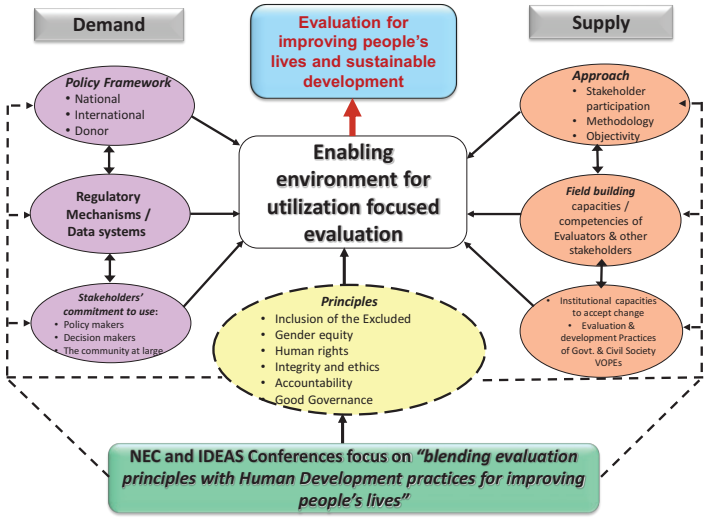
The current situation can be reflected upon in the context of two fundamental questions for better understanding:

- How can the SDGs empower evaluators to provide evidence that the lives of people are being improved in a more sustainable way?
- How can the paradigm shift toward sustainability ensure that people can achieve a balance between their economic, social, and environmental needs, both for the present and for the future?

A conceptual outlook on these two questions can be shown in diagrammatic form, depicting the demand and supply aspects within a utilization-focused evaluation framework, as captured in figure 3.1.

[3] See http://www.oecd.org/dac/effectiveness/thehighlevelforaonaideffectivenessahistory.htm for Paris, Accra, and Busan; and http://effectivecooperation.org/events/2016-high-level-meeting/ for Nairobi.

FIGURE 3.1 **Demand and supply aspects of a utilization-focused evaluation framework**



SOURCE: M.R. Samaranayake.

Demand for Evidence-Based Evaluation

The demand for evaluation and the use of its findings are very much dependent on the perceptions and understanding of various stakeholders regarding the added value evaluation can bring to the policy formulation and decision-making processes. The commitment of policy makers to transparency and the use of evaluation findings in policy formulation is considered a significant factor underlying the development agenda. In this context, there is an increasing focus on the role of parliamentarians in these processes.

Demand for evaluation is created through an enabling environment comprised of policy frameworks at the national, international, global, and donor levels. In this context, cooperation and partnership for promoting evaluation of both public and private sector programs, including the commitment of policy makers to promote and use evaluation findings in policy formulation, become critical factors. Thus:

- **Regulatory mechanisms** such as results-based monitoring and evaluation (M&E) systems need to be in place, to ensure that the milestones are being achieved as planned, with evidence being used for decision making.
- **Stakeholders' commitment** to using evaluation findings is another crucial factor in making evaluations matter. This involves policy makers, decision makers, implementers, and the overall community. There is an emerging trend of civil society organizations moving

from service delivery to advocacy, in order to secure more sustainable, widespread change. More important in the context of the SDGs is the involvement of the people themselves, since the final goal is improvement of their lives. A crucial element to be considered is the willingness and commitment of the commissioners of evaluations (i.e., government and donor agencies) for independent evaluation findings, both positive and negative.

- **Community participation** in the evaluation process becomes an important underpinning factor, allowing the voices of the people to be heard.
- **Political will** for the use of evaluation findings and for providing space for the involvement of the people should prevail. The involvement of parliamentarians as policy makers is crucially significant in order for political will to use evaluation findings and influence policy.

A PARLIAMENTARIAN MOVEMENT FOR EVALUATION

Developing and strengthening evaluation policies in countries is important for good governance and effective development. Moreover, it is implicit in UN General Assembly Resolution GA/RES/69/237 (UN 2015a) and the SDGs and their guiding framework, the 2030 Agenda for Sustainable Development (UN 2015b). The SDGs focus on country-led evaluation in line with identified priorities for the SDG targets that are most relevant to national and local contexts. This has emphasized the need for countries to strengthen their data collection, analysis, and review processes. The importance of evaluation is highlighted in the 2030 Agenda, which states that review of the SDGs will be "rigorous and based on evidence, informed by country-led evaluations," and calls for "strengthening of national data systems and evaluation programs." One of the key principles of the SDGs, "no one left behind," points to the importance of achieving equity-focused sustainable development. In many countries, one of the challenges is that disadvantaged communities sometimes do not receive the benefits of development. This is why equitable development needs to be emphasized through equity-focused and gender-responsive evaluation.

According to the Parliamentarians' Forum for Development Evaluation mapping study, as of 2015, only 20 countries had established national evaluation policies (PFDE 2015). This shows how far there is to go. The Global Parliamentarians Forum for Evaluation (GPFE) plans to advance this important work on national evaluation policies and systems (box 3.1).

Among Asian countries, only two, the Philippines and Malaysia, have endorsed national evaluation policies. In addition, Malaysia has a strong integrated results-based management system that is used in all governmental ministries. Although the Philippines has endorsed a national evaluation policy (NEDA and DBM 2015), it is yet to be operationalized. Afghanistan, Bhutan, Nepal, and Sri Lanka have draft policies that were developed through stakeholder consultations and have been submitted to their governments for endorsement. Out of these four draft policies, the Afghanistan policy includes a section on equity and gender (section 3.3). In Nepal and Sri Lanka, the draft

BOX 3.1 **Key facts on the Global Parliamentarian Forum for Evaluation**

A movement by parliamentarians toward the use of evaluation has grown rapidly in the past few years. Particularly during 2014–15, regional parliamentarian forums were created in the Africa, East Asia, Latin America, and Middle East and North Africa regions. The first-ever parliamentarians' panel on evaluation was held at the Community of Evaluators of South Asia's Evaluation Conclave 2013 in Nepal. This was a historic milestone, as it was the first time parliamentarians raised their voices to advocate for national evaluation policies and to commit to put evaluation at the core of the country-level agenda. The Parliamentarians Forum for Development Evaluation (PFDE) was established in South Asia in early 2013. Thereafter, parliamentarians were featured in many international evaluation events promoting national evaluation capacities. One of the key milestones in this regard is a study mapping the status of national evaluation policies, which was conducted by PFDE with support from EvalPartners, the global movement to strengthen national evaluation capacities (PFDE 2015). This helped promote national evaluation policies, including through regional consultations.

The African Parliamentarians Network on Development Evaluation (APNODE) was initiated at the African Evaluation Association conference held in Yaoundé, Cameroon, in March 2014, a year after the initiation of PFDE. APNODE is hosted and supported by the African Development Bank, and is the most formal group among all the parliamentarian forums currently active. In 2015, regional parliamentarians' forums were initiated in other regions.

More importantly, the first-ever national parliamentarians' forum for evaluation was initiated in Nepal by a group of parliamentarians representing all political parties. In Kenya, a caucus for evaluation was initiated to advocate for evaluation in the Kenyan Parliament. In this context, the Global Parliamentarians Forum for Evaluation (GPFE) was launched on November 25, 2015, at the Parliament of Nepal, on the occasion of celebrating the International Year of Evaluation (EvalYear).

SOURCE: GPFE website, https://globalparliamentarianforum.wordpress.com/; used with permission.

policies are being reviewed by stakeholders for inclusion of equity-focused and gender-responsive evaluation. Nepal is the only country in the region that has evaluation included in its Constitution, and in which the national evaluation policy will be formalized through an act of Parliament. The Evaluation Community of India has formed a task force to work on its national evaluation policy. Bangladesh and Pakistan are planning to work on their national evaluation policies as well.

In Nepal, all stakeholders, including parliamentarians, the government (through the National Planning Commission), VOPEs, development partners, academia, and the media are supporting the national evaluation policy and its formulation process. Nepal was the first country to initiate a National Parliamentarians Forum for Development Evaluation. In this forum, most Nepalese political parties are represented, and it actively supports the case for evaluation in the country. Another example of working with stakeholders is in Sri Lanka, where the National Parliamentarians Forum for Evaluation, the government, VOPE, and development partners all work together in stakeholder consultation meetings. Cambodia, Indonesia, Laos, Mongolia, Myanmar, Thailand, and Vietnam have established in-country VOPEs. In Cambodia, the VOPE, the government (through the Ministry of Planning), and parliamentarians are working together. Pol Ham, who is a member of the National Assembly and of the Steering Committee of the GPFE, is also the chair of the Parliament's standing committee on planning. He therefore will be able to officially support evaluation activities through the parliamentary system.

The VOPE in Mongolia is very new, and VOPEs in Laos, Myanmar, and Vietnam are also in their emerging stages. None of these countries have national evaluation policies and systems in place, nor have they even begun the process. Indonesia, Laos, Myanmar, Thailand, and Vietnam have been part of EvalPartners' Peer-to-Peer projects; and Laos, Mongolia, Myanmar, and Vietnam are part of one project in the fourth round of the Peer-to-Peer program. This participation will give these countries a chance to further advance evaluation culture within their countries and to strengthen VOPEs.

Cambodia, Nepal, and Sri Lanka are the three countries in the region that are involved in the EvalGender+ network-supported projects on evaluating SDGs through an equity and gender lens. Cambodia has developed guidelines on equity-focused and gender-responsive evaluation, and the final draft of these is available. In Nepal, under this project, a workshop was conducted on "Evaluating SDGs in Equity and Gender Lens" for all stakeholders; the country has also developed a national evaluation agenda and conducted various other activities, including development of an online repository that documents tools and methods, national networking meetings, etc. The Sri Lankan project has conducted a national stakeholder consultation to develop a national evaluation plan, held two meetings for parliamentarians, initiated the Sri Lanka Parliamentarians Forum, and conducted a training of trainers program for potential evaluators from the public sector. Interestingly, a young parliamentarian from Sri Lanka has submitted two motions, one on national evaluation policy, and one on the allocation of resources for evaluation from the national budget, to the Parliament.

SUPPLY FOR EVIDENCE-BASED EVALUATION

In order to meet the demand for evidence-based evaluation, there is a dire need for evaluation capacities to be developed and made available. Methodologies that ensure the active involvement of the people require an approach

that allows space for them to voice their views. Evaluators need to integrate participatory evaluation methods with systems analysis. Economic, social, and environmental sustainability requires that these systems interact with each other. To manage such integration, evaluators need to provide evidence of what people need, together with evidence of how far a particular solution would work both for the present and the future. The methodology should be comprehensive enough to allow for free expression of views by all stakeholders concerned: this is the challenge for evaluators.

Evaluation field building refers to the process of improving an organization's ability to use evaluation to learn from its work and improve results. Organizational evaluation approaches and practices need to be strengthened and the knowledge, attitudes, and skills of individual evaluators improved. The competencies of evaluators include coaching and training on the principles and techniques of evaluation, experience in conducting evaluations, on-the-job training, the appropriate use of tools and techniques, and the exchange of experiences among peers in different projects and different countries for the purpose of learning from each other. In this context, Fred Carden, Evaluation Director of the International Development Research Centre in Canada, emphasizes building evaluation capacity, as well as the need for improving the evaluation quality (Carden 2010). Concerns about evaluation capacity in international development are raised at many different levels: the small number of trained evaluators in many developing countries; the nature and location of training; the cultural, linguistic, and political differences that have to be considered in the evaluation process; and the capacity of evaluators to meet the needs of both donors and the specific countries.

Institutional capacity building to accept change also requires attention. Change that strengthens the organization's standing, influence, formal presence, or ability to achieve its goals, and the ability to address the demands from the state, civil society, the private sector, and the community when necessary are essential capacities to be developed in order to meet the demand for evidence-based evaluations. There is a demand for more training: funds are being established by some donors to support evaluation capacity building, and more organizations are trying to understand how they can play a useful role in addressing these gaps.

Evaluators can reflect on the principles, and suggest strategies to ensure integration with the human development practices that are included in the Global Evaluation Agenda 2016–2020, and that contribute to developing national capacities to evaluate sustainable development. It is pertinent to consider how inclusion of the excluded, gender equity, and human rights could be included in the evaluation process. Evaluator integrity and ethics are important aspects of moral character and involve a commitment to intellectual honesty and personal responsibility. Evaluation for accountability and governance are significant aspects on which the capacity building of evaluators needs to be focused.

These are some of the key challenges that evaluators will face:

- How policy change can be addressed to prevail positively for evidence-based decision making

- Bridging the gaps between the commissioners, practitioners, and users of evaluation
- How evaluators can provide evidence that improves the lives of people in a more sustainable way
- How the shift toward sustainability can ensure that civil society and people can achieve a balance between their economic, social, and environmental needs for both the present and the future, and can use data to hold the state accountable
- How to engage with people concerning their future needs, and how they see a sustainable balance emerging from their economic, social, and environmental needs
- How to bring the time dimension into their evaluations in blending evaluation principles and practices to support the SDGs
- How to provide evidence for the continuity of changes that need to be taken into account in improving people's lives, as required by the SDGs.

THE GLOBAL EVALUATION AGENDA 2016–2020

The Global Evaluation Agenda 2016–2020, also called EvalAgenda 2020, was formally launched at the Parliament of Nepal on November 25, 2015 (EvalPartners 2016). This agenda explains the demand as well as the supply implications for evidence-based evaluation. It is clear that evaluation as a tool for effective governance is becoming increasingly respected and implemented: evaluation has become so embedded in the vision of good governance that no policy maker or manager will imagine excluding it from the decision-making toolbox, dare to hold an important meeting, or reach an important decision without having first reviewed relevant evaluation information. It is also increasingly true that evaluators, whether internal or external, will use whatever methods and approaches are most appropriate to the situation to generate high-quality, ethical information that is pertinent to the issues at hand.

At the same time, it is envisaged that evaluation will help to amplify the voice of all stakeholders, particularly the marginalized and disadvantaged. Experience shows the difference evaluation can make in illuminating the realities of specific contexts by unpacking the complexity that people, organizations, and communities face as they struggle to address a variety of economic, social, and environmental issues. Experience shows the beneficial impact that principled evaluation can have in democratic settings when evaluators work in a neutral way, with all stakeholders contributing data, analysis, and insights in order to assess results, identify innovations, and synthesize learning toward improved outcomes.

Four essential dimensions of the evaluation system make up the core of EvalAgenda 2020: the enabling environment for evaluation; institutional capacities; individual capacities for evaluation; and links among these first three dimensions.

A **strong enabling environment** reflects the demand for evidence-based evaluations:

- All sectors of society understand and appreciate the value of evaluation
- Evaluation is explicitly required or encouraged in national evaluation policies and other governance and regulatory instruments
- Sufficient resources are allocated for evaluation, at all levels
- Credible, accessible data systems and repositories for evaluation findings are readily available
- Stakeholders are eager to receive and use evaluation information
- Evaluation receives due recognition as a profession and
- The ownership of public sector evaluations rests with national governments based on their distinctive needs and priorities, and with full participation of civil society and the private sector

Strong institutional capacities include the following:

- A sufficient number of relevant institutions, including but not limited to VOPEs, government agencies, civil society organizations (CSOs), academia, and institutions generate and share relevant data to develop and support evaluators and evaluation
- These institutions are capable of appreciating and facilitating quality evaluations
- They are skilled at collaborating with other relevant and involved institutions
- They are able to resource quality data generation and evaluations as required, make information readily accessible, and are ready to follow up on evaluation findings and recommendations
- They are able to continually evolve and develop as the evaluation field advances and
- Academic institutions have the capacity to carry out evaluation research and run professional courses in evaluation

Strong individual capacities for evaluation include the following:

- Developing individual capacities for evaluation will be relevant not only to evaluators, but also to the commissioners and users of evaluation
- Commissioners and users of evaluation have a sound understanding of the value of evaluation the processes for conducting high-quality, impartial evaluations and more commitment to using evaluation findings and recommendations
- Sufficient numbers of qualified evaluators, drawn from a diversity of relevant disciplines, are available to conduct high-quality evaluations in all countries and all subject areas
- These evaluators have the knowledge, skills, and dispositions to make appropriate use of generally accepted evaluation principles, theories, methods, and approaches

- Evaluators have integrated the values discussed above, and are culturally sensitive and
- Evaluators continually learn, and improve their capabilities

Strong links among these first three dimensions include the following:

- Governments, parliamentarians, VOPEs, the United Nations, foundations, civil society, the private sector, and other interested groups dedicate resources to joint ventures for the conducting of evaluations, innovation in the field of evaluation, and evaluation capacity building
- A common set of terms exists in all languages to disseminate and share evaluation knowledge
- Multiple partners in evaluation regularly attend national and international learning opportunities
- The "no one left behind" principle stated in the SDGs is embedded as a key value that goes across the three fundamental building blocks of an evaluation system: an enabling environment institutional capacities and individual capacities for evaluation

These four dimensions do not operate in isolation, but are connected in diverse ways in different countries, sectors, and situations. The relationships are dynamic, with overlapping influences, partners, and drivers; yet at the same time, all dimensions are working like a vortex, pulling the various dimensions ever closer to better outcomes. Each of the partners (institutions, individuals, and evaluation users) contribute a distinct part to the whole through the mutually supportive and interconnected dimensions of the agenda.

It is a collective hope and intention that by advocating for the many initiatives and activities outlined in the Global Evaluation Agenda, the global evaluation community will be able to make significant contributions to attaining EvalAgenda 2020, and all the SDGs, for the benefit of humankind. Each partner in the global community, including but not limited to parliamentarians, donors, governments, VOPEs, CSOs, the media, and the private sector, will have their role to play, and all of the stakeholders will be willing to work with parliamentarians to promote evaluation.

Parliamentarians can play a significant role in this process by demanding high-quality evaluations to ensure accountability. Parliamentarians can take the lead in promoting national evaluation policies and systems, and all parliamentarians and parliaments are expected to join hands with the evaluation community in this effort. Together, parliamentarians and evaluators can proceed toward achieving EvalAgenda 2020.

REFERENCES

Carden, Fred. 2010. "Introduction to the Forum on Evaluation Field Building in South Asia." *American Journal of Evaluation* 31 (2): 219–21. doi: 10.1177/1098214010366049

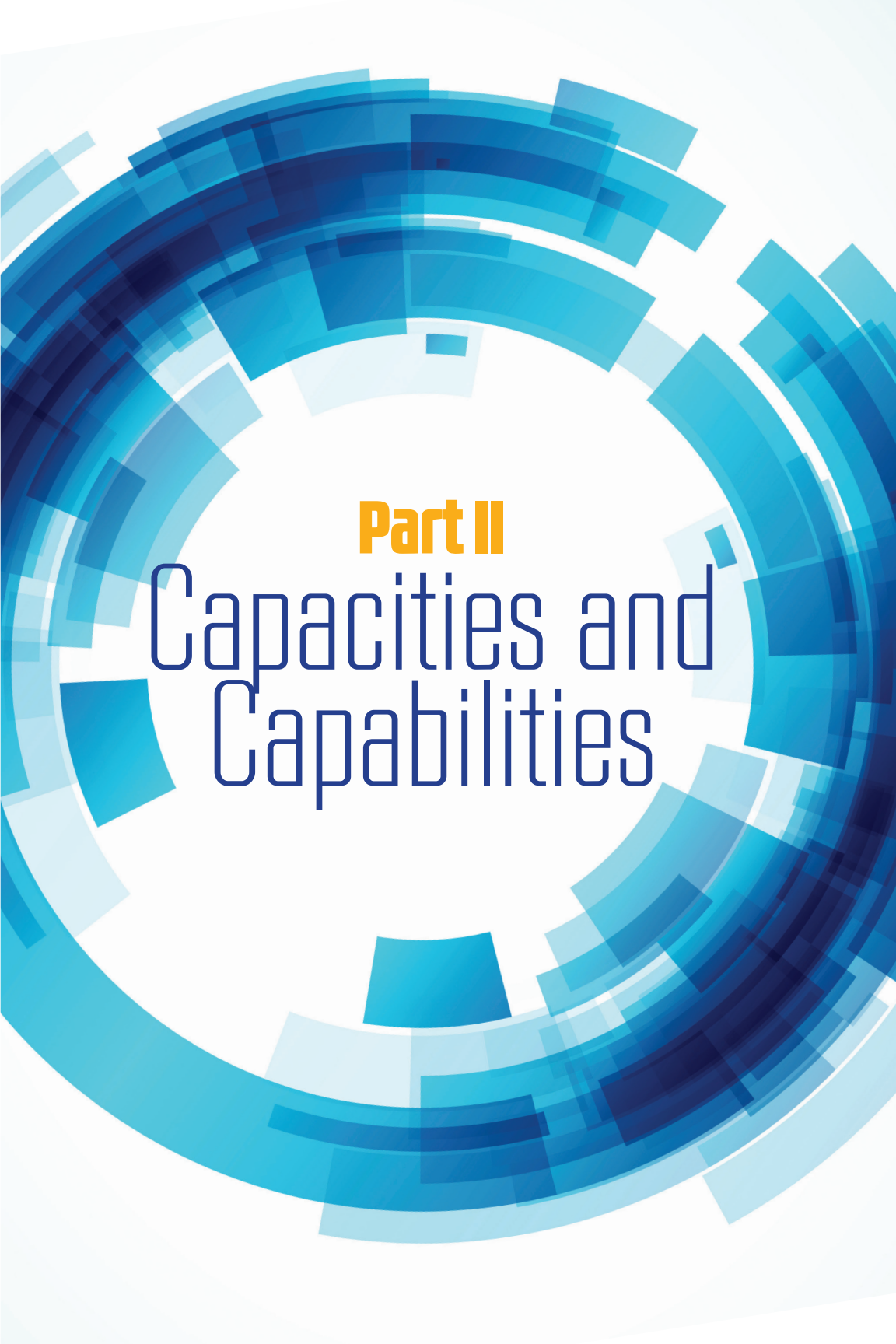
EvalPartners. 2016. *EvalAgenda 2020. Global Evaluation Agenda: 2016–2020*.

NEDA and DBM (National Economic Development Authority and Department of Budget and Management) 2015. "National Evaluation Policy Framework of the Philippines." Joint Memorandum Circular No. 2015-01. Republic of the Philippines.

PFDE (Parliamentarians Forum for Development Evaluation). 2015. "Status of National Evaluation Policies: Global Mapping Report."

UN (United Nations). 2015a. "Building Capacity for the Evaluation of Development Activities at the Country Level." A/RES/69/237. UN General Assembly, New York.

———. 2015b. "Transforming Our World: The 2030 Agenda for Sustainable Development." A/RES/70/1. UN General Assembly, New York.


Part II
Capacities and Capabilities

Chapter 4

Incorporating the Sustainable Development Goals in National Evaluation Capacity Development

Indran Naidoo and Ana Rosa Soares

Abstract. *This chapter discusses the efforts of the United Nations Development Programme to develop national evaluation capacities through the biannual conferences and actions promoted by its Independent Evaluation Office. The paper also looks into lessons learned from implementing the Millennium Development Goals that could be useful in evaluating the Sustainable Development Goals (SDGs). It further outlines directions and priorities for incorporating the SDGs in national evaluation capacity development efforts, building on what emerged from the consultations that took place during the Fourth International Conference on National Evaluation Capacities in Bangkok, which was jointly organized with the 2015 Global Assembly of the International Development Evaluation Association.*

At a United Nations (UN) summit in September 2015, member states adopted the 2030 Agenda for Sustainable Development, "a plan of action for people, planet and prosperity" that "seeks to strengthen universal

Indran Naidoo, United Nations Development Programme, indran.naidoo@undp.org; Ana Rosa Soares, United Nations Development Programme, ana.soares@undp.org.

peace in larger freedom" (UN 2015c). The 2030 Agenda commits all countries and various stakeholders to work together to "free the human race from the tyranny of poverty and want and to heal and secure our planet," to address inequality and injustice, and to ensure "that no one will be left behind." The 2030 Agenda presents an integrated plan of action with a vision and principles for transforming our world as set out in the results framework of 17 Sustainable Development Goals (SDGs) and 169 targets, with quantitative and qualitative objectives for the next 15 years; a means of implementation and global partnership; and a follow-up and review process. The follow-up and review framework calls for accountability to the people, national ownership, and country-led evaluative processes. Evaluation practice will provide an important means for raising the voice of stakeholders in this process to inform, support, measure, and assess whether development progress around the SDGs is relevant, sustainable, and equitable. Developing national evaluation capacities will be necessary in order to ensure that the follow-up and review process adds value to the implementation of the SDGs.

UNDP DEVELOPING NATIONAL EVALUATION CAPACITIES TO EVALUATE DEVELOPMENT GOALS

In 2015, the Independent Evaluation Office (IEO) of the United Nations Development Programme (UNDP) and the Royal Thai Government cohosted the Fourth International Conference on National Evaluation Capacities (NEC) in Bangkok, in collaboration with the UNDP Regional Bureau for Asia and the Pacific. The conference was organized jointly with the 2015 Global Assembly of the International Development Evaluation Association (IDEAS).

This conference was the fourth in a series of NEC conferences that have recognized UNDP for its distinct focus on supporting the governments with which UNDP works across the globe, as part of an IEO strategy to support the development of national evaluation capacities. The NEC conferences are held by UNDP every two years, each time in a different region, in partnership with a host government.

The model of the NEC conference has evolved over the years, drawing on lessons learned and emerging demands: but of essence is the focus on supporting governments to build their accountability capacities, of which evaluation is a key part. Over the last 10 years, the event has involved different partners, each of which has provided a particular emphasis and served to enrich the discussions, making it a key global evaluation event.

Each time around, support is focused on a specific region and uses different formats of exchange to promote commitment, cooperation, and action between and beyond senior government officials and to encourage the sharing of responsibility with other key players in the evaluation community. Much effort has also been invested in promoting continued engagement with past participants and institutions to deepen dialogues and foster continuity, partnership, and learning.

The NEC conferences are part of a broader architecture, in which the United Nations Evaluation Group (UNEG) plays a role by bringing together UN agencies and development partners to collaborate with each other. These

occasions serve to enhance the understanding and appreciation of evaluation as a powerful tool of public accountability and learning. They also help to advance the evaluation discourse globally and to align it with a strong call for cooperation in "building capacity for the evaluation of development activities at the country level" highlighted in the UN General Assembly Resolution 69/237 (UN 2015a). This resolution invites the entities of the UN development system, with the collaboration of national and international stakeholders, to support, upon request, efforts to further strengthen the capacity of member states for evaluation, in accordance with their national policies and priorities. Through UNEG, the UN promotes professional norms and standards for evaluation. In addition, UN entities and partners use evaluation to support accountability and program learning; to inform UN systemwide initiatives and emerging demands; and to benefit from and contribute to an enhanced global evaluation profession. The UN plays a particularly important role in enhancing national capacities to monitor and evaluate progress in poverty eradication and other internationally agreed-upon development goals, and therefore its key responsibility in supporting the SDGs.

UNDP's value added in evaluation has been its contribution as the secretariat and cochair of UNEG, but above all the strong independent mandate of its IEO to evaluate its contributions to development. IEO works with UNDP country offices and bureaus for national evaluation capacity development with a clear division of roles and responsibilities. IEO is responsible for developing guides and standards, and for promoting national evaluation capacity development through discussion forums such as the NEC conferences and associate initiatives of knowledge exchange. UNDP program units are further responsible for following up on the outcomes and partnerships of the conferences, and for supporting more specific programs and partnerships to develop national evaluation capacities in the medium to long term. This division of roles and responsibilities ensures IEO's independence and ability to credibly evaluate the results of UNDP's contributions.

UNDP believes that when appropriately tailored to national circumstances and priorities, the evaluation function can be an effective country-led vehicle for greater citizen accountability that can accelerate progress toward national SDG priorities, drawing on contributions from indigenous peoples, civil society, the private sector, and other stakeholders, including national parliamentarians (UNDP 2016a). Governments are engaged by UNDP to identify national evaluation partners, especially during the NEC conferences, but also as partners in certain evaluations. In these processes they develop their national evaluation capacities to promote greater accountability, learning, and development effectiveness in their countries.

UNDP has been supporting a range of activities to promote national evaluation capacity development, based on the UNDP definition of capacity development as an endogenous process through which individuals, organizations, and societies obtain, strengthen, and maintain the capabilities to set and achieve their own development objectives over time. Such a process can be described as country-owned if it is operated in a dynamic change process with reflection and learning, and if it is gradual, opportunistic, and adaptive to varying circumstances.

Using this frame of reference, the IEO NEC strategy approaches evaluation capacity development and the purpose and meaning of evaluation from a country, as opposed to a donor, perspective. In this regard, the purpose of evaluation goes beyond accountability to donors, to encompass public sector efficiency and accountability to the citizens of the country. The purpose of evaluation embraces other significant institutional and national goals for learning, and the development of innovation and social capital, knowledge assets, and the intellectual capital needed for growth, development, and contribution to global advancement.

UNDP has been successful in linking theory with practice, vision and ideals with realities, and in the methods we have sought to do the bridging. The Fourth NEC Conference provides a clear example of this approach with the theme "Blending Evaluation Principles with Development Practices to Change People's Lives." Together with IDEAS, this conference was an important opportunity to engage decision makers, academics, practitioners, and the UN community in global dialogue and advocacy around evaluation and the SDGs. More than 450 participants from 100 countries and from three key evaluation networks—UNEG, the Evaluation Cooperation Group of the Multilateral Development Banks, and the Evaluation Network of the Development Assistance Committee of the Organisation for Economic Co-operation and Development (OECD DAC)—participated in the conference, indicating the important role evaluation should play in shaping and contributing to the SDGs during the upcoming 15 years. The conferences are also an important opportunity for countries interested in South-South and South-North cooperation to find solutions together for challenges that have no ready-made answers. All can learn from previous experiences, such as what was advanced during the decade of the Millennium Development Goals (MDGs), that can be useful for the SDGs.

In Brazil in 2013, in a previous iteration of the conference, participants discussed solutions to challenges related to the independence, credibility, and use of evaluations. The conference produced 18 NEC commitments to further enhance national evaluation capacities, and encouraged creating greater accountability by setting goals for each country's NEC journey.[1] The 18 NEC commitments centered around four main strategies to build national evaluation capacities:

- Promote evaluation use through in-country and global advocacy
- Define and strengthen evaluation processes and methods
- Engage existing and new stakeholders in exchange and collaboration
- Explore options for different institutional structures for managing evaluations

In 2015, IEO published a baseline assessment of the countries that have participated in the NEC conferences series in order to document where each

[1] The NEC commitments can be found at http://www.nec2013.org/.

country stood, and in what direction they were moving regarding national evaluation capacities (IEO UNDP 2015c). This assessment found a variety of institutional settings and legal frameworks among the countries, reflecting a variety of government interests, political contexts, and national developmental stages. These granular aspects of national evaluation capacities are complex, and intrinsically linked to each country's development agenda: therefore they need to be taken into consideration and incorporated into the development of future evaluation agendas.

Over the years, one key lesson was learned through the NEC conferences and the process of promoting and implementing NEC commitments: without clear goals and the appropriate follow-up, governments and partners have a hard time focusing their attention on, and committing to the changes and long-term investments needed in order to build national evaluation capacities. The 18 NEC commitments served as a conversation starter for NEC participants to go back to their countries and reconsider their key national evaluation capacities and needs. These commitments were not necessarily the 18 initial commitments of the 2013 NEC conference, but included specific commitments that the countries agreed made more sense for their unique national contexts, and were therefore equally important.

Expanding on the 18 NEC commitments, in a global partnership effort for 2015, the International Year of Evaluation (EvalYear), the NEC 2015 conference focused on gathering information and commitments from participants to develop a new set of NEC commitments. The outcome was the Bangkok Declaration, a much expanded format that went beyond NEC and incorporated elements that also focused on the evaluation profession and global issues.[2] The declaration later contributed to another relevant document, the Global Evaluation Agenda, which was the first ever long-term global vision for evaluation.[3]

The Bangkok Declaration was a collective statement of all participants of the joint 2015 NEC conference and IDEAS Global Assembly: it is an expression of aspirations grounded in the community of practice of professional development evaluation. It is not legally binding on individuals or governments, but it seeks to capture key principles, give a sense of common purpose and understanding, and frame a vision of joint action in future support of individual, professional, and national evaluation capacity as countries shape their responses to the 2030 Agenda.

CONVERGENCE OF THE MDGS AND THE SUSTAINABLE DEVELOPMENT TRACKS FOR THE SDGS

Recognizing the intrinsic linkage between poverty eradication and sustainable development, during the General Assembly Special Event in September 2013, UN member states requested that the Open Working Group and the

[2] The Bangkok Declaration can be found at http://web.undp.org/evaluation/nec/nec-2015_declaration.shtml.

[3] The Global Evaluation Agenda 2016–2020 can be found at http://www.evalpartners.org/global-evaluation-agenda.

Committee of Experts on Sustainable Development Financing produce inputs for the post-2015 negotiations of the SDGs.

In August 2014, the Open Working Group submitted its proposal for a set of 17 SDGs, along with 169 associated targets. At the same time, the Committee of Financing Experts produced a set of recommendations on sustainable development financing. In December 2014, the Secretary-General submitted to UN member states his synthesis report, combining the intergovernmental proposals and the full range of inputs from both tracks.

UN member states agreed that the proposed SDGs would form the basis for intergovernmental negotiations of the post-2015 agenda with a text of the new agenda entitled "Transforming Our World: The 2030 Agenda for Sustainable Development," for adoption by the Post-2015 Summit held in New York September 25–27, 2015. The text included a declaration, 17 SDGs and 169 targets, and components on the means of implementation, the global partnership, and a follow-up and review process. The 2030 Agenda was structured around five "Ps"—people, planet, prosperity, peace, and partnership—and its set of 17 SDGs were officially adopted by the summit on September 25. The comprehensive nature of this new agenda has effectively reaffirmed this convergence, aligning the processes and the scope, and leading to a holistic approach to development.

Three other complementary processes ran alongside the Post-2015 process: one with a focus on disaster risk reduction, another on financing for development, and a third focused on climate change. The 2030 Agenda became an umbrella agreement for these other three agreements as well.

In March 2015, UN member states adopted the Sendai Framework for Disaster Risk Reduction (2015–2030) during the Third World Conference on Disaster Risk Reduction held in Japan. This framework, which is the result of several years of consultations and several months of intergovernmental negotiations, contains seven targets and four priorities for action to reduce negative impact, build resilience, and strengthen related international cooperation. The 2030 Agenda explicitly recognizes the importance of disaster risk reduction, and makes reference to the Sendai Framework and the need for development of holistic disaster risk management at all levels in its Goal 11.

In July 2015 in Addis Ababa, UN member states held the Third International Conference on Financing for Development, organized as a follow-up to the Monterrey Consensus and Doha Declaration. This conference led to an agreement entitled "Addis Ababa Action Agenda," which identifies the key action areas needed in order to provide the means and create an enabling environment for implementing the SDGs. The text of the adopted 2030 Agenda recognizes the concrete policies and actions agreed to in Addis Ababa as supporting, complementing, and contextualizing the means of implementation targets of the SDGs, and the Addis Ababa Action Agenda itself as an integral part of the 2030 Agenda for Sustainable Development.

The third complementary process is the UN Framework Convention on Climate Change (UNFCCC). The 21st Conference of the Parties (COP21) that was held in Paris in late 2015 featured negotiations toward the first universal, legally binding global agreement on climate change, now known as the Paris Agreement. The Paris Agreement, which is due to enter into force in 2020,

contains an action plan that could allow UN member states to limit global warming to well below 2°C, and aims to limit it to 1.5°C. In addition to actions aimed at the reduction of emissions, it also covers issues related to adaptation, support, loss and damage, and transparency and stocktaking. Such an agreement is explicitly mentioned in the proposed SDGs, and the UNFCCC is acknowledged as the primary forum for these negotiations.

APPLYING LESSONS LEARNED FROM THE MDGS TO THE SDGS

The post-2015 negotiations, and the work to design the SDGs, were led by member states from the outset. The outcome is the result of a consistent global participatory process in which representatives from countries, academia, civil society, and the private sector together formulated the SDGs. The SDGs contain much that critics said was missing in the MDGs.

One thing that was learned from the experience of the MDGs was the importance of reporting and monitoring. However, the 2030 SDG Agenda has a much wider scope than the largely social goals of the MDGs, and takes into greater consideration the need for economic, social, and environmental sustainability. It also recognizes the importance of peaceful societies. With the MDGs, the question was: What are the goals that are lagging the most, where are the gaps, and how can we fill them? With the SDGs, given the fact that the breadth of the 2030 Agenda implies a need to break down silos and adopt an integrated approach to development interventions, the question has become more evaluative. Reporting and monitoring are insufficient; countries need evaluations in order to answer this question: What are the actions required to accelerate progress across a broader range of interlinked goals? Addressing this question requires thinking through the connections and synergies across the goals, and pointing out how actions in one area affect other areas. Evaluative tools are also required to assess and manage trade-offs, and in this context "evaluation methods will need to determine whether the right choices were made to achieve possibly conflicting desirable outcomes, and how the different outcomes should be valued" (Heider 2015).

While much has been achieved during the MDG implementation period, a key criticism of the MDGs was that there was insufficient attention paid to generating evidence on achievements and particularly learning from challenges.[4] Much greater focus has been on monitoring and reporting, with many countries publishing national and also subnational MDG progress reports, while evaluation of which policies and interventions have worked and which have not were often only conducted at a later stage, and as part of designing MDG acceleration frameworks. MDG progress was largely tracked at the aggregate level, masking disparities in performance and disguising rising inequalities. In order to move forward in such a way as to ensure that no one is left behind, a better understanding of why and how certain policy choices and interventions affect different segments of society will be imperative. Recognizing that "only

[4] See EvalSDGs, http://www.unicef.org/evaluation/files/EvalSDG_Overview_Paper_8-12-15_1-pager.pdf.

by counting the uncounted can we reach the unreached" (UN 2015b), SDG targets should be met for all nations, peoples, and segments of society, and should "reach the furthest behind first," but only by evaluating trends, and contributing and hindering factors, can we assess whether progress has been or can be made relevant, sustainable, and equitable.

The 2030 Agenda is also a much more ambitious agenda than the MDGs, aspiring toward the goals of the elimination of poverty and universal access to benefits: this requires addressing the root causes of exclusion and deprivation, which are often deeply embedded in economic, social, and political marginalization. Another key lesson of the MDG implementation was that early strategic planning is important in laying the groundwork for long-term progress, because putting into place priority actions at an early stage can have multiplier effects on development outcomes (IEO UNDP 2015b). Targets associated with the MDGs were only shaped over time, while financing the MDGs was discussed in Monterrey two years after the Millennium Declaration. In contrast, the inclusion from the outset of a detailed results framework in the 2030 Agenda presents an opportunity for early action to link results and resources for results-based management.

From the beginning, the follow-up and review mechanism of the SDGs will also allow for early adjustments, course corrections, and enhanced results. In addition, the "MDG monitoring experience has clearly demonstrated that effective use of data can help galvanize development efforts, implement successful targeted interventions, track performance and improve accountability" (UN 2015b, 10). The MDG framework also strengthened the use of robust and reliable data for evidence-based decision making, with many countries integrating the MDGs into their national priorities and development strategies. Country ownership, leadership, and the participation of a wide range of stakeholders have been vital to ensure MDG progress and accountability.

PRIORITIES FOR NATIONAL EVALUATION CAPACITY DEVELOPMENT IN THE POST-2015 SDG DEVELOPMENT AGENDA

The 2030 Agenda states that "Governments have the primary responsibility for review, at the national, regional and global levels, in relation to progress made in meeting the goals and targets over the next fifteen years" (UN 2015c). With the explicit follow-up and review mechanism of the SDGs, countries will need to go beyond the usual monitoring and tracking of the MDGs, and tackle evaluations.[5]

Given the complexity of the SDGs—17 goals, 169 targets, and 230 indicators—the evaluation community has to be prepared to support an SDG platform for measurement, and for improving national evaluation capacities to contribute to accountability and learning. In addition, investment in qualitative

[5] The universal nature of the 2030 Agenda and the changing dynamics of development finance and development cooperation also present an opportunity to move from donor-driven to country-led evaluation.

assessment and careful design of national and international platforms and networks for dialogue, information sharing, and debate, with particular attention given to evidence provided by diverse domestic actors, may become central to achieving the SDGs.

The challenge of implementation points to the need for learning what works and what does not; which factors influence and hamper success; which aspects can risk sustainability, under which contexts; and how to break down silos and promote an integrated approach in order to achieve the most effective and efficient results. With that in mind, four overarching priorities emerged from the consultations that took place during the last NEC conference in Bangkok, building on the discussions from previous NEC conferences about independence, credibility, and the use of evaluations.

Promoting country-owned, country-led evaluations, with an emphasis on their use in influencing policies. One important priority is to respond to national circumstances, to support existing national systems and to avoid duplication of efforts and the famous "reinventing the wheel." Doing this entails a shift from donor-driven evaluations to country-owned evaluations and developing national evaluation capacities. This process should not be donor-driven but rather localized, contextualized, and culturally sensitized.

Critical for national ownership of evaluations is the need to raise the demand for evaluations, and not just focus on supply. A successful use of evaluations to inform policy, and to promote a change in mindsets within organizations and governments, can be used to advocate for a prominent role for evaluation in the implementation of the SDGs, for learning, and ultimately, to bring about improvement in people's lives.

Developing and strengthening evaluation process and methods. A second priority identified at the conference is developing new methods for evaluating progress toward, and the impact of, the SDGs. The 2030 Agenda is committed to developing broader measures of progress to complement gross domestic product. But how do we measure sustainability with the SDGs in mind? Environmental protection is only one of the means to achieve continued ecosystem services to mankind of clean air, water, healthy food, and freedom from disease. Sustainability requires an adaptive dynamic balance between the social, economic, and environmental domains. The SDGs require seeing economic growth, social inclusion, and environmental protection as mutually reinforcing. Whether our measurement and evaluation tools are sufficiently sophisticated to provide evidence on whether such a dynamic balance has been reached, or is within reach, and whether it is adaptive enough to change when necessary are great challenges. Methods that capture social inclusion and environmental protection need to be found in order to assess and evaluate sustainability.[6]

[6] There has been increasing interest from governments around the world in using innovative techniques to get better feedback from citizens on the effectiveness of their policies and programs, and to improve equity, sustainability, and accountability.

Engaging existing and new stakeholders in exchange and collaboration. A third priority identified was the importance of promoting more diverse partnerships and greater cooperation between governments, civil society, parliaments, and the private sector, in order to increase the awareness and use of evaluations. Traditional North-South aid models are playing a increasingly small role as private sector and national government resource flows increase, and the evaluation community advocates for more country-driven evaluations. There is a need for more dialogue in order to improve cooperation between the public and private sectors, to create networks and platforms for information and knowledge sharing, and to involve representatives of the private sector, parliamentarians, policymakers, legislators, and individual citizens.

The conference also stressed the importance of citizens as stakeholders, and the importance of raising awareness among citizens of the SDGs and the role of evaluation. There is growing awareness of the importance of people's engagement in monitoring and evaluation, and in accountability mechanisms.

Institutional structures for the evaluation of the SDGs. The NEC Conference revealed that we still have a long way to go in understanding how we integrate the evaluation of the SDGs into institutional structures. Almost every SDG is covered by national policy, so the question governments now face is how to monitor and evaluate all these policies and SDGs without duplicating and wasting resources. The holistic and integrative nature of the SDGs is not reflected in its structure and division into 17 goals. Governments may be tempted to divide out responsibility for the SDGs to respective line ministries, and the integrative perspective may be lost as a result. We need to sustain the discourse on the need to work on all SDGs also in evaluation. In an environment where the resources needed to deliver on the SDGs are scarce, evaluation will continue to gain prominence as a means of ensuring accountability for the use of those resources, and can help nations learn what works best under which contexts, in order to ensure effectiveness. The right institutional structures and national evaluation capacities will be key to the success of these processes.

CONCLUSION

The 2030 Agenda spells out the ideals and goals that will require evaluators and the development community to engage with in addressing a variety of interrelated, complex, and challenging issues, and to be competent at multiple levels in order to make significant contributions. If evaluators are to help give voice to people and countries in a global context where inequality persists at multiple levels, there is a need to start thinking about evaluation of the SDGs now, rather than as an afterthought. The SDGs contain a vision that combines a human capability approach to development with modern

UNDP, through its Innovation Fund, has implemented a number of prototypes with partners that harness technology to improve sustainability and accountability.

reconstructions of traditional economic models of growth. In responding to the SDGs focus on inequity, and in service to the principle of "no one left behind," the evaluation function can bring methodological validity as well as the legitimacy to empower people as effective evaluation processes help promote social action for development.

Achieving the SDGs depends on country-led evaluations that will produce evidence of whether the outcomes and impacts of policies, programs, and projects are equitable, relevant, and sustainable. Such evidence is useful not only in demonstrating public sector accountability, but also in focusing the attention of civil society and governments on enhancing learning, adaptive management, and innovation. Evaluation does not only identify "what works and what doesn't," or simply answer the question of whether we did or did we not achieve our objectives. Its real value is that it can be coupled to learning. For that, the learning and knowledge highlighted in evaluations needs to be used beyond simple accountability for strategic planning and adaptive management. Evaluation is a dynamic and ongoing process that continues to evolve, and is vital to support improving efforts, results, and development.

Therefore, supporting national evaluation capacity development is key to enabling mutual accountability among countries, and promoting learning to further the effective achievement of the SDGs, while ensuring that no one is left behind.

IEO is proud of the role UNDP has played in supporting development, and stands ready to work with partners to advance in supporting the development of national capacities for evaluating progress toward the SDGs.

In 2017, the Fifth International Conference on National Evaluation Capacities will take place. The evaluation community should be intensely engaged in discussing how to assess the equitability of SDG outcomes for marginalized populations; how to measure and evaluate new themes that are integral to the SDGs; and how to assess the effectiveness of integrated approaches, in order to understand what works best and under which contexts, to expedite progress toward and the achievement of the SDGs.

BIBLIOGRAPHY

Bester, Angela. 2015. "Capacity Development." Report prepared for the United Nations Department of Economic and Social Affairs 2016 Quadrennial Comprehensive Policy Review.

BetterEvaluation. n.d. "Steps in Planning and Managing an Evaluation."

Goldman, I., J.E. Mathe, C. Jacob, A. Hercules, M. Amisi, T. Buthelezi, et al. 2015. "Developing South Africa's National Evaluation Policy and System: First Lessons Learned." *African Evaluation Journal* 3 (1). http://dx.doi.org/10.4102/aej.v3i1.107.

Hauge, A. 2003. "The Development of Monitoring and Evaluation Capacities to Improve Government Performance in Uganda." Operations Evaluation Department ECD Working Paper No. 10, World Bank, Washington, D.C.

Heider, Caroline. 2015. "Evaluation beyond 2015: Implications of the SDGs for Evaluation." Blog post. Independent Evaluation Group, World Bank, September 15.

IEO UNDP (Independent Evaluation Office of the United Nations Development Programme). 2015a. "Blending Evaluation Principles with Development Practices to

Change People's Lives." Proceedings from the Fourth International Conference on National Evaluation Capacities. UNDP, New York.
———. 2015b. *Evaluation of the Role of UNDP in Supporting National Achievement of the Millennium Development Goals*. New York: UNDP.
———. 2015c. *Towards a Baseline Study: Insights on National Evaluation Capacities in 43 Countries*. New York: UNDP.
Mackay, K. 2007. *How to Build M&E Systems to Support Better Government*. Washington, DC: Independent Evaluation Group of the World Bank.
Naidoo, I. 2014. "The Contribution of Country Level Evaluation in Strengthening UNDP's Role in State-Building." African Development Bank Group, Abidjan.
Nilsson, Måns, Dave Griggs, Martin Visbeck, and Claudia Ringler. 2016. "A Draft Framework for Understanding SDG Interactions." International Council for Science.
OECD DAC (Organisation for Economic Development Development Assistance Committee). 1991. "DAC Principles for the Evaluation of Development Assistance." OECD, Paris.
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.
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.
Schwandt, Thomas, Zenda Ofir, Stefano D'Errico, Kassem El-Saddik, and Dorothy Lucks. 2016. "Realizing the SDGs by Reflecting on the Way(s) We Reason, Plan and Act: The Importance of Evaluative Thinking." IIED Briefing, International Institute for Environment and Development, London.
UN (United Nations). 2015a. "Building Capacity for the Evaluation of Development Activities at the Country Level." A/RES/69/237. UN General Assembly, New York.
———. 2015b. *The Millennium Development Goals Report 2015*. New York: UN.
———. 2015c. "Transforming Our World: The 2030 Agenda for Sustainable Development." A/RES/70/1. UN General Assembly, New York.
———. 2016a. "Interim United Nations Development Assistance Framework Guidance." UN, New York.
———. 2016b. "Ministerial Declaration of the High-Level Segment of the 2016 Session of the Economic and Social Council on the Annual Theme "Implementing the P-2015 Development Agenda: Moving from Commitments to Results." UN, New York.
———. 2016c. "Progress towards the Sustainable Development Goals." Report of the Secretary-General. UN, New York.
———. 2016d. "Report of Executive Summaries of Volunteer Countries on Voluntary National Reviews for the High-Level Political Forum." UN, New York.
———. 2016e. "Report of the High-Level Political Forum on Sustainable Development Convened under the Auspices of the Economic and Social Council at Its 2016 Session." UN, New York.
UNDP (United Nations Development Programme). 2009. *Capacity Development: A UNDP Primer*. New York: UNDP.
———. 2013. *Solutions Related to Challenges of Independence, Credibility and Use of Evaluation: Proceedings from the Third International Conference on National Evaluation Capacities*. New York: UNDP.
———. 2016a. "The UNDP Evaluation Policy." DP/2016/23. United Nations, New York.

———. 2016b. "UNDP Support to the Implementation of the 2030 Agenda for Sustainable Development." Policy and Programme Brief, UNDP, New York.

UNEG (United Nations Evaluation Group). 2012. "National Evaluation Capacity Development: Practical Tips on How to Strengthen National Evaluation Systems." UNEG, New York.

———. 2016. "Norms and Standards for Evaluation." UNEG, New York.

World Bank. 1999. *Evaluation Capacity Development: A Diagnostic Guide and Action Framework*. Washington, DC: World Bank.

———. 2012. "Guide to Evaluating Capacity Development Results." A collection of guidance notes to help development practitioners and evaluators assess capacity development efforts. World Bank, Washington, DC.

Chapter 5

Professionalizing Evaluation – A Golden Opportunity

Linda Morra Imas

Abstract. *This chapter considers the strong mandate for evaluation provided by the 2030 Agenda for Sustainable Development; and there are laudable efforts to strengthen national evaluation capacities in line with the mandate. However, there is still a lack of clarity on what such capacities look like in the context of the Sustainable Development Goals (SDGs), and what this means for evaluator competencies. Evaluators can view this situation as an opportunity to move forward on establishing core competencies for evaluators that reflect the SDGs, as well as to develop a clearer vision of national capacities to evaluate them.*

The good news is that the field of evaluation has a strong mandate, and is responding to it: on January 1, 2016, the 17 Sustainable Development Goals (SDGs) of the 2030 Agenda for Sustainable Development—which were adopted by world leaders in September 2015 at a historic UN Summit—officially came into force (UN 2015). Over the next 15 years, these new goals, which universally apply to all countries, will mobilize efforts to end all forms

Linda Morra Imas, International Program for Development Evaluation Training, lindag1000@gmail.com.

of poverty, fight inequalities, and tackle climate change, while ensuring that "no one is left behind."

Ensuring that no one is left behind means strengthening the voices and power of the most marginalized members of society—the disabled, the young, women, the poor—and challenging some of the most vested interests, such as those of energy producers. These are no small tasks, and there are some who believe the goals will not be realized by 2030.[1] But even to know how we are progressing toward these goals, evaluation of poverty and inequity, both within and across countries, is clearly needed.

The SDGs also provide an international mandate for evaluation, and continue to propel an evaluation capacity-strengthening movement that began with the Millennium Development Goals (MDGs). From these and from other ongoing influences, evaluation societies, known as voluntary organizations for professional evaluation (VOPEs) have sprung up at all levels—national, subnational, regional, and international—from just 15 in 1999 to 151 verified VOPEs by 2016.[2] Some have developed, or are in the process of developing, competency systems for their memberships, and a few have moved to credentialing or qualification systems. Both are ways of attempting to increase the professional status of evaluation, among other things.

But is this too little, and is it moving too slowly? The SDGs are multidimensional and tend to require mixed methods and complex, system-level insights. Without agreement on basic core competencies for the profession, and by letting "a thousand flowers bloom," are we missing a golden opportunity to advance the professionalization of evaluation? This chapter explores these issues.

A STRONG INTERNATIONAL MANDATE FOR EVALUATION

It cannot be overemphasized that the 17 SDGs, with their 169 targets, each with multiple indicators, are a first step in requiring all countries—not just the so-called developing countries—to set their own national agendas and strategies in collaboration with stakeholders. In this context, *all* countries are "developing" countries, facing common issues. The 2030 Agenda for Sustainable Development calls for follow-up and review processes that examine progress toward achieving the SDGs at the country, regional, and international levels. Follow-up and review processes are to be "rigorous and based on evidence, informed by country-led evaluations and data which is high-quality, accessible, timely, reliable and disaggregated by income, sex,

[1] See, e.g., Berliner (2015). Berliner and his team selected one key target for each of the 17 goals. Using projections from leading organizations, they predicted that not a single goal will be reached by 2030 if current trends continue.

[2] The International Organization for Cooperation in Evaluation (IOCE), the umbrella organization for evaluation organizations, lists the number of VOPEs as 188 in 2013 on its home page (ioce.net). Creating VOPEs may be easier than sustaining them. IOCE maintains an Excel database on VOPEs and, as of May 7, 2016, reports 151 verified VOPEs. This still represents huge growth from 1999.

age, race, ethnicity, migration status, disability and geographic location and other characteristics relevant in national contexts" (UN 2015). As indicated by EvalSDGs, a network of policy makers, institutions, and practitioners who advocate for evaluation of the SDGs, the initial focus has been necessarily on how to define and measure progress using indicators. However, it is acknowledged that measurement is not enough: "monitoring must be accompanied by evaluation that addresses the complexity of the SDGs and how they are achieved" (Schwandt et al. 2016).

WHAT DOES EVALUATION OF THE SDGS ENTAIL?

How to evaluate the SDGs is not so clear when it comes to the specifics, and relatively few have tried to articulate a clearer vision. Taking the lead, the International Institute for Environment and Development (IIED), in partnership with the network EvalSDGs, has been producing a series of briefs on this topic. At a basic level, the April 2016 brief indicates that evaluation uses monitoring data, but adds that it is "primarily concerned with how well implementation, outputs, and development outcomes were achieved, as well as with determining long-term development impact." It also says that "Evaluation asks why targets were achieved or not achieved and what can be done to improve the likely success of future initiatives" (Schwandt et al. 2016a, 2). But these generic statements about evaluation are followed with an emphasis on the interconnectivity of the SDGs that leads to the need to think about the evaluation of complex systems, rather than the evaluation of a single policy, program, or project.

More specifically, in terms of the SDGs, the brief notes that because the SDGs are interrelated in such complex ways, they present "wicked" problems for evaluation. For example, the aim of reducing income inequality (SDG 10) cannot be neatly separated from the aim of ensuring healthy lives and well-being (SDG 3). Initiatives to address such problems are themselves complex. They may involve "long causal chains with many intermediate outcomes, or outcomes that can only be understood using a 'causal package' approach that examines contributions from multiple interventions, contexts, or agencies…" (Schwandt et al. 2016a, 3). The implication for evaluation is that skills in new evaluation methodologies that draw on systems thinking may be needed. Another brief focuses on critical thinking skills as essential for conducting evaluations that analyze arguments, weigh evidence, and assess claims (Schwandt et al. 2016b). Being able to conduct country-led evaluations that assess sectoral, thematic, and holistic national policies, and that reflect whether a problem was correctly identified, the intended effects achieved, and whether unintended effects—either positive or negative—occurred is another part of the skills picture. Also, one must determine that outcomes and impacts are equitable, relevant, and sustainable. Because evaluations are not only at the national level but country led, stress is additionally placed on partnerships and evaluation capacity building.

Others—for example, Patton's "blue marble evaluation"—have put an emphasis on the need for evaluation from a global perspective, and the ability to evaluate adherence to principals such as human rights, gender equity,

inclusiveness, and sustainability, as well as the importance of maintaining the independence of evaluations (UNEG 2016c).[3] Yet others focus on the evaluation process itself, stressing the involvement of stakeholders, and qualities such as mindfulness, inclusiveness, and facilitation skills (Catsambas 2016).

Those who favor experimental and quasi-experimental approaches to evaluation have raised their voices to stress that information must be "evidence-based" and "rigorous." Some, such as the organization 3ie (International Initiative for Impact Evaluation), are promoting the use of findings from systematic reviews as a sounder base than individual studies for policies and programming (White 2015). This is because many lessons come from evaluation anecdotes, correlational data, and strong counterfactual data. These lessons need to be sorted out in order to use the most rigorous findings, and to see how the effects in one setting hold in other settings and over time.

From these writings, we could draw up an impressive and long list of skills that might be needed by those seeking to evaluate SDGs. But we also know that countries currently have not only widely different levels of evaluation capacity, but also wide variability in the availability of evaluation training. Variability in evaluation and in the quality of evaluation likely will be the main stories in efforts to evaluate the SDGs today, such that it might be difficult to have a coherent picture beyond indicator data. The ability to look at and address SDG issues from a global, regional, or even subnational perspective is limited. Professionalization of evaluation could provide an opportunity for a more level playing field by identifying global core competencies, and focusing training on building those competencies.

THE NEED FOR NATIONAL EVALUATION CAPACITY STRENGTHENING

Even at a basic level, it has been apparent for many years that many countries need assistance in developing their national evaluation capacities. The MDGs had building national evaluation capacity as an emphasis, and that emphasis is continued with the SDGs. For example, the World Bank, through its Independent Evaluation Group, fostered the development and implementation of the long-running International Program for Development Evaluation Training (IPDET), and it also partners in the Regional Centers for Learning on Evaluation and Results (CLEAR). With the SDGs, UN organizations are making increased efforts to support VOPEs. To illustrate, the International Organization for Cooperation in Evaluation (IOCE) has as its mission to support VOPEs in contributing to good governance, effective decision making, and strengthening the role of civil society. Under an agreement with UNICEF, IOCE

[3] At a lunch presentation at IPDET in June 2015, Patton floated the idea of "blue marble evaluators." He was received enthusiastically and based on the reaction developed a proposal to move forward the perspective of a complex, dynamic, and interconnected world system. The "blue marble" perspective means thinking globally, holistically, and systematically. Evaluators need special perspectives and competencies to engage and evaluate these global change efforts.

launched the EvalPartners Peer-to-Peer Program, which encouraged two or more VOPEs to form partnerships to strengthen their capacities. Thirty-two national and six regional VOPEs have formed 25 partnerships to design and implement Peer-to-Peer projects.[4]

The largest evaluation association today is the American Evaluation Association (AEA), with a membership of 7,100, followed by the Canadian Evaluation Society (CES), with about 1,800 members; the Latin American and Caribbean Network of Monitoring, Evaluation and Systematization (ReLAC) has about 1,600 members; the International Association for Development Evaluation (IDEAS) about 900 members; Australasia 860 members; and the European Evaluation Society (EES) about 550 members.[5] But then the numbers drop substantially. Other national and regional associations generally struggle, with from about 150 members to just a handful (UNDP 2015). Thus, efforts to support VOPEs, which are often easier to create than to sustain, continue to be needed.

The Global Evaluation Agenda, EvalAgenda 2020 (EvalPartners 2016), sets out four key areas where evaluation capacity needs to be strengthened if it is to fully realize its potential in supporting the new development agenda and beyond: an enabling environment for evaluation, institutional capacities for evaluation, the capabilities of individual evaluators, and the links between these three elements. And while all of them are important, it is the third area, concerning the capabilities of individual evaluators, that relates most directly to professionalism.

A strong **enabling environment** is described as one where all sectors of society understand and appreciate the value of evaluation; where evaluation is explicitly recognized or encouraged in national evaluation policies and other governance and regulatory instruments; where sufficient resources are allocated for evaluation at all levels; where evaluation findings are used; and where evaluation receives due recognition as a profession.

Strong **institutional capacities** involve strong VOPEs, as well as government agencies, civil society organizations, academia, and other institutions that generate and share relevant data to support evaluation.

Developing **individual capacities** for evaluation is relevant not only to evaluators but to commissioners and users of evaluation as well. The latter need sound understanding of the value of evaluation, and commitment to using evaluation findings and recommendations. In terms of individual capacities, the goals are to have:

[4] More information on the program and reports from its first projects can be found under P2P on the EvalPartners website, EvalPartners.org.

[5] Membership numbers are from the various association websites, accessed May 2017: AEA, http://eval.org; CES, https://evaluationcanada.ca; the Australasia Evaluation Society, http://aes.asn.au; IDEAS, http://ideas-global.org; and EES, http://europeanevaluation.org. ReLAC membership data are from Rodriguez-Bilella and Lucero (2016).

- Sufficient numbers of qualified evaluators, drawn from a diversity of relevant disciplines, who are available to conduct high-quality evaluations in all countries and all subject areas;
- Evaluators who have the knowledge, skills, and dispositions to make appropriate use of generally accepted evaluation principles, theories, methods, and approaches;
- Evaluators who have integrated the values discussed above, and are culturally sensitive.

But despite the vision presented in this agenda, a lack of clarity and agreement exists today on what it would look like to have national evaluation capacity in the context of the SDGs. UNDP, for example, has contracted for a study to try to further understand what this would entail, and what it would look like if a country had it.[6] The issues are many. For example, does national evaluation capacity mean having a sufficient number of qualified evaluators who are able to conduct high-quality evaluations in all relevant subject areas? Does it mean having the governmental capacity to do cross-cutting evaluations (i.e., evaluations that cross different ministries, so that interrelatedness can be addressed)? Does it mean having government policy makers and parliamentarians who are not only able to use evaluative information, but also consistently do so? Is it about having the capacity to conduct independent evaluations? Inclusive evaluations? Self-evaluations? Rigorous counterfactual evaluations? Complex systems evaluations? Systematic reviews? All the above and more?

THE CHALLENGES OF PROFESSIONALIZATION

It is difficult to build evaluation capacity in a profession that remains fractured, and lacking in agreement on how to define competencies for evaluators. Much has been written about evaluators and development evaluators still lacking professional status and visibility; evaluators not feeling recognized as professionals; and the belief that the lack of control over access to the ranks of evaluators resulting from lack of professional standards (and lack of enforcement of those standards) too often yields poor quality evaluation work.[7]

Is the Problem That Evaluation Is a Young Profession?

Evaluation is often referred to a young discipline. For example, Robert Picciotto, one of the gurus of development evaluation, calls it a "fledgling profession" (Picciotto 2015). Others have called it the "new kid on the block" among the social sciences. Its "youth" is often given as the reason for its struggles with professionalizing, and its general lack of agreement on evaluator

[6] Personal communication with Charles Lufthaus, Universalia, January 13, 2017.

[7] See, e.g., Altschuld and Engle (2015), King and Podems (2014), Morra Imas (2010), and Picciotto (2011).

competencies. But it is interesting to compare two young professions in the United States—school psychologists and evaluators—both of which have developed over similar time frames.

Rossi, Freeman, and Lipsey have credited the boom in evaluation to the demand for knowledge of results that accompanied large public expenditures for major programs in urban development and housing, education, occupational training, and preventive health services following World War II (Rossi, Freeman, and Lipsey 1999). They indicate that major commitments were also made during this time to international programs for family planning, health, and nutrition, and rural development. They conclude that by the end of the 1950s, evaluation research was commonplace.

According to Hogan, we can thank Russia's launch of Sputnik in 1957 and the ensuing space race for the discipline of evaluation (Hogan 2007). The National Defense Education Act poured money into new education projects and programs in math and science, and evaluations were funded to measure the success of the new curricula. The passage of the Elementary and Secondary Education Act of 1965 is commonly considered the birth of modern program evaluation in the United States, because it required evaluation, and thus helped evaluation to emerge as a profession. Once federal monies began to flow, universities began to offer courses in evaluation methods.[8] The Evaluation Research Society emerged in 1976, and so did evaluation journals such as *Educational Evaluation and Policy Analysis, New Directions for Program Evaluation,* and *Evaluation News*. By this time, evaluation clearly had emerged as a distinct specialty field within social science. In 1986, the Evaluation Research Society and the Evaluation Network merged to become the AEA.

Evaluation has been largely practice-based, with the people who do program evaluation coming from many different backgrounds, such as education, sociology, psychology, economics, social work, and public policy. Most evaluation degrees are still awarded out of departments such as education or psychology.

But as described by Stevahn et al. (2005), in the development of competencies for evaluators, the field of program evaluation has been decidedly less than can-do. They indicate that most fields recognized as professions, such as health care, teaching, counseling, and so on, have typically developed competencies for their practices by asking a group of distinguished practitioners—often on behalf of a professional organization—to first generate a category scheme and initial list of competencies, then to institute an expert review process to edit and refine them (Stevahn et al. 2005). The competencies are then made available to professionals in the field so that they can structure training programs for novice practitioners; continuing education programs for experienced professionals; and periodic reviews to update the competencies as theory, research, and practices evolve over time. But this has not happened in the field of program evaluation. Because there has been no

[8] For example, at Western Michigan, the University of Virginia, and the University of Illinois.

standardization, anyone can claim to be an evaluator, and can still do so to this day.

By contrast, the field of school psychologists was recognized as a division of the American Psychological Association (APA) in 1945.[9] But it was a nine-day conference in 1954 with 48 APA participants representing practitioners and trainers of school psychologists that began to advance the profession. Their task was to develop an official position on the roles, functions, and necessary training and credentialing of school psychologists. One of the goals of the conference was to define school psychologists, and the agreed definition was that school psychologists were psychologists who specialize in education, and have specific knowledge of the assessment and learning of all children. Participants at the conference felt that since school psychology is a specialty, individuals in the field should have a completed either a two-year graduate training program or a four-year doctoral program. They also felt that states should be encouraged to establish certification standards to ensure proper training. It was also decided that a practicum experience should be required, to help facilitate experiential knowledge within the field.

The National Association of School Psychologists (NASP) was formed in 1969, and in 1988-89 they moved to a national credentialing system. NASP is linked to state education agencies and to their credentialing boards. The NASP Standards for Training and Field Placement Programs in School Psychology identify the critical training experiences and competencies needed by candidates preparing for careers in school psychology. Ten domains are laid out, with standards in each domain. These standards serve to guide the design of school psychology graduate education. They provide a foundation for the recognition of programs that meet national quality standards through the NASP program approval process. The Standards for the Credentialing of School Psychologists are intended as a model for state education agencies or other state or local entities that employ school psychologists, and have the statutory authority to establish and regulate credentialing for school psychologists' titles and practices. Included are recommended criteria for initial credentialing, consisting of graduate coursework, practicums, and internship requirements, as well as recommendations for credential renewal.

While the world of school psychologists has its own debates, such as whether a doctorate should be required for entry into the profession, there is no debate about professionalism. Access to the profession is clearly highly controlled.

This is not to suggest that the credentialing of school psychologists should serve as a model for global evaluation, where access to training is not equal, and equity is a major concern. Additionally, evaluation is trans-sector: it does not have a history of state licensure, nor is accreditation a goal. But what we can conclude from looking at the development of school psychology as a profession is that the youth of a profession does not necessarily

[9] Much of this section is drawn from the National Association of School Psychologists (NASP) website and from Wikipedia.

correlate with professionalism, or hinder the degree of control over access to the profession through the setting of standards.

Letting a Thousand Flowers Bloom

Today any person or group can create their own set of evaluation competencies. And indeed, that is not only what is happening, but also what is being encouraged. The Global Evaluation Agenda summarizes the general view that "Given widely different cultural contexts and operating requirements, no standard blueprint approach would be appropriate for all VOPEs in all countries" (EvalPartners 2016, 24). Given the widespread contexts and sectors that evaluators cover, it is believed that each VOPE should design its own qualification system within a set of general, internationally accepted, guiding principles. Such principles are currently part of evaluator capabilities framework pilots being implemented by the EES and the United Kingdom Evaluation Society (UKES). These principles address voluntariness, autonomy, legitimacy, pluralism, transparency, equity, and quality assurance.[10]

As recognized by King and Stevahn, there are advantages to letting a thousand evaluation-competency framework "flowers" bloom (King and Stevahn 2015). This provides room for adaptation to unique contexts and content, and it may generate creativity and innovative ideas. The good principles referred to above would not restrict the bloom, and they can help guard against the possibility of elitism and continuing exclusion that is feared, especially in the context of developing countries, if formal qualifications are overemphasized over other indicators of competence, such as on-the-job training or relevant experience (Levin 2015).

But unless there is an agreed-on core of competencies that have some part in competency frameworks, it is hard to see how practitioners will be able to advance the argument that evaluation is indeed a discipline. Without an agreed-on core set of competencies that can be augmented by specialist and context-laden additions, it is difficult to see how the field of evaluation will be able to move toward increased professionalism. While there is value in diversity, more coherence is needed in order to advance the professionalism of evaluation. And without a core set of competencies, there is no sound basis for the exclusion of unqualified practitioners—a basic qualification for any profession.[11]

COMPARING SYSTEMS OF COMPETENCY

Today evaluator competency systems are rapidly being developed and adopted around the world not only in VOPEs, but also in organizations such as the United Nations Evaluation Group (UNEG 2016a, 2016b) and graduate

[10] These principles are more fully described in *EvalAgenda2020* (EvalPartners 2016), 84–86.

[11] See Wilcox and King (2014), 3, describing Worthen's nine criteria for judgments of evaluation's professional status.

school programs in universities, for example, Western Michigan University's doctoral program in evaluation. While only two credentialing systems exist—those of the CES (which is peer review–based) and the Japanese Evaluation Society (which is training course and exam–based)—the EES and the UKES are conducting pilots of peer review–based systems. IDEAS, the international VOPE, has been considering whether to also pilot a peer review–based system and/or to investigate other options. Given the more rapid growth of competency programs over credentialing programs, the next questions to consider are how fundamentally different these competency systems really are, and whether any of them have been developed in the context of the SDGs.

Table 5.1 compares the competency domains of five different international, regional, and national associations. The associations have separate ethical standards for their memberships, and some have separate standards for the commissioners of evaluations.

The CES has the only approved and operating peer review–based, credentialed evaluator system.[12] Launched in 2010, it is a voluntary designation, which means that the holder has provided evidence of the education and experience required by CES to be a competent evaluator. Competencies for Canadian evaluation practice (along with ethical standards) are the foundation for the credentialed evaluator program. Their 49 key competencies were placed into five competency domains: reflective practice, which focuses on fundamental norms and values, and awareness of one's evaluation expertise and need for growth; technical practice competencies, which focus on specialized aspects of evaluation; situational practice, which covers the application of evaluative thinking and the contextual circumstances in which evaluation skills are being applied; management practice competencies, which focus on the process of managing evaluations; and interpersonal practice competencies, which focus on "people skills."

To qualify for the designation, applicants must provide evidence of a graduate-level degree or certificate related to evaluation; evidence of two years (full-time equivalent) of evaluation-related work experience within the last 10 years; and indicators of education and/or experience related to 70 percent of the competencies in each of the five domains. As a peer review–based system, applications are reviewed by the CES credentialing board. There are special provisions for those who do not have a graduate degree or certificate, but they do carry additional out-of-pocket costs.

The Aotearoa New Zealand Evaluation Association (ANZEA) established its evaluator competencies in 2011, adding them to a system that included ethical guidelines and evaluation standards for undertaking and

[12] See Kuji-Shikatani (2015). Also, the Japan Evaluation Society has not only developed a competency framework, but has also implemented a certification program that is based on completion of a training program and a passing score on the related exam. As of early 2017, the Eurasian Alliance of National Associations, which includes evaluation associations from the Russian Federation, Ukraine, Kazakhstan, Kyrgyzstan, Tajikistan, and Armenia, was seeking to partner with academic institutions to further professionalization.

TABLE 5.1 **Comparison of competency/capability framework domains of five evaluation associations**

General description	CES	ANZEA	IDEAS	EES	AEA (draft)
Enhances/advances professional practice/ continuous learning	Reflective practice	Reflective practice and professional development	Promoting a culture of learning	Dispositions and attitudes	Professional domain
Knows and applies appropriate design and methods	Technical practice	Systematic evaluative inquiry	Professional foundations; monitoring; planning and design; conducting the evaluation	Evaluation knowledge: appreciates role played by evaluation in society; masters antecedents of evaluation quality; understands potential/ limits of evaluation instruments and tools	Methodological domain
Considers/analyzes issues, interests, context, and adapts practice	Situational practice	Contextual analysis and engagement	Professional foundations; monitoring; planning and design; conducting the evaluation	Evaluation knowledge: masters antecedents of evaluation quality	Context domain
Conducts/manages evaluations skillfully	Management practice	Evaluation project management and professional practice	Managing the evaluation	Professional practice: capacity to manage and deliver evaluations	Management domain
Inclusive/communicates effectively and respectfully/ negotiation	Interpersonal practice		Managing the evaluation; communicating evaluation findings	Professional practice: displays interpersonal skills	Interpersonal domain

commissioning evaluations (ANZEA 2011). Cutting across four competency domains and approximately 100 competencies are values and cultural competency. These are meant to ensure the inclusion and participation of indigenous groups and all marginalized subgroups. The first domain, contextual analysis and engagement, describes the abilities critical to undertaking analysis of the context; engaging with people as part of developing this understanding; and identifying the people, skills, knowledge, and experience needed to carry out the evaluation. The second domain, systematic evaluative inquiry, describes the knowledge, skills, and abilities required to undertake a systematic evaluative inquiry. The third domain, evaluative project management and professional evaluation practice covers the competencies needed to manage an evaluation in a professional manner. The fourth, reflective practice and professional development, includes competencies that support the development of the evaluation practitioner and the profession.

In 2012, IDEAS also approved a set of professional competencies. IDEAS is the only association for individual development evaluators, and it can proudly state that its competencies were developed by an international volunteer group from 40 countries, and ratified by the IDEAS membership, which spans at least 105 countries (Morra Imas 2010). Competencies were developed for those who conduct evaluations as well as for those who manage evaluations but do not conduct them directly. For those evaluators conducting evaluations, 25 competencies across six domains were identified. Additional supporting documentation breaks the competencies down even further. This is the only one of the five competency frameworks reviewed that identifies monitoring skills and capabilities as a domain. Additionally, unlike with the other competency systems, technical skills are broken into stages. Adapting to and knowing the context and the culture—what other systems refer to as situational practice or context—is incorporated by IDEAS into professional foundations; planning and design; and conducting the evaluation. Interpersonal practice also overlaps several domains. IDEAS treats all the competencies as core competencies.

The EES has attempted to make a distinction between a capabilities, or input-based, system such as their own, and ANZEA's and other outcome-based competency systems, such as the CES system. Outcome-based systems require evaluators to demonstrate their competencies: they are in effect testable, or results-based. Input-based systems are viewed as having a more deliberate learning orientation that focuses more on capabilities than on the demonstration of skills. While this distinction may not be quite clear at this point, and while the framework is being used to implement the Voluntary Evaluator Peer Review Pilot, it is certain that this is the only framework that stresses evaluator dispositions and attitudes, rather than reflective practice or professional development. Other models tend to make evaluator dispositions and attitudes a focus of separate ethical standards, or they include a competency on compliance with ethical standards.

After years of discussing competencies (Stevahn et al. 2005), in 2015, the AEA began to formally develop a set of competencies for its membership (Altschuld and Engle 2015). Several drafts have been produced and reviewed by the membership since that time, with review continuing into

September 2017. The specific goal of the AEA is to develop the general competencies that every evaluator or team of evaluators should have, regardless of context. As of September 2017, the competencies remained in draft form, with 44 competencies comprising the five domains. AEA continues to debate certification and credentialing.

SURPRISING COMMONALITIES

Looking across the five frameworks, at least in terms of the domains, there is more consistency than might be expected. While the specific words used may differ, five domains seem to be central to all of these systems: reflective practice; professional or methodological skills; contextual understanding; evaluation management; and interpersonal communication. A few domains are unique, such as the IDEAS identification of monitoring practice, or EES's dispositions and attitudes.

More variation seems to exist in the specification of competencies, and in their placement in domains. Some associations specify more methodological evaluator competencies, while others focus on competencies in the interpersonal domain. Still, this comparison suggests that identifying and agreeing on five or six core domains, and core competencies within each domain that skilled evaluators in various contexts should have, is a realizable goal. The core piece could then be added to and adapted for evaluators working in different specific contexts, but the core would remain the same.

None of these five systems have been developed with the SDGs or MDGs as a driving force. Most acknowledge that they will require review and revision from time to time in order to remain current. If the SDGs, and the commitment to them, are to be taken seriously, then they should provide the imperative for a review against the core competencies required by those who seek to evaluate the SDGs. It may be only a dream now, but the SDGs could provide the impetus to move forward toward the professionalization of evaluation in a more directed way.

CONSIDERING CORE COMPETENCIES IN THE CONTEXT OF THE SDGS

Thinking through core competencies in the context of the SDGs will likely be a three-step process. What is needed first is a review and agreement on the handful of core domains and concomitant competencies that are most important for skilled evaluators to have, whether they are specialists in empowerment evaluation or randomized designs, HIV/AIDS evaluation experts, or country evaluation specialists. These are the base competencies that those who call themselves evaluators should have, even though they may also specialize in particular evaluation methods, sectors, or countries. Like any set of competencies, these would not be set in stone, but would have provisions for periodic review and renewal as the field evolves. This step in itself would advance the professionalism of the field.

Second, as discussed earlier, is the envisioning of what it would look like at the national level, for countries to have the capacity to evaluate the

SDGs. The third would be to add to the core competencies and extend them to the SDG context as needed.

This will not be relevant for all evaluators. Not all evaluators work on an everyday basis in the direct context of the SDGs. But it would be a critically important paradigm for the many evaluators who are struggling with the SDGs. For example, working across sector boundaries might become a core SDG competency under the interpersonal domain. Mastery of complexity theory and systems approaches might be deemed important SDG competencies under professional or methodological practice; or new uses of technology for better data capture; or the use of big data. Some key questions will always need to be asked—for example, is this a necessary competency? or is it teachable?, as suggested by King and Stevahn (2015).

What are some of the advantages of professionalizing evaluation in the context of the SDGs? By creating agreement on what the evaluation of the SDGs entails, and delineating the core competencies necessary to undertake it, clear core standards would enable evaluators to work across geographical boundaries. This could level the playing field in the sense of knowing what the expectations are, and enabling the targeting of evaluation training programs where they are lacking, and where they are most urgently needed. Flexibility could be retained for noncore competencies, so that customization is still possible. Two of the largest benefits might be increased comparability of evaluation findings from shared methods and approaches, and an increased quality of evaluations.

How to start? That the convening power exists today to accomplish this goal is clear. EvalPartners, in partnership with IOCE, IDEAS, UNEG, and others, could continue a series of global multistakeholder consultations, whether face-to-face, virtual, or some combination of the two. This has already started to some extent, with the Third Global Evaluation Forum, held in April 2017 in Bishkek, Kyrgyz Republic, and organized by IOCE, EvalPartners, UNEG, the Global Parliamentarian Forum for Evaluation, the Kyrgyz Monitoring and Evaluation Network, and the Eurasian Evaluation Network, with the support of the Kyrgyz government. This forum brought together some 150 delegates, representing governments, parliaments, development partners, foundations, the private sector, universities, civil society, and the evaluation community, to advance implementation of the SDGs through review and implementation of Eval2020. Much could be gained by involving larger groups of evaluators, taking advantage of everyone's need to understand and advance evaluation of the SDGs; and with it, the opportunity for professionalization of the field. As SDG competencies are developed and agreed upon, ready access to high-quality professional training opportunities on SDG evaluation can follow.

REFERENCES

Altschuld, J.W., and M.E. Engle, eds. 2015. *Accreditation, Certification, and Credentialing: Relevant Concerns for U.S. Evaluators.* New Directions for Evaluation, No. 145. San Francisco: Jossey-Bass.

ANZEA (Aotearoa New Zealand Evaluation Association). 2011. *Evaluator Competencies.*

Berliner, Tom. 2015. "Can the SDGs be Achieved by 2030?" *Independent Research Forum* blog.

Catsambas, T.T. 2016. "The 5 Practices of SDG-Ready Evaluation." *Evaluation* blog.
EvalPartners. 2016. *EvalAgenda 2020. Global Evaluation Agenda: 2016–2020*.
Hogan, R.L. 2007. "The Historical Development of Program Evaluation: Exploring the Past and Present." *Online Journal of Workforce Education and Development* 11 (4): 1–14.
King, J.A., and D. Podems. 2014. "Introduction to Professionalizing Evaluation: A Global Perspective on Evaluator Competencies." *Canadian Journal of Program Evaluation* 28 (3): vii–xv.
King, J.A., and L. Stevahn. 2015. "Competencies for Program Evaluators in Light of Adaptive Action: What? So What? Now What?" In *Accreditation, Certification, and Credentialing: Relevant Concerns for U.S. Evaluators*, J.W. Altschuld and M.E. Engle, eds., 21–37. San Francisco: Jossey-Bass.
Kuji-Shikatani, K. 2015. "Credentialed Evaluator Designation Program, the Canadian Experience." In *Accreditation, Certification, and Credentialing: Relevant Concerns for U.S. Evaluators*, J.W. Altschuld and M.E. Engle, eds., 71–85. San Francisco: Jossey-Bass.
Levin, R.M. 2015. "Professionalizing Monitoring and Evaluation for Improved Performance and Integrity: Opportunities and Unintended Consequences." Presentation at the SAMEA Conference.
Morra Imas, L.G. 2010. "The Movement for Global Competencies for Development Evaluators." Paper presented at meeting of the Canadian Evaluation Society, Victoria, B.C.
Picciotto, Robert. 2011. "The Logic of Evaluation Professionalism." *Evaluation 17* (2): 165–80.
———. 2015. "What Is Evaluation?" Lunch presentation, International Program for Development Evaluation Training.
Rodríguez-Bilella, P., and M.A. Lucero. 2016. "Evaluation as a Global Phenomenon: The Development of Transnational Networks." In *The Future of Evaluation: Global Trends, New Challenges, Shared Perspectives*, R. Stockman and W. Meyer, eds., 66–80. London: Palgrave MacMillan.
Rossi, P.H., H.E. Freeman, and M.W. Lipsey. 1999. *Evaluation: A Systematic Approach*. 6th edition. Thousand Oaks: Sage Publications.
Schwandt, Thomas, Zenda Ofir, Dorothy Lucks, Kassem El-Saddick, and Stefano D'Errico. 2016a. "Evaluation: A Crucial Ingredient for SDG Success." IIED Briefing, International Institute for Environment and Development, London.
Schwandt, Thomas, Zenda Ofir, Stefano D'Errico, Kassem El-Saddik, and Dorothy Lucks. 2016b. "Realizing the SDGs by Reflecting on the Way(s) We Reason, Plan and Act: The Importance of Evaluative Thinking." IIED Briefing, International Institute for Environment and Development, London.
Stevahn, L., J.A. King, G. Ghere, and J. Minnema. 2005. "Establishing Essential Competencies for Program Evaluators." *American Journal of Evaluation* 26 (1): 43–59.
UN (United Nations). 2015. "Transforming Our World: The 2030 Agenda for Sustainable Development." A/RES/70/1. UN General Assembly, New York.
UNDP (United Nations Development Programme). 2015. *Towards a Baseline Study: Insights on National Evaluation Capacities in 43 Countries*. New York: UNDP.
UNEG (United Nations Evaluation Group). 2016a. *Evaluation Competency Framework*. NY: UNEG.
———. 2016b. "Professionalization of Evaluation." Concept Paper. UNEG, NY.
———. 2016c. "Evaluation in the SDG Era: Lessons, Challenges and Opportunities for UNEG." Working Paper. UNEG, NY.
White, Howard. 2015. "Toward Evidence-Informed Policies for Achieving the Sustainable Development Goals." *Evidence Matters* 3ie blog.

Wilcox, Y., and J.A. King. 2014. "A Professional Grounding and History of the Development and Formal Use of Evaluator Competencies." *Canadian Journal of Program Evaluation* 28 (3): 1–28.

Chapter 6

From Evaluation Capacity Building to Evaluation Capacity Development – A Paradigm Shift

Michele Tarsilla

Abstract. *In acknowledging the current limitations of contemporary evaluation capacity-building practice, and in an effort to promote an innovative and equity-focused contribution to the current discourse on evaluation capacity, this chapter suggests a new framework for conceptualizing, planning, implementing, and assessing the results of capacity development (as opposed to capacity building) in evaluation more effectively in the future. The first part of the chapter proposes a new definition of capacity in evaluation, and encourages readers to embrace and adopt the more encompassing term of "evaluation capacity development" as opposed to that "evaluation capacity building." The second part offers funders and planners an overview of those contextual and process-related factors that need to be taken into account in order to enhance the effectiveness of capacity development activities and programs. The third part highlights the specific and innovative contribution of the International Development Evaluation Association (IDEAS) to the ongoing discourse on evaluation capacity.*

Michele Tarsilla, International Development Evaluation Association, mitarsi@hotmail.com.

Often associated with the delivery of short-term training and technical assistance funded by international agencies, evaluation capacity building (ECB) has fallen short of its intended objectives in many different development and humanitarian contexts (Tarsilla 2014a). Despite the copious resources allocated by international development organizations to enhance the capacity of low- and middle-income countries to evaluate the performance and impact of their development programs and policies, most ECB activities on the ground have failed on multiple fronts. They have not been able to target a sufficiently large critical mass of individuals and organizations that could truly foster systemic change in the uptake and use of evaluation. The agencies that fund evaluation capacity-building programs in the Global South have not coordinated with each other as closely and systematically as they should have and, in so doing, have undermined their own ability to foster a more efficient and strategic use of resources. Finally, time and energy have been focused on the measurement of short-term effects, while the quest for long-term results has largely remained elusive.

Well aware of such weaknesses, and in response to the need expressed by many actors for the roll-out of more innovative ECB strategies, the International Development Evaluation Association (IDEAS) is promoting a paradigm shift in contemporary thinking and practice in the area of evaluation capacity development (ECD) (Tarsilla 2012). In particular, it is calling for a shift from the current focus on short-term training activities to the adoption of ECD strategies that are more contextually relevant and are better geared toward equitable, systemic and sustainable learning in evaluation.

EVALUATION CAPACITY BUILDING: KEY ISSUES AND LIMITATIONS

It is well understood that having individual practitioners and organizations' staff participate in two- or three-day workshops on evaluation theories and methods cannot foster dramatic change either in the quality of evaluation practice or the use of evaluation products. However, despite this, most ECB interventions, supported by both national and international funders in many different countries over the last two decades, have consisted mainly in the implementation of a finite number of activities. Such reification of capacity building, illustrated by the tendency to equate capacity building with training, as well as the tendency to implement evaluation workshops in a vacuum—that is, without accurate knowledge of how power and resources are distributed and contested at the local level—is indeed one of ECB's main weaknesses. This phenomenon, which has had egregious effects on the way international and national funders have planned and budgeted for in this area of development in the past, has been so prevalent that the meaning of ECB has been watered down, and its potential significantly compromised.

A second limitation in the way ECB has been conducted in the past is the dismissal of organizational processes in the planning, implementation, and evaluation of capacity-building efforts. While too much focus has been given to advancing the technical skills of individuals and organization staff, ECB planners and workshop facilitators have often failed to assess and act

upon the environment in which their target groups are operating. In particular, ECB planners and implementers have not systematically addressed the following as part of their capacity needs assessment:

- Environmental factors that influence the use of evaluation (e.g., the type of political system in place at the national and subnational levels; existing governance and accountability norms; and the degree of openness to accept failure/s and learn from them)
- Institutional or organizational processes that either benefit or hinder the evaluation function (e.g., lines of reporting and communication across different levels of a governmental agency, or data quality assurance regulations within the national statistics office)
- The quantity and quality of incentives available to conduct and use evaluation (e.g., the systematic publication of evaluation reports on a public portal, and/or the practice of organizations' executives to develop a management response in reaction to the recommendations included in an evaluation report)

Given inadequate understanding of ECD ecology, most funders and planners have failed to get many of the relevant actors from either the public or private sectors involved, either as partners or beneficiaries, in their past ECD efforts.[1] Evaluation training programs, for instance, are generally aimed at a limited number of evaluation technical officers from one or more organizations without the strategic involvement of their supervisors and directors. Furthermore, numerous ECB interventions provide participants with evaluation toolkits and checklists but they often dismiss the environmental factors that influence the adoption and use of such knowledge product—what I have defined as the "political economy" of ECD.

A third limitation has been the more or less inadvertent perpetuation of the old development paradigm, according to which donors' needs and interests prevail over anybody else's. For more than a decade, ECB activities have been geared toward increasing the level of knowledge, skills, and attitudes of project staff in the field with the primary de facto objective to enhance regular results reporting to funders (e.g., on a quarterly basis). What has been particularly fallacious is the assumption that retrofitting existing practices within established boundaries identified by donors would enable empowerment and social change. Unfortunately, this donor-centric strategy, which I tend to classify as "functional evaluation capacity building," or F-ECB, not ECD, has gradually become the norm (Tarsilla 2014a). In one case, four small cultural organizations that I worked with in the Democratic Republic of Congo had strategic objectives that, as spelled out in the plans and logical frameworks formulated for them by an international funder, were out of sync not only with their organizations' own vision, but also with their country's

[1] Public sector actors would include, among others, staff in ministries and members of parliaments. Private sector actors would include, among others, training institutes and consulting firms.

national cultural policies. It was very disheartening to discover that—according to the contract signed with the funder—I needed to enhance the capacity of these four organizations to measure the attainment of their objectives, which had been imposed from the outside, by using a number of indicators that the organizations in question did not really understand, or assign any credit to. In particular, it was very difficult to talk to them about logic models, theory of change, and rigorous evaluation designs, especially given the fact that the impact indicators showing at the top of the logical framework template—which had been distributed by the funder—rested on the assumption that grantees would be in a position to effectively measure the extent to which some of their activities, which were targeting less than 20 participants per year, had contributed to improved attainment of two of the Millennium Development Goals in the whole country. Unfortunately, this example shows once again that ECB programs often provide participants with the knowledge and tools that facilitate timely reporting to funding agencies, but rarely foster true organizational learning and increased results-oriented agency.

HOW TO OVERCOME EXISTING LIMITATIONS WITH THE NEW DEFINITION OF CAPACITY IN EVALUATION: FROM EVALUATION CAPACITY BUILDING (ECB) TO EVALUATION CAPACITY DEVELOPMENT(ECD)

ECB in International Development: Key Assumptions and Real-World Considerations

The evaluation policies and strategic evaluation plans currently in use among several development agencies around the world are predicated on the assumption that international development evaluation serves two primary functions (GIZ 2013; Norad 2006; Sida 2007; UNESCO 2015; USAID 2011). The first is to enhance the accountability of those who manage and implement international development projects, especially vis-à-vis their respective funders and expected beneficiaries (Wiesner 1997). The second is to foster learning among those who commission, manage, conduct, and use evaluation, on what works well and what needs to be improved in international development projects and programs (Argyris and Schön 1996; Bamberger 2009; Pasteur 2006; Rist, Boily, and Martin 2011; Solomon and Chowdhury 2002). Based on such assumptions, any activity aimed at strengthening the evaluation function—locally, nationally, or globally—should, therefore, be able to contribute to strengthening both the performance and the effectiveness of international development projects in a variety of countries. However, this is easier said than done.

Typically, funders and international development agencies attempt to strengthen the evaluation function by developing the technical skills of local development practitioners. However, such strategies do not always translate into stronger development effectiveness. One reason for this is the lack of a genuine evaluative culture—that is, the systematic conduct of evaluation, and the use of findings for decision making—which often results from the limited ability of ECB to foster ownership and inclusiveness of evaluation processes.

The scenario, however, is not as bleak as it would first appear. There are several examples of countries that have put monitoring and evaluation (M&E) systems into place that are both prospering and serving accountability and learning purposes. Chile, Colombia, Malaysia, Mexico, South Africa, and Sri Lanka provide good illustrations of how the creation of a supportive environment to enhance the evaluation functions at both the organizational level (Stevenson et al. 2002; Wijayatilake 2011) and the institutional level can foster the development of a strong national evaluative culture (Boyle, Lemaire, and Rist 1999; Mayne 2008; Trochim 2006). One feature that all of these countries have in common is the buy-in of different stakeholder groups both within and outside the national government into the discourse on evaluation, and their subsequent involvement in all of the related processes. Such success stories, though, have not been capitalized on in the international development arena as frequently as they should, or could, have been. The result is that the pursuit of non-inclusive targeting strategies has hindered the success of ECD interventions in many countries around the world. The simultaneous involvement of actors operating both within and outside of national governments as part of an ECB intervention has been very rare: this is also a result of the rigidity featured by the mission and the scope of work among the majority of funding agencies. On the one hand, for over a decade bilateral donors and philanthropic foundations have been able to fund a plethora of initiatives and programs specifically aimed at strengthening the knowledge and skills of stakeholders supplying evaluation services (e.g., representatives from academia, the private sector, and nongovernmental organizations) (OECD 2006). On the other hand, multilateral agencies, such as the United Nations and the World Bank, have allocated the largest share of ECB resources to national governments; that is, the agents that for the most part demand evaluation services.

ECB or ECD: Does the Terminology Really Matter?

Far from becoming rhetoric in a purely academic exercise, an exploration of the language used to describe ECB would be particularly useful to enhance future programming in this area. In particular, reviewing the central attributes, modalities of implementation, and evaluative variables of ECB would be beneficial. The same is true for another popular term used in the international development arena: evaluation capacity development. An in-depth review of ECD appears even more relevant than that of ECB, given that capacity development—in evaluation as well as in a variety of other fields—is not the only "missing link in development" (World Bank 2005, 24). Even more importantly, it is part of the overall goal of development cooperation (Fukuyama 2004).

A certain confusion or lack of consensus exists over the meaning of both terms, and has been attested to by a stream of peer-reviewed literature produced by evaluation scholars. Among the most recent contributions on this topic, Bohni and Attström's (2011) appears particularly relevant. According to these Danish authors, more serious reflection and debate on the distinction between ECD and ECB is needed, as it would allow the addressing of four main issues affecting the practice of evaluation in a number of

countries: the widespread conceptual pluralism in the area of ECB and ECD; the increased number of discordant opinions regarding the purposes of ECB; the lack of a comprehensive empirical base for most ECB and ECD models; and the relatively greater focus on the approaches implemented in tackling ECB rather than ECD.

When referring to capacity building, Morgan, one of the most prolific authors on capacity, has defined it as:

> ...a risky, murky, messy business, with unpredictable and unquantifiable outcomes, uncertain methodologies, contested objectives, many unintended consequences, little credit to its champions and long time lags. (Morgan 1998, 6).

Likewise, in defining capacity development, Lusthaus, one of the most well-respected Canadian experts in institutional evaluation and change, has described it as follows:

> ...a concept still in its infancy. Its definition is still forming. Research describing how people use the concept is sparse. So is research, which tests its assumptions and predicts its consequences. There are few evaluations of projects that are claiming to use approaches to capacity development. (Lusthaus et al. 2002, 34)

The discourse on capacity development (including over its definitions) has continued over the years. However, it has traditionally been dominated by the voices of northern scholars. Only recently have researchers and academic institutions in the Global South become engaged in this area of work. Among some of the most recognized actors who have contributed to advancing capacity development-related definitions are the following:

- The African Union and the New Partnership for Africa's Development, which in 2010 introduced an African Capacity Development Strategic Framework (AU and NEPAD 2012)
- The African Capacity Building Foundation (ACBF 2016)
- The Capacity Development for Development Effectiveness (CDDE) forum in Asia[2]
- The Task Team on South-South Co-operation (TT-SSC) in Latin America[3]

The Main Attributes of ECD

Rather than being a purely semantic issue, the distinction between ECB and ECD appears all the more relevant due to the unique political and ideological connotations ascribed to each of the two terms. As confirmed by the

[2] http://www.southsouthcases.info/casosasia/caso_14.php.

[3] https://www.oecd.org/dac/effectiveness/46080462.pdf.

increasing use of the term "evaluation capacity development" in some of the evaluation offices of international organizations, such as the World Bank and GIZ, ECD is characterized more *consistently* and *intentionally* by a stronger emphasis on inclusiveness, flexibility, development result focus, and context responsiveness to an already existing evaluation capacity. An Institute of International Development Studies report commissioned by UNESCO draws an important distinction between capacity building and capacity development that appears particularly useful for the discussion in this chapter (Ortiz and Taylor 2009). When applied to the evaluation context, the report seems to corroborate and amplify the differences between ECB and ECD. As discussed in one of the report's most salient passages, capacity building and capacity development are not described as simply different terms, but rather as two opposite development paradigms:

> Much of the capacity development literature stresses the fact that development is already happening before the arrival of any project, donor, program or initiative, and not to recognize this as an irresponsible error and ultimately a precursor to an ineffective use of resources. Too many donors and executing agencies are determined that their projects be executed in any event, yet when those projects are severely out of tune with the development processes already in motion, they are likely to fail. They fail because:
>
> a) Capacity development programming that does not recognize development in motion is quite literally a foreign object; that is, it pushes ideas that aren't likely to take hold because they are out of step with local realities;
>
> b) They do not build on momentum; that is, positive development initiatives and processes already in motion;
>
> c) The motivation needed to take forward a strategy that does not fit will in turn require a push strategy to convince people to carry it out. Even when the appropriate incentives are in place, true motivation will be dubious because participation will likely be led by the possibility of short-term gain. The fundamentals required for sustainability will be lacking and therefore the project activities and desired behavior changes are unlikely to develop deep roots" (Ortiz and Taylor 2009, 26)

Based on such foundational work, and following global research on the understanding of ECB and ECD among evaluation planners, managers, and practitioners in many different countries, the need for a new definition of ECB and ECD becomes apparent (Tarsilla 2012). However, rather than seeing the two terms as opposite, the two definitions that I came to develop after my exchanges with more than 150 practitioners in over 40 countries situate themselves along an ECB-ECD continuum, where ECB generally accounts for one of the preliminary phases of a broader and long-term ECD strategy.

Evaluation Capacity Building: A New Definition

The new definition of ECB, which was developed toward the end of a long series of consultations with practitioners around the world, reads as follows:

> A necessary (but not sufficient) condition for ECD to take place. ECB mainly consists of a vast array of trainings and coaching activities (some of which are short-term in nature) aimed at building capacity, especially where capacity is either very low or thought not be in place yet, among a discrete number of individuals working either for or within organizations and/or institutions that develop, commission, manage, conduct and/or use evaluation. Although it is an integral component of most national and international development projects today, ECB has often been viewed (especially outside of the United States) as a relatively limited accountability-driven tactic rather than a full-fledged strategy aimed at attaining organizational learning as well as other developmental objectives. As a result, ECB scope and modalities of delivery have often been considered too narrow. (Tarsilla 2012)

Evaluation Capacity Development: A New Definition

ECD, on the other hand, is defined as:

> A process consisting in both the integrated enhancement and maintenance over time of: (a) Individuals' knowledge, skills and attitudes; (b) Organizations' capabilities; and (c) Institutions' readiness; toward contextually relevant planning, management, implementation, and use of evaluation at any level-global, regional, national or sub-national. More specifically, ECD is aimed at both individual and collective transformational learning in the pursuit of three primary goals: strengthening the technical quality and ownership of national evaluation processes; enhancing the local authenticity and cultural appropriateness of evaluation approaches, methods and tools used in-country; and increasing the use of evaluation findings as a way to improve development interventions in a variety of sectors. (Tarsilla 2012)

In order for ECD to be successful, it is critical that ECD strategies be implemented either in a simultaneous, or an intentionally sequenced fashion. ECD-savvy strategies (such as the ones adopted by IDEAS) are specifically aimed at promoting the conditions that support ECD among a variety of actors operating in two different spheres (both within and outside of national government), and characterized by different functions (operational, and policy or decision making) and roles (both consumers and providers of evaluation). ECD strategies consist of a combination of short, medium, and long-term activities (including training, mentoring, coaching, peer exchange, and the creation of evaluation units). Otherwise, ECD appears to be a systemic and adaptive process rather than the combination of stand-alone activities aimed at enhancing capacity at the individual, organizational, or institutional levels. As stressed by the new definition provided above, ECD emerges as an endogenous process that builds upon existing levels of knowledge, skills, and attitudes (individual), capabilities (organizational), and readiness (institutional) either simultaneously or sequentially, and in a variety of contexts (global, regional, national, and subnational), as opposed to building from scratch.

ENVIRONMENTAL AND PROCESS-RELATED FACTORS INFLUENCING THE OUTCOMES OF CAPACITY DEVELOPMENT (ECD) ACTIVITIES AND PROGRAMS

ECD: Gaps in Targeting and Possible Solutions

Activities aimed at strengthening technical capacity within national ministries and central planning agencies (Compton, Baizerman, and Stockdill 2002) have enhanced the knowledge of evaluation within the governmental sphere, but have not necessarily contributed to the development of skills needed to either formulate key evaluation questions or use evaluation findings (Bamberger 2009; OECD 2006). Furthermore, activities aimed at enhancing national evaluation capacity have rarely been customized to the specific functions (operational or strategic) and roles (commissioners, implementers, policy makers) of individual officers operating within the government, and have instead favored the implementation of the same standardized approach at several levels within the government, as if it were a monolithic block. Addressing the limitations of current ECB targeting is all the more relevant, as the currently biased allocation of funding between governmental and non-governmental actors has three primary consequences.

First, it has hindered the mainstreaming of evaluation at a more systemic level, as predicated by a number of studies, including a recent work funded by the U.K. Department for International Development (Gaventa and Barrett 2010). Change happens through multiple types of citizen engagement: not only through formal governance processes, even participatory ones, but also through associations and social movements that are not created by the state. Strengthening these broader processes of social change and their interactions can, in turn, create opportunities for state reformers to respond to demands, build external alliances, and contribute to state responsiveness (Gaventa and Barrett 2010).

Second, the identification of individual evaluation champions within host governments that are characterized by high employee turnover has not always contributed to either the uptake of an evaluation culture or the sustainable promotion of the use of evaluation findings in other sectors (Lennie 2005). There are certainly some good examples of the contributions of national evaluation champions. However, the tendency among politicians to cater to their constituencies' needs and interests regardless of what the available evidence suggests confronts ECD planners and implementers with a real hurdle to overcome.

Third, the greater focus on the demand for evaluation, which some authors critically refer to as "elite domination" (Fung 2003, 340), has ignored the potential contribution of evaluation "suppliers" (e.g., national evaluators), and has not sufficiently leveraged their wealth of knowledge and practical experience during the undertaking of evaluations. That notwithstanding, ECD targeting is already gradually evolving, as attested to by the support provided by such initiatives as EvalPartners and the strengthening of voluntary organizations of professionals in evaluation (VOPEs) over the last five years.

ECD Processes: What's New Compared to Past ECB Practices?

As per the new definition of ECD, the focus of ECD is not on either activities or products (e.g., a training on mixed methods, or the timely submission of a midterm review to the project funder), but rather on processes, interactions, incentives, leadership, organizational learning, and organizational development. Furthermore, ECD is characterized as a particularly inclusive process that is able to respond to the continually emerging needs and interests not simply of individuals, organizations, or institutions; but rather of individuals situated within organizations, and institutions interacting with each other, both in the governmental and nongovernmental spheres. Likewise, rather than resting upon linear and mechanistic planning, ECD is understood and defined as a process grounded in both a realistic understanding of the world's complexity, and the need to adopt more flexible and iterative planning processes. However, my analysis of contemporary ECB practice shows that the latter has focused on developing individual technical skills—how to write evaluation terms of reference, or how to develop sampling strategies—rather than on developing organization-level capabilities and institutional readiness, based on relatively linear and results-based planning processes (Tarsilla 2012). Moreover, in contrast to the descriptions of capacity development as an incessant endogenous process in the peer-reviewed literature, the way the term ECB has been understood suggests that it rests on the main assumption that in-country capacity is static (you either have it or you don't), and that targeted interventions, often funded from external development partners, are the most effective. As a result, ECB does seem to fail to recognize the inherent institutional processes and social dynamics of the settings where its activities are being implemented.

Other Contextual Factors Influencing the Outcomes of ECD Activities and Programs

If one takes the organization as the main unit of analysis of any reflection on ECD,[4] then it is relevant to consider those unique distinct organizational features (its structures and processes) that are likely to affect the outcomes of an ECD program. The adequate consideration of organizational infrastructures and underlying dynamics is, therefore, critical to the success of any ECD strategy. For each of the relevant factors identified in the left column of table 6.1, a series of ECD strategies are recommended in the right column.

Based on a review of the organizational factors listed in the table, a key conclusion is that, despite the size of one's own organization, the planning of any ECD program cannot overlook the context in which that organization operates. As harmonization, relevance, and ownership are some of the principles that any sound ECD program should feature, it could be useful to link organizations targeted by ECB efforts with each other as if they were

[4] Most individual evaluation practitioners work either within or for one or more organizations.

partners and not simply grantees sharing the same donor, so as to allow an alignment of their internal M&E systems.

It is important to note that what is being advocated here is not the development of a cookie-cutter approach, whereby individual organizations share exactly the same set of indicators or evaluation strategies: after all, creativity and flexibility to adjust to emerging changes are two typical features of successful ECD programs. By promoting ECD alignment, what is being suggested instead is to encourage organizations to find common and cost-effective solutions to their information needs and operational questions. During the planning of an ECD program, for instance, each organization with a vested interest in ECD could look at what tools similar organizations have used to measure a certain construct that they are interested in measuring and—for the sake of avoiding replication and the wasting of resources—might adopt those very same existing tools. Besides freeing up resources for the development of other tools, or the set-up of an information management system within one's own organization, for example, ECD alignment would foster the use of identical indicators and, as a result of increased data aggregation, the availability of data that could then inform ECD-related decision and policy making. This is even more relevant if such alignment not only takes place locally, but also at the provincial, district, regional, and national levels.

Likewise, it is important that ECD strategies acknowledge more effectively what the real interests and needs of any organization are, independently of the donors' interests and needs. With that in mind, the following key recommendations should be taken into account during the development of a new ECD strategy:

- Although it is tempting to introduce radical changes within the realm of organizational practices when ECD programs are being implemented, it is critical that donors and contractors implementing ECD strategies on the ground recognize the speed of local organizations to "digest" new evaluation methods and tools.
- Development organizations should understand that assigning a prominent role to funders' evaluation requirements and needs, and building upon them to design an ECD program, is a conventional form of evaluation capacity building. For evaluation capacity development to occur, the centrality of the organizations (each with their own interests and needs) within the system where they are operating needs to be recognized.
- A broader and more systemic targeting of ECD is needed. Two new possible scenarios could be envisaged. On the one hand, funders and implementing organizations should promote the conduct of evaluation awareness-raising among actors who, despite not being directly targeted by the ECD intervention, still gravitate within the system where the latter is being implemented. On the other hand, funders and implementing partners should ensure more active involvement of the entities working in both the private and public sector anytime a large-scale ECD program sponsored by a national

TABLE 6.1 **How an organizational diagnostic could contribute to effective ECD programming**

Key organizational identity traits	Considerations to integrate into an ECD strategy
Historical development (informal development, formalization, expansion, regional consolidation, transition to national ownership, stagnation/implosion, self-reflection, nominal/effective revitalization)	▪ Learn about the history of the organization(s) targeted by your ECD strategy (e.g., key milestones, original founders, individuals promoting and/or challenging transformations within the organizations in question) ▪ Make sure to conduct a stakeholders' mapping, as well as a political economy analysis to assess how power and resources are distributed in the context of the organization(s) where the capacity is expected to develop further as a result of your intervention
Organizational development phase (pioneer, differentiated, integrated) (Ubels 2010)	▪ Assess the extent to which the staff of the organization(s) targeted by your ECD strategy mainly rely on one only leader to find their direction; or if they are driven by clearly articulated organizational policies and job descriptions ▪ In the latter case, explore to what extent evaluation tasks and responsibilities could be integrated into the existing processes and procedures
Membership diversity (low, medium, high)	▪ Learn about the staff making up the organization(s) targeted by your ECD strategy ▪ Assign staff members to different groups according to their specific role and responsibilities (e.g., top leaders and decision makers; managers; technical officers) ▪ Be sure to combine activities aimed at the whole staff with more specific strategies tailored to the needs and interests of each one of the identified subgroups ▪ Try to learn about past training programs offered to each of the identified subgroups so as to build on the examples and language used in past training events
Compliance with internal government rules (low, medium, high)	▪ Look for any organizational and performance audits that have been conducted in relation to the organization(s) targeted by the ECD strategy ▪ Identify organizational deficiencies observed in the past with respect to conformity with the established organizational procedures

(continued)

TABLE 6.1 *(continued)*

Key organizational identity traits	Considerations to integrate into an ECD strategy
Degree of internal networking (low, medium, high)	▪ Reconstruct the lines of communication and reporting among the staff of the organization(s) targeted by your ECD strategy ▪ Liaise with the management information system officer (if available) in the organization(s) targeted by your ECD strategy, to better understand the flow of information, both bottom-up and top-bottom
Resilience (low, medium, high)	▪ Understand how the organization(s) targeted by your ECD strategy has been able to respond to external challenges (even those threatening the survival of the organizations in question) in the past ▪ Highlight the self-defense mechanisms, values, practices that have proved instrumental in allowing the organization(s) to stay abreast of difficulties encountered ▪ Measure to what extent the organization(s) has/have been able to absorb, adapt, and transform
Leadership type (concentrated, decentralized, shared) (Ubels 2010)	▪ Meet with the leader(s) of the organization(s) targeted by your ECD strategy, and try to assess the degree to which their decision-making processes are participatory and inclusive of all staff perspectives ▪ Organize structured conversations with such leaders before implementation of the strategy begins, and try to learn what their respective frameworks of reference are (this might include assessing the type of literature, or the sources from which they draw the information that is informing their decisions)
Ownership (low, medium, high)	▪ Explore the extent to which the organization(s) targeted by your ECD strategy has/have actively participated in the design of the policy or project in relation to which your ECD strategy is being undertaken ▪ Measure the degree to which the different subgroups identified with the organization(s) in question have contributed to, and are still contributing to, the development, implementation, and evaluation of your ECD strategy ▪ Identify the opportunities for scaling up the strategy (this will include the analysis of available options to turn implementation into a sector-wide, multi-actor, inclusive endeavor, through which the roles of the funder and the external process facilitator can be gradually reduced)

government or a consortium of large funders is being planned. Too often ECD funding has concentrated on either the supply side (the individuals providing evaluation services) or the demand side (the commissioners of evaluations) of the equation. However, at a time when roles often overlap such a distinction no longer appears to be reasonable.

- For organizations that are implementing ECD strategies, as well as local organizations being targeted by ECD programs, ECD program objectives or organizational visions for evaluation should fit well with their respective institutional or organizational vision.

Enhancing Evaluation Capacity: The Equity Paradox

Considering the observed gaps in the targeting of ECD programs across funding agencies in the past, and based on the results of key ECB and ECD theoretical frameworks in use, this section provides a list of suggestions on how to make ECD targeting more inclusive and effective in the future (box 6.1).

IDEAS'S CONTRIBUTION TO A PARADIGM SHIFT IN CONTEMPORARY CAPACITY DEVELOPMENT THINKING AND PRACTICE

A Radically New Perspective on Evaluation Capacity Development

The role that IDEAS has played so far, and intends to play, in the area of ECD in the future is important and timely for three main reasons. First, it allows revitalization of the discourse on ECD among IDEAS members eight years after the IDEAS Global Assembly that was organized around this theme in Johannesburg, South Africa in 2009.[5] Second, it is likely to provide guidance in the production of some concrete tools and checklists that IDEAS members could use to enhance the effectiveness of the ECD work in their respective fields in the future. Third, it would build and expand upon the work that IDEAS has already done on evaluation competencies. Overall, the work of IDEAS in this area, as envisioned in the mission of the newly created Evaluation Capacity Development International Topical Interest Group (ECD ITIG), aims to build more consensus among members of IDEAS from different regions on what it means to work with organizations and governments on evaluation capacity in a more contextualized and sustainable manner.

[5] "Getting to Results: Evaluation Capacity Building and Development. For more details, please visit the conference website: https://ideas-global.org/2009-conference/.

BOX 6.1 Key suggestions to make ECD strategy more inclusive and effective

1. **Understand the specific ECD ecology where you are working.** In conducting a mapping of the major institutions—both within and outside of the government sphere, including VOPEs, academia, and the private sector that have demonstrated interest in ECD in the past—the identification of individuals as well as specific units with a more vested interest in evaluation is strongly recommended, so as to avoid the personalization of the evaluation function, which would then be exposed to the risk of collapse in case of staff turnover.
2. **Identify some common nationwide goals and objectives that all ECD stakeholders could be encouraged to contribute to.** Such goals, ideally aligned with the Sustainable Development Goals, do not need to be perceived as imposed from the outside, and should be consistent to the largest extent possible with the each targeted actor's mission and objectives.
3. **Build institutional incentives from within.** The incentives that are made available, including the identification and rewarding of champions, should not focus on meeting preset performance agreements, and should promote instead the ECD actor's ability to wonder and ask questions about how to turn short-term and isolated tactics into long-term and sustainable strategies.
4. **Conduct a participatory ECD diagnostic assessment.** Through such assessment, it will be important to foster opportunities for self-reflection and mutual learning. In this vein, it would be important that such assessment focus on a selected number of the organizational and environmental factors and processes discussed in this chapter.
5. **Develop a national ECD strategy and, depending on the scope of your intervention, put into place a national ECD task force.** Far from being a logical framework or road map, a national evaluation strategy is to be regarded as a work in progress and a living document, setting general objectives and leaving ECD stakeholders space to come up with creative and innovative ways to achieve the agreed-upon objectives. This phase might require addressing some key issues, such as the creation of dedicated evaluation units with three specific responsibilities—compiling a database of evaluation data; conducting data analysis; and, fostering dissemination of evaluation findings—as well as the establishment of partnerships between different departments within the same organization. With respect to an ECD task force, it is advisable that a variety of actors with different functions and roles (from both the public and private spheres), as well as entities whose membership cuts across different spheres, be involved.

(continued)

BOX 6.1 *(continued)*

6. **Focus on strengthening the capacity of local actors (both users and providers of evaluation services), and develop an opportunistic joint exit strategy**, in close collaboration with the individuals, organizations, and institutions involved in an ECD program. Although not operationalized from the outset, an exit strategy needs to be conceptualized from the very beginning of an ECD intervention. In order to advance sustainability, it is of utmost importance to have a very strong leader in place, who is capable of innovation and available to promote internal structural changes and the reallocation of budget resources based on the organization's needs and the changing contextual opportunities (what I have called "responsible systemic-ness").
7. **Evaluate the progress of your ECD over time.** In order to measure the effectiveness of your ECD support program, you need to make sure that your target audience as well as the type of capacity development you are trying to support (according to the interests and needs expressed by in-country ECD stakeholders) have been clearly defined at the time of ECD strategy development.

The Organizational Principles Driving IDEAS's ECD Efforts

Cognizant of the peer-reviewed literature produced to date, and building upon the understanding of contemporary evaluation practices, IDEAS is striving to create an enabling environment in which an authentic evaluation culture can flourish among its members in the future. The IDEAS ECD strategy rests on six key organizational principles that have been identified as among the most influential in the development of an evaluative culture, especially at the global level:

Membership diversity. The more diverse the membership of IDEAS in terms of roles and functions, the more likely it is that IDEAS will be able to affect the national evaluation discourse within both the public and private spheres.

Decentralized leadership. The more that IDEAS leadership is shared, and the larger the availability of channels through which members can contribute to IDEAS decision-making processes, the better the compliance with internal governance rules will be.[6]

[6] According to IDEAS current organizational set-up, all of the world's regions are equally represented in the Board. Individual board members coordinate, too, with the national and regional evaluation associations falling within their respective geographical spheres of competence.

Frequent diagnostics of both IDEAS capabilities and organizational processes. The more frequently capability assessments are conducted—for example, annually—and the more promptly identified weaknesses are addressed, the more successfully IDEAS will be able to fulfill its mission.

Availability and continued monitoring of the IDEAS theory of change. The more often the IDEAS theory of change is available and is revised, based on its program development and the findings of capability assessments carried out among its members, the more often the assumptions underlying it will be monitored, and programmatic improvements will be made.

Promotion of effective international and external communication. The more well-articulated a communication strategy is, the more social and political legitimacy IDEAS ECD work could benefit from it.

Availability of opportunities to members for disseminating, exchanging, and developing ideas, theories, and concepts about evaluation. The approval of a new publication and dissemination policy by the IDEAS Board in May 2017 is an important milestone in the association's history. The process of publishing high-quality, peer-reviewed papers is an integral part of the IDEAS ECD ITIG work plan, and of that of all other actors within the association who have an interest in capacity development and professionalization.

By enabling its members to publish original work on topics related to international development evaluation, IDEAS will be able to attain the following objectives in the short term:

- Give visibility to practitioners and decision makers with no prior track record of publications
- Disseminate evaluation-related ideas on topics and/or countries rarely discussed in the mainstream peer-reviewed evaluation literature
- Encourage its members to collaborate in documenting and writing about their own evaluation practices on a more regular basis—that is, not only in response to the call for conference proposals launched every two years before the IDEAS General Assembly
- Use the publication of articles and other items posted on the IDEAS website as an entry point for further dialogue, and for mutual intellectual and professional enrichment among its members

In the medium and long term, the new policy is expected to elicit a stronger sense of personal belonging to IDEAS, and that, as a result, will assist in advancing and furthering IDEAS members' practice of development evaluation, through strengthening their capacities, and their uptake of innovative evaluation methodologies.

The Vision and Values Driving IDEAS's ECD Efforts

IDEAS makes a constant effort to promote a more just, equitable, and democratic design for the management, funding and evaluation of ECD work. By questioning paradigms that have dominated the international development discourse over the last 10 years, IDEAS has called upon evaluation practitioners and international partners to engage in evaluation and in capacity development with justice, sustainability, and transformative change in mind. In so doing, IDEAS is encouraging more reflection among members on what kinds of postures evaluation practitioners could and should have toward issues of equity, cultural competence, type and quality of evidence, sustainability, and the use of evaluation. Furthermore, this work promotes a more systematic appreciation of the cultural and linguistic diversity (what I call the "equality" of differences) within the IDEAS as well as the rest of the evaluation community. The ECD ITIG, for instance, complements the other IDEAS ITIGs, and helps to strengthen the cultural and linguistic diversity of the association's membership, as a strategy for enhancing the sustainability of future evaluation endeavors. As part of such a strategy, some of the key IDEAS resources will be translated into languages other than English (French and Spanish, among others), and stronger links with other regional evaluation conferences through the establishment of joint projects will be strengthened.

Technical Considerations Driving IDEAS's ECD Efforts

Besides the foundational and more vision-related principles inspiring its future ECD strategy, a number of technical considerations are driving IDEAS's ECD work, as follows:

- Given the lack of an effective decentralization of the M&E function in many countries, it has proven extremely difficult to promote a defused culture of evaluation through a top-down approach. Therefore, through the involvement of members at the country and subnational levels, IDEAS is attempting to extend the in-country boundaries of the evaluation culture.
- Due to the relatively weak data analysis skills among national evaluators, IDEAS aims to enhance the statistical as well as the qualitative analysis of their members through the use of webinars and other publications.
- In response to misconceptions about evaluation, or "reductive" understanding of the purposes of evaluation,[7] IDEAS will

[7] Many evaluation commissioners and planners believe that evaluation consists in verifying the compliance between what happens on the ground and the original objectives spelled out in the project logframe, with almost no reference to unexpected impact. This is what I refer to as the "RBM-ization of the evaluation function" (Tarsilla 2014b). As a result of such limited interpretation of evaluation objectives, the risk is that evaluation could become highly centralized, with little room left for conducting independent and equity-based evaluations of programs and policies.

increasingly promote critical publications, blogs, and online discussions on the IDEAS website or on LinkedIn, as a way to promote a more exhaustive definition of the evaluation function.

CONCLUSION

This chapter encourages funders as well as practitioners to rethink the way they plan, implement, and evaluate capacity development efforts. Based on a review of existing gaps in contemporary capacity-building practices in the evaluation arena, this chapter calls on all international development actors globally: to adopt a more contextually relevant, adaptive, equitable, inclusive, and democratic definition of ECD; to conduct more exhaustive diagnostics of both capacity and processes among the organizations and entities targeted by their ECD strategies; and to assess the distribution of power and resources within the systems where ECD strategies are expected to be implemented. Lastly, by documenting the current IDEAS initiatives that are aimed at maintaining and promoting evaluation capacity at several levels (national, regional, and global), this chapter attests to the association's leadership in the ECD arena.

REFERENCES

ACBF (African Capacity Building Foundation). 2016. *African Union Agenda 2063. Capacity Development Plan Framework: Buttressing Implementation of the First 10-Year Plan "The Africa We Want."* Harare, Zimbabwe: ACBF.

Argyris, C., and D. Schön. 1996. *Organizational Learning II: Theory, Method and Practice*. Reading, MA: Addison Wesley.

AU and NEPAD (African Union and New Partnership for Africa's Development Planning and Coordinating Agency). 2012. *Africa's Capacity Development Strategic Framework*. Midrand, South Africa: NEPAD.

Bamberger, M. 2009. *Institutionalizing Impact Evaluation within the Framework of a Monitoring and Evaluation System*. Washington, DC: The World Bank.

Boyle, R., D. Lemaire, and R.C. Rist. 1999. "Incentives, Constraints and Culture-Building as Instruments for the Development of Evaluation Demand." In *Building Effective Evaluation Capacity*, R. Boyle and D. Lemaire, eds., 153–77). New Brunswick, NJ: Transaction Publishing.

Bohni, Nielsen S., and K. Attström. 2011. "Perspectives on Evaluation Capacity Building." In *Influencing Change: Building Evaluation Capacity to Strengthen Governance*, Ray C. Rist, Marie-Helene Boily, and Frederic Martin, eds. Washington, DC: World Bank.

Compton, D.W., M. Baizerman, and S.H. Stockdill, eds. 2002. "The Art, Craft, and Science of Evaluation Capacity Building." *New Directions for Evaluation* 93. San Francisco: Jossey-Bass.

Fukuyama, F. 2004. *State-Building: Governance and World Order in the 21st Century*. Ithaca, NY: Cornell University Press.

Fung, A. 2003. "Recipes for Public Spheres: Eight Institutional Design Choices and Their Consequence." *Journal of Political Philosophy* 11: 338–67.

Gaventa, J., and G. Barrett. 2010. "So What Difference Does It Make? Mapping the Outcomes of Citizen Engagement." IDS Working Paper 347. Institute of Development Studies, Brighton.

GIZ (German Agency for International Cooperation). 2013. *GIZ's Monitoring and Evaluation Policy*. Bonn: GIZ.

Lennie, J. 2005. "An Evaluation Capacity-Building Process for Sustainable Community IT Initiatives: Empowering and Disempowering Impacts." *Evaluation* 11: 390-414.

Lusthaus, C., M.H. Adrien, G. Anderson, F. Carden, and G.P. Montalvan. 2002. *Organizational Assessment. A Framework for Improving Performance*. Ottawa/Washington, DC: International Research Development Centre/Inter-American Development Bank.

Mayne, J. 2008. "Building an Evaluative Culture for Effective Evaluation and Results Management." ILAC Working Paper 8. Institutional Learning and Change Initiative.

Morgan, P. 1998. *Capacity and Capacity Development: Some Strategies*. Hull: Policy Branch, CIDA.

Norad (Norwegian Agency for Development Cooperation). 2006. "Instruction for Evaluation Activities in Norway's Aid Administration."

OECD (Organisation for Economic Co-operation and Development). 2006. *The Challenges of Capacity Development: Working towards Good Practice*. Paris. OECD.

Ortiz, A., and P. Taylor. 2009. "Learning Purposefully in Capacity Development: Why, When and What to Measure?" Opinion Paper. Brighton, Sussex: Institute of Development Studies.

Pasteur, K. 2006. "Learning for Development." In *Relationships for Aid*, R. Eyben ed. London: Earthscan.

Rist, Ray C., Marie-Helene Boily, and Frederic Martin, eds. 2011.*Influencing Change: Building Evaluation Capacity to Strengthen Governance*. Washington, DC: World Bank.

Solomon, M.J., and A.M. Chowdhury. 2002. "Knowledge to Action: Evaluation for Learning in a Multi-Organisational Global Partnership." *Development in Practice* 12: 346–54.

Stevenson, J.F., P. Florin, D.S. Mills, and M. Andrade. 2002. "Building Evaluation Capacity in Human Service Organizations: A Case Study." *Evaluation and Program Planning* 25: 233–43.

Sida (Swedish International Development Cooperation Agency). 2007. *Looking Back, Moving Forward. Sida Evaluation Manual*. 2nd revised ed. Stockholm: Sida.

Tarsilla, Michele. 2012. "From Building Evaluation Capacity to Supporting Evaluation Capacity Development: The Cases of Democratic Republic of Congo, Niger, and South Africa." PhD dissertation, Western Michigan University.

———. 2014a. "Evaluation Capacity Development in Africa: Current Landscape of International Partners' Initiatives, Lessons Learned and the Way Forward." *African Evaluation Journal* 2 (1).

———. 2014b. "From RBM-ization to Normalization: A Field Practitioner's Reflection on ECD Current Trends." OECD Newsletter.

Trochim, W. 2006. "An Evaluation Culture." *Research Methods Knowledge Base* website.

Ubels, J. 2010. "Stimulating the Provision of Local Capacity Development Support." In *Capacity Development in Practice*, J. Ubels, Na-A. Acquaye-Badoo, and A. Fowler, eds. New York: Earthscan.

UNESCO (United Nations Educational, Scientific and Cultural Organization). 2015. *UNESCO Evaluation Policy*. 196 EX/24.INF. UNESCO.

USAID (United States Agency for International Development). 2011. *Evalution: Learning from Experience. USAID Evaluation Policy*. Washington, DC: USAID.

Wiesner, E. 1997. "Evaluation, Markets, and Institutions in the Reform Agenda of Developing Countries." In *Evaluation for the 21st Century: A Handbook*, E. Chelimsky and W. R. Shadish, eds., 189–200. Thousand Oaks, CA: Sage Publications.

Wijayatilake, D. 2011. "Working toward Development Results: The Case of Sri Lanka." In *Influencing Change: Building Evaluation Capacity to Strengthen Governance*, Ray C. Rist, Marie-Helene Boily, and Frederic Martin, eds., 13–22. Washington, DC: World Bank.

World Bank. 2005. *Building Effective States, Forging Engaged Societies*. Washington, DC: The World Bank.

Chapter 7

Mentoring Emerging Evaluators – Sharing Experiences from the Global South

Awuor Ponge, Taiwo Peter Adesoba, Ahmed Tammam, and Tara Devi Gurung

Abstract. *This chapter describes a mentoring program the International Development Evaluation Association (IDEAS) is implementing to support emerging evaluators. Building a business case for mentoring, the authors attempt to establish evidence for what has and has not worked, and suggest how the program can be carried out effectively. Links between mentoring and the professionalization of evaluation, and the potential benefits of the IDEAS program are discussed, as well as different models of mentoring; mentoring policies and procefdures; how they should be developed, and by whom; the importance of recognizing mentors for the work they do; and reverse mentorship, with young evaluators mentoring older professionals in social media and the use of digital technology. Summaries of presentations made by three of the authors at a panel discussion of young and emerging evaluators at the 2015 IDEAS Global Assembly describe mentoring experiences in Nigeria, the Middle East and Eurasia, and Nepal. The*

Awuor Ponge, African Policy Centre, awuorponge@gmail.com; Taiwo Peter Adesoba, East Tennessee State University College of Public Health, oluwasesan@gmail.com; Ahmed Tammam, Independent Consultant, dr.egypt85@gmail.com; Tara Devi Gurung, Nepal Red Cross Society, gurung.tara@gmail.com.

chapter offers conclusions drawn from discussions about the pilot mentoring program, and what it portends for young and emerging evaluators, especially in the Global South, as they position themselves in readiness for the evaluation of the Sustainable Development Goals.

Emerging evaluators often lack the portfolio of experience required of professionals. One solution to this challenge lies in strengthening the professional capacity and credibility of less experienced evaluators. The International Development Evaluation Association (IDEAS) is piloting a mentoring program for young and emerging evaluators to create opportunities to help them build specific professional skills.

Many young evaluators, as well as those entering the field in a career shift, are disadvantaged in terms of consultancy and job placement because of a lack of experience. Over the past decade, mentoring has proliferated as an intervention strategy for addressing the needs of young people for adult support and guidance throughout their career development. This chapter draws on experiences shared at a panel discussion at the Global Assembly of IDEAS in Bangkok in October 2015. The specific aims of this discussion were to allow young and emerging evaluators to share some of the challenges they are experiencing as they enter the evaluation profession; to brainstorm on how context-specific challenges can be addressed in such a way as to set emerging evaluators on a stable footing in the profession; and to share best practices from across the continents so that emerging evaluators can learn from the challenges of others, and how they have been addressed.

There is ample evidence of the positive contribution of mentoring to improvement in skills development, social and professional competence, and intellectual development, as well as the development of the vocational skills needed for professionalization of the evaluation discipline, while positioning mentees for professional satisfaction in their practices.

IDEAS REACHING OUT TO YOUTH

IDEAS is a global membership organization focused on the evaluation of development that brings together evaluation practitioners from all the corners of the world to help develop their professional skills while enhancing networking among members and recruiting others into the evaluation profession. In order to promote the inclusion of youth, IDEAS has initiated a youth membership category for evaluators up to age 30, with a reduced annual fee. The mentoring program is to be spearheaded by young evaluators themselves, under the guidance of senior evaluators as mentors.

ABOUT EMERGING EVALUATORS

Emerging evaluators often lack the portfolio of experience that is required by potential employers and the commissioners of consultancy assignments during the hiring process. One possible solution for enhancing the skills set and credibility of less experienced evaluators involves the establishment of

a mentoring scheme that pairs emerging evaluators with more experienced counterparts. Established evaluators are able to mentor upcoming evaluators in various ways, including working with them on assignments. Through involving youth in evaluation, we are also enhancing their individual development and encouraging their active involvement in decisions that will affect their lives. The strategy involves setting up partnerships with development organizations globally in order to create opportunities for students and youth evaluators to benefit from the advice of senior evaluators, and give them the hands-on experience they need to build specific professional skills. A strategy known as reverse mentorship will also be employed, in which youth evaluators will mentor older members in new and emerging trends in development evaluation, including but not limited to the use of digital technology and social media.

THE CONCEPT OF MENTORING

Mentoring programs for youth are commonplace in today's society: more than 5,000 such programs in the United States serve an estimated 3 million young people (MENTOR/National Mentoring Partnership 2006, as quoted in DuBois et al. 2011). In a typical program, each youth is paired with a volunteer from the community, with the aim of cultivating a relationship that will foster the young person's positive development and wellbeing (DuBois et al. 2011).

A mentoring relationship is most often defined as a professional relationship in which an experienced person (the mentor) assists a less-experienced person (the mentee) in developing specific skills and knowledge that will enhance the mentee's professional and personal growth. Evaluators are educators: their success is judged by what others can learn from their work. Mentoring is generally viewed as one component of a more comprehensive youth development strategy: these strategies can help youth gain the competencies they need to meet the challenges of adolescence and become successful adults (Foster 2001).

Over the past decade, mentoring has proliferated as an intervention strategy for addressing the needs of young people for adult support and guidance in the development of their careers. Widespread expansion of youth mentoring programs in the United States was inspired by the release of a report on an evaluation of the Big Brothers Big Sisters of America, a community-based mentoring program. Findings from this research provided evidence of associations between mentoring and a range of positive youth outcomes, and were widely embraced by policy makers and practitioners (Tierney, Grossman, and Resch 1995). Sharing experiences is a way of building community, highlighting commonalities, engaging in practices of cathartic release, and often shining light on other matters that might otherwise remain hidden. To share experiences is to articulate them, and to articulate them is to gain power over them, rather than to feel "stuck" (Vallabha 2015).

As human beings, we live in social groups where we learn our norms, values, and behaviors by the example and coaching of others. Mentoring happens in all organizations, whether it is fostered as a development strategy or encouraged as an informal process. People are constantly learning

from others, adopting modeled behaviors and attitudes, and absorbing the culture and perceived values of the organization or professional environment through their personal interactions with other members of the organization.

In the context of the IDEAS mentoring program, it is not age alone that affects the relationship between a mentor and a mentee. A mentee may be older, but changing careers. Or he or she might be an experienced evaluator who needs help learning how to use qualitative analysis software from a colleague who has experience with this skill.

DEVELOPING THE CASE FOR MENTORING: ESTABLISHING THE EVIDENCE

Because individuals may experience mentoring at various life stages, it is not surprising that there are three distinct streams of mentoring: youth mentoring; academic mentoring; and workplace mentoring. Youth mentoring involves a relationship between a caring, supportive adult and a child or adolescent (Rhodes 2002).

Mentee motivation and involvement in the evaluation profession may be greatly influenced by mentoring. Role modeling can expose mentees to the field of evaluation and to social opportunities that can open their eyes to different possibilities and motivate them to seek out new experiences. Those who have been mentored are more likely to see the need to go the extra mile in order to be fully engrossed in the profession: this can involve many different matters currently under debate regarding professionalization, including certification and accreditation, as well as other forms of recognition.

A mentoring relationship can promote career success. Mentors can impart specific knowledge and expertise that contributes to mentee learning and skill development (Eby et al. 2008; Kram 1985). Mentors can also facilitate professional networking by introducing mentees to influential individuals within academic and organizational contexts: these important career contacts can in turn lead to career success in terms of salary, promotions, and job offers (Kram 1985). Anecdotal evidence abounds of those who have been successful in evaluation practice as a result of the networking and informal mentoring they have received through working with more advanced or senior evaluators. This certainly is a positive step toward professionalizing not only the individual, but the discipline as well.

A study of youth development interventions concluded that a wide range of youth development approaches, including mentoring, result in positive behavioral changes, such as improved interpersonal skills and relationships, and increased self-control and academic achievement (Foster 2001). For example, the mentored professional will most certainly see the need to pursue academic or professional courses that are geared toward building their capacity in evaluation.

At IDEAS, the emphasis is on workplace mentoring, which occurs in an organizational setting, and the purpose of which is the personal and professional growth of the mentee (Kram 1985). The mentor may be a supervisor; someone within the organization who is outside the mentee's chain of command; or an individual in another organization (Eby 1997). Mentoring at

different developmental stages tends to serve different functions or purposes. Youth mentoring is often aimed at reducing risky behavior, or improving social and academic functioning; academic mentoring tends to target student retention, academic performance, and adjustment to college life; while workplace mentoring aims to enhance the personal and career development of employees (Eby et al. 2008).

A study of the top 200 executives in 50 large U.S. companies revealed that mentoring is one of the most effective tools in personnel development (Chambers et al. 1998). Another study, by the Institute of Management, involving 1,500 U.K. managers revealed that mentoring is one of the two most powerful leadership development tools used in organizations. It has also been reported that 71 percent of Fortune 500 and private companies use mentoring in their organizations, and that 77 percent of U.S. companies surveyed said that mentoring had improved both the retention and performance of employees.[1]

In addition to correlations with higher educational aspirations, the research shows that mentoring is correlated with other positive developmental outcomes, including changing attitudes (higher self-esteem, and stronger relationships with adults, including with teachers and peers) and better behavior (Bruce and Bridgeland 2014). In 2002, DuBois and colleagues published a meta-analytic synthesis of findings from 55 evaluations of youth mentoring programs that had been published through 1998 (DuBois et al. 2002). These findings indicated that, on average, youth participating in mentoring programs had benefited significantly in each of five outcome domains: emotional/psychological, problem/high-risk behavior, social competence, academic/educational, and career/employment (DuBois et al. 2002, 2011).

Studies have shown that many young adults are entering the labor force without even the limited skills that are necessary to attain a job in the first place, such as interviewing skills, conflict resolution, and effective communication (Eccles and Gootman 2002). Mentoring has been linked with a myriad of intellectual skills and development, including good decision-making skills, in-depth knowledge of more than one culture, knowledge of both essential life skills and vocational skills, and rational habits of mind, such as critical thinking and reasoning skills. One study found that nearly all young adults who had formal mentoring relationships (95 percent) found these experiences to be "helpful," half of which (51 percent) found the relationship to be "very helpful." Similarly, nearly all youth in informal mentoring relationships (99 percent) said their experience was "helpful," seven in 10 (69 percent) reporting it as "very helpful" (Bruce and Bridgeland 2014).

All of these studies can be summed up in one statement: there is ample evidence of the positive contribution of mentoring to improvement in skills development, social and professional competence, and intellectual development, as well as the development of the vocational skills that are needed for professionalization of the discipline: it also positions mentees for professional satisfaction in the practice.

[1] Chronus, https://chronus.com/how-to-use-mentoring-in-your-workplace.

WHAT IS THE ADDED VALUE OF A MENTORING PROGRAM FOR IDEAS?

Organizations recognize that workforce demographics have changed dramatically in recent years. Youth are becoming employed in large numbers, thanks to the global youth bulge. More and more graduates are joining the workforce as junior professionals in evaluation practice, while some are developing interest in the profession while they are still in university. There is a need to provide sufficient growth opportunities for potential future professionals in the practice. Senior evaluators also need to be mentored in new and emerging evaluation methodologies, including the use of digital technology and social media: they could benefit from the proposed reverse-mentoring program.

These are the potential benefits of the IDEAS mentoring program for the organization:

- Recruitment of new members to the organization and discovery of talent
- Development of leadership for the future survival and prosperity of IDEAS
- Communication of values, goals, and plans of the organization globally
- Demonstration of personal and professional standards among members
- Implementation of equity initiatives
- Fostering of shared values and teamwork
- Building a strong global learning organization in evaluation practice
- Development of cross-organizational networks
- Increase in morale and motivation among both junior and senior professionals

DEVELOPMENT OF A MENTORING PROGRAM

The IDEAS Board has appointed a mentoring program coordinator, who proposed the program. The program coordinator then formed a task force of five people, which has been approved by the Board to develop the mentoring program policy, as well as to oversee its implementation. Members of the task force represent a cross-section of the organization, including potential mentors and mentees, as well as stakeholders who bring value to the process. The duties of the task force include the following:

- Determining the goals of the program
- Choosing the proper mentoring model
- Developing a memorandum of understanding (MOU) template for the mentoring partnership
- Negotiating the MOU with major institutions commissioning development evaluations
- Selecting eligible evaluation training programs/institutions
- Defining criteria for mentors and mentees

- Defining other critical components of the program
- Matching the participants
- Monitoring the pilot
- Evaluating the results at the end of the pilot program

The mentoring panel at the Global Assembly was sponsored by the U.K. Department of International Development (DFID), among others, but no other funds could be raised. Thus, many of the more ambitious plans turned out to be unrealistic. Similar initiatives in the European Evaluation Society and the American Evaluation Association were largely voluntary in nature; this seems to be a more realistic way forward for IDEAS as well.

PURPOSE AND OBJECTIVES OF THE IDEAS MENTORING PROGRAM

The mentoring program at IDEAS aims to achieve the following:

- **Induction and skills enhancement.** Helps new members settle into the organization; facilitates potential and skill development for new members, both young evaluators and those just entering a career in evaluation; seeks to enable skills to be passed on by experienced, highly competent professionals to others who need to develop specific skills.
- **Networking and career development.** Helps both mentors and mentees in the planning, development, and management of their careers; helps them become more resilient in times of change; and more self-reliant in their careers. Offers young professionals visibility and the opportunity for networking, which helps them to explore and plan their career pathways. Also helps both mentors and mentees gain a greater awareness of opportunities and activities that can broaden their professional experience.
- **Education support and practical orientation.** Helps bridge the gap between theory and practice; complements formal education and training through sharing the knowledge and hands-on experience of competent practitioners. Offers mentees the opportunity to acquire new knowledge and skills by observing and understanding the mentor's practical experience.
- **Leadership and development of competencies.** Encourages the development of leadership and professional competencies that are more easily gained through example, guided practice, or experience than through theoretical education and training.
- **Global visibility and organizational development.** Will expand IDEA's culture of cooperation and commitment through sharing the values, vision, and mission of the organization, and will give IDEAS an enhanced visibility globally.

The program proposes the following models of mentoring:

- **One-to-one mentoring**—matches one mentor with one mentee
- **Group mentoring**—assigns one mentor to work with several mentees
- **Team mentoring**—involves more than one person working with the same mentee
- **Computer online mentoring**—uses computer-based opportunities to develop relationships through online communication
- **Peer mentoring**—young people experienced in evaluation mentor other young people who are just entering the evaluation profession

In order to realize and sustain the program, IDEAS needs to find mechanisms for arranging mentoring on a voluntary basis, without giving up on the effort to mobilize resources for the program that are needed for purchasing mentoring software; covering the administrative costs for staff managing the program; and the costs of training for mentors and mentees, among other things. There will also be a need to provide mentors with formal recognition for all they do for young and emerging evaluators. This recognition can take many forms, such as awards—for example designating a mentor/mentee match of the year—as well as gifts and/or letters of appreciation.

Three of the case studies presented at the Global Assembly in Bangkok in October 2015 are summarized below.

NIGERIA CASE STUDY

In a presentation titled "Evaluation Capacity Development for Emerging Evaluators: A Nigerian Experience," Taiwo Peter Adesoba observed that describing an emerging evaluator is sometimes a tricky task, because there seems to be no globally agreed-upon definition in terms of age, educational requirements, job experience, and so on. Just as a plant has hurdles to overcome when emerging from the soil, so emerging evaluators have particular needs while they are trying to establish themselves in the evaluation profession. A major challenge for young evaluators, especially in low- and middle-income countries, is weak evaluation capacity, which leads to their exclusion from evaluation activities. With the growing number of youth-led organizations, especially in Africa and Asia, more attention is being given to the monitoring and evaluation (M&E) components of their projects. Young development professionals are keenly interested in evaluating the impact of their own work: the demand for accountability from policy makers and other project implementers in their communities is also a factor. Although young evaluators are passionate about implementing evaluation assignments, the technical skills required to properly fulfill this passion is poor. This makes evaluation capacity development (ECD) a necessity for young and emerging evaluators.

The ECD Project in Nigeria is targeted at young evaluators aged 18-30 with less than two years' experience in M&E. The project was conceptualized following a survey among young evaluators in which about 81 percent of respondents said they had never had formal training in M&E: 92 percent of them did not belong to any voluntary organizations for professional evaluation (VOPEs), and 100 percent of them said they needed mentoring. The

objectives of ECD programs are to improve youth inclusion not only in project implementation, but in its evaluation; to promote and advance the practice of the evaluation profession; and to increase youth participation in decision making for sustainable development. The participants in the first phase of the ECD were 24 (7 males; 17 females) emerging evaluators ranging from 21 to 33 years who work for local nongovernmental organizations (NGOs) in the Ondo and Ekiti states. They take on M&E roles in their organizations, but have only limited experience. About 20 percent of the trainees had attended M&E training previously. None of them belonged to any VOPEs in Nigeria, and many had never attended an evaluation conference. The ECD covered the basics of M&E, and its application to their organizations' projects.

Some of the lessons learned during the two-day training were tailored toward the project areas of the trainees. Assignments and group work were aimed at the areas of HIV prevention, economic empowerment, and gender programs. After two days of intensive training, group work, and discussions, the participants showed a good understanding of the basics of M&E. Some have gone on to find online evaluation webinars, and two of the trainees have gotten internship opportunities with NGOs in other parts of the country. However, better results could be achieved if the trained young evaluators had access to mentors who could provide them with additional guidance.

Adesoba concluded that capacity development of emerging evaluators for sustainable development will bring about better results if well-qualified evaluation professionals are available to provide on-the-job mentoring, through either short-term internships or evaluation job placements.

THE MIDDLE EAST AND EURASIA: SAVE THE CHILDREN CASE STUDY

In his presentation "The Future of Evaluation," Ahmed Tammam discussed the challenges faced by emerging evaluators in the Middle East and Eurasia (MEE) region. He observed that M&E is a critical component in the advancement of evidence-based interventions, and that it consequently enhances efficiency, transparency, and accountability. Evaluation is a profession, and grooming new leadership for the future of this profession is needed: thus, investment in developing the capacities of young evaluators is an emerging priority.

Tammam noted that the evaluation profession is far ranging. Therefore, support from experienced evaluators is key in order for young evaluators to advance in such a robust profession, which has so many specific and different areas of work. Evaluation is a growing career path, but visibility, employment, and networking opportunities can be very limited for young evaluators.

He further noted that talented young evaluators in the MEE region are increasingly lobbying donors to fund only successes, and described a case study of Save the Children in the MEE region as an example. Save the Children is a child rights–based organization, working in 12 countries in the Middle East and Eurasia, with operations that support children in developmental contexts (as in Egypt and Georgia), as well as in an emergency contexts (as in Syria, Ukraine, and Yemen). With the expansion of needs in the

region and the limitation of resources, there is a push to focus only on what works best for children. Consequently, there is a need for effective evaluation, and thus for evaluators.

In 2014, the program quality department of Save the Children at the MEE regional office began a process of investing in the talents of local emerging evaluators by recruiting an evaluator from a country office to manage the M&E tasks on the regional level. One of the mandates of the recruited evaluator was to begin to establish a cadre of talented young evaluators. A capacity-building process for emerging young evaluators was undertaken, promoting intra-learning and providing opportunities for these individuals to gain more knowledge and enhance their experience as emerging evaluators. This interactive learning program was mentored by senior program-quality staff who helped the emerging evaluators map the real needs in their countries.

As a result, an interactive mapping exercise of local M&E needs and capacities was conducted by the emerging evaluators, and a regional think tank was created. Through this process all of the young evaluators gained on-the-job experience. Two of them offered to support other country offices; another five were given the authority to be part of the operationalization process of the Vision and Position Paper of the Eurasia region (Albania, Armenia, Georgia, Kosovo, and the Northwest Balkans), which aims to align the efforts of five countries to work together on subregional programs in order to maximize benefits for children and to efficiently utilize the limited funding opportunities, by pushing donors to fund success.

In conclusion, Tammam noted that strengthening young and emerging evaluators through enhancing the learning environment and providing opportunities for mentorship are key not only to developing the young evaluators' future, but the future of the evaluation profession as well.

NEPAL CASE STUDY

Tara Devi Gurung presented a case based on her anecdotal experience as an emerging evaluator in Nepal, and also from a desk review of the available literature about development evaluation, and the role of the young and emerging evaluators in Nepal.

In her presentation, "Evaluation Challenges and Opportunity for Emerging Evaluators: A Nepalese Experience/Case," Gurung noted that evaluation, which assesses the results of policies, programs, and projects, is an integral part of the development process. Evaluation is particularly critical in the context of Nepal, which has a complex social structure, a high poverty rate, gender discrimination, dynamic forces of globalization sweeping traditional societies, and numerous development projects aimed at the population. These factors, in addition to greater competition for limited resources available for international development are pushing donors, program participants, and evaluators to seek more rigorous, but still flexible, systems for monitoring and evaluating development and humanitarian interventions.

Many of the current approaches to evaluation are unable to address the changing structure of development assistance, and the increasingly complex environment in which it operates. Innovative evaluation approaches

and practices are particularly important in such situations. However, it is challenging for emerging evaluators to carry out innovative evaluation, since this calls for a high level of expertise.

Emerging evaluators are those who have recently joined the profession and have limited experience. They are not necessarily young, although most of them are relatively young. Many of them are disadvantaged in terms of consultancy and job placement because they lack sufficient experience. There are only limited forums for sharing and disseminating best practices; developing quality and ethical protocols; enhancing the capacity of new evaluators; and researching evaluations. There are limited resources for evaluation that collects lessons drawn from evaluations around the world; produces knowledge through research undertakings; and supports the development of curricula for and carries out basic and advanced training in evaluation. Moreover, there is no academic institution in Nepal that offers a university degree in evaluation. Lack of specific acts, rules, and regulations for evaluation has posed even more challenges for evaluators in Nepal. Available guidelines focus only on governmental M&E, and do not cover other sectors.

In this context, it is difficult to raise funds for evaluation research, capacity building, and activities related to the promotion of evaluation in Nepal. The funding agencies have a tendency to support already established organizations and firms rather than new evaluators. Often the potential for innovation, and the expertise of emerging evaluators, are overlooked by the commissioners of evaluations. Better representation, and the active engagement of young and new evaluators in the decision-making process are needed in order to bring their ideas and perspectives into evaluation.

Gurung concluded that effective evaluation is crucial in order to assess the progress and impact of the efforts of the government policy to "Build a New Nepal" through accelerated development inclusive of all castes, ethnic groups, and genders. Evaluation is gradually becoming an integral part of development plans, projects, and emergency operations in Nepal. Gurung predicts that this will lead to a rise in skilled evaluators, including emerging evaluators in the near future.

CONCLUSION

The IDEAS mentoring program aims to enhance induction and skills development, networking and career development, education support and practical orientation, leadership and competencies development, global visibility, and organizational development for young and emerging evaluators.

Mentoring at a variety of developmental stages tends to serve different functions or purposes. Workplace mentoring is aimed at enhancing employees' personal and career development: this is the kind of mentoring that IDEAS is piloting. In general, mentoring has been linked with a myriad of intellectual skills and development, including good decision-making skills, in-depth knowledge of more than one culture, knowledge of both essential life skills and vocational skills, and rational habits of mind such as critical thinking and reasoning skills. There is ample evidence of the positive contribution of mentoring to improvement in skills development, social and professional

competence, and intellectual development, as well as the development of the vocational skills that are needed for professionalization of the evaluation discipline, while also positioning mentees for professional satisfaction in their practices.

Studies have shown that, among other benefits, youth participating in mentoring programs have benefited significantly in social competence, academic and educational progression, as well as career or employment advancement. The three cases discussed in this chapter demonstrate that there is a country-level need for supporting young and emerging evaluators through mentoring in order to advance their professional competence and career prospects. These cases demonstrate further that capacity development of emerging evaluators will ensure that more well-qualified evaluation professionals are available to provide on-the-job mentoring through short-term internships or evaluation job placements. They also demonstrate that strengthening young and emerging evaluators through enhancing the learning environment, and providing opportunities for mentorship are key, not only to developing the future of young evaluators, but for the future of evaluation as a profession in general. Overall therefore, these studies build a strong business case for the mentoring of young and emerging evaluators.

ACKNOWLEDGMENTS

Special thanks to the members of the IDEAS Executive Board namely: Rob D. van den Berg (President); Susan Daet Tamondong (Vice-President), and Cristina Magro (Secretary), who have all supported this initiative since inception. We also wish to acknowledge the valuable comments received from Taiwo Peter Adesoba, Kruno Karlovcec, Jérôme Gandin, and Halcyon Louis, who are all members of the IDEAS Mentoring Policy Development Team which is leading the piloting of this initiative at IDEAS. Lastly, thanks to the panelists at the Mentoring Panel Session in Bangkok, Thailand: Awuor Ponge, Taiwo Peter Adesoba, Ahmed Tammam, Tara Devi Gurung, and Ana Borges.

REFERENCES

Bruce, M., and J. Bridgeland. 2014. "The Mentoring Effect: Young People's Perspectives on the Outcomes and Availability of Mentoring." A report for MENTOR: The National Mentoring Partnership.

Chambers, E.G., M. Foulon, H. Handfield-Jones, S.M. Hankin, and E.G. Michaels III. 1998. "The War for Talent." *McKinsey Quarterly* 3: 44–57.

DuBois, D.L., B.E. Holloway, J.C. Valentine, and H. Cooper. 2002. "Effectiveness of Mentoring Programs: A Meta-Analytical Review." *American Journal of Community Psychology* 30, 157–97.

DuBois, D.L., N. Portillo, J.E. Rhodes, N. Silverthorn, and J.C. Valentine. 2011. "How Effective Are Mentoring Programs for Youth? A Systematic Assessment of the Evidence." *Psychological Science in the Public Interest* 12 (2):57–91.

Eby, L.T. 1997. "Alternative Forms of Mentoring in Changing Organizational Environments: A Conceptual Extension of the Mentoring Literature. *Journal of Vocational Behavior* 51: 125–44.

Eby, L.T., T.D. Allen, S.C. Evans, T. Ng, and D. DuBois. 2008. "Does Mentoring Matter? A Multidisciplinary Meta-Analysis Comparing Mentored and Non-Mentored Individuals." *Journal of Vocational Behavior* 72 (2): 254–67.

Eccles, J., and J.A. Gootman, eds. 2002. *Community Programs to Promote Youth Development*. Washington, DC: National Academy Press.

Foster, L. 2001. "Effectiveness of Mentor Programs: Review of the Literature from 1995 to 2000." CRB-01-004. California Research Bureau, Sacramento, CA.

Kram, K.E. 1985. *Mentoring at Work*. Glenview, IL: Scott Foresman.

MENTOR/National Mentoring Partnership. 2006. "Mentoring in America 2005: A Snapshot of the Current State of Mentoring." Alexandria, VA: MENTOR/National Mentoring Partnership.

Merrick, L. n.d. "How Coaching and Mentoring Can Drive Success in Your Organization." White Paper, Chronus Mentoring and Talent Development Solutions.

Rhodes, J.E. 2002. *Stand by Me: The Risks and Rewards of Mentoring Today's Youth*. Cambridge, MA: Harvard University Press.

Tierney, J.P., J.B. Grossman, and N.L. Resch. 1995. *Making a Difference. An Impact Study of Big Brothers Big Sisters*. Philadelphia: Public/Private Ventures.

Vallabha, Bharath. 2015. "The Value of Sharing Experiences." *The Rough Ground* blog.

Part III
Regional
Perspectives

Chapter 8

Evaluating Sustainable Development in SIDS – Lessons from the Pacific and the Caribbean

Juha I. Uitto, Jeremy Kohlitz, and David Todd

Abstract. *Small island developing states (SIDS) are particularly vulnerable to environmental stresses, and especially to the impacts of climate change. This is due to numerous factors, including limited geographic size and extensive coastal areas; remote locations; fragile economies that are often dependent on narrow sectors; limited natural resources and access to fresh water and energy; small populations; and weak institutional capacity. Managing sustainable development requires coherent and effective policies and strategies. An essential part of the formulation and implementation of such policies and strategies is effective monitoring and evaluation (M&E). Evaluation is also needed to ensure that interventions, policies, and strategies are achieving their goals and contributing to sustainable development. This chapter reviews experiences with M&E in the Pacific and Caribbean SIDS. It identifies a number of challenges that*

Juha I. Uitto, Global Environment Facility Independent Evaluation Office, juitto@thegef.org; Jeremy Kohlitz, Institute for Sustainable Futures, University of Technology Sydney, Jeremy.Kohlitz@uts.edu.au; David Todd, International Development, Environment and Disasters, deemtuk@yahoo.com.

need to be overcome, including limited human and institutional capacities, and the perception that evaluation is only important for donor-funded programs, which results in low priority being given to M&E. These challenges are best addressed by crafting M&E systems that are appropriate for a variety of SIDS contexts, that are country led, and that are supported by external agencies in a coherent manner.

Small island developing states (SIDS) are facing unique and often severe challenges to sustainable development. Their small size in terms of geography, economy, and population, and their limited capacities render them vulnerable to external shocks. The age-old limitations pertaining to natural resources, water, energy, and waste management have been exacerbated by global environmental change. Although their role in causing climate change has been minimal, they are at the frontlines of facing its impacts. Consequences ranging from increased weather variability and intensified storms to sea-level rise and salinization of groundwater pose serious threats to the sustainability of SIDS.

There are 57 countries classified by the United Nations (UN) as SIDS.[1] Most of them are located in the Caribbean and Pacific regions, but a number of them can also be found elsewhere—in the Atlantic and Indian Oceans, in the Mediterranean, and in the South China Sea. The UN has long recognized the special development situation of SIDS. The Barbados Programme of Action (BPOA), adopted in 1994 at the Global Conference on the Sustainable Development of Small Island Developing States, identified the unique nature of the vulnerabilities and characteristics of SIDS, including their small size, remoteness, and narrow resource and export base, as well as their exposure to global environmental challenges and to external economic shocks (UN 1994). Since then, a series of conferences under UN auspices has focused on devising concrete ways to further sustainable development in SIDS. For example, the 2005 Mauritius Strategy of Implementation (MSI) built upon the BPOA.[2] In 2014, the Third International Conference on SIDS was held in Apia, Samoa, and the SIDS Accelerated Modalities of Action (SAMOA) Pathways was adopted.[3] Similarly, global processes have taken special note of the situation of SIDS. The 2012 UN report on "The Future We Want" dedicated a section to SIDS, noting that "small island developing States have made less progress than most other groupings, or even regressed, in economic terms especially in terms of poverty reduction and debt sustainability," with member states reaffirming their commitment to providing assistance to implementing BPOA and MSI (UN 2012).

The 2030 Agenda for Sustainable Development, adopted by the UN member states in 2015, also focused on the specific plight of SIDS, especially in terms of climate change and associated issues related to sea-level rise,

[1] https://sustainabledevelopment.un.org/topics/sids/list.

[2] https://sustainabledevelopment.un.org/conferences/msi2005.

[3] http://www.sids2014.org/index.php?menu=1537.

ocean acidification, and other impacts that are particularly affecting low-lying countries and coastal areas (UN 2015b). The attendant Sustainable Development Goals (SDGs), in particular Goal 13 (Climate Action) and Goal 14 (Life Below Water), are specifically relevant to SIDS. One of the challenges facing SIDS pertains to their generally limited capacities in terms of human and institutional resources. Consequently, the SDGs also call for "raising capacity for effective climate change-related planning and management."[4]

Effective implementation of sustainable development strategies calls for effective monitoring and evaluation (M&E) in order to determine that processes are on track and that interventions, policies, and strategies are leading to desired change. Evaluating sustainable development in SIDS must take into account the economic, social, and environmental dimensions, while dealing with the considerable risk and uncertainties caused by global climate change, as well as possible discontinuities and tipping points in environmental trends.

Establishing effective M&E systems requires systematic effort and overcoming capacity constraints. An evaluation conducted by the Global Environment Facility Independent Evaluation Office (GEF IEO) in the Pacific found that all GEF projects have M&E protocols, and that the systems have been used effectively for adaptive management in the context of the projects. Yet institutionalizing M&E within the regular operations of the involved ministries and departments has proven challenging, primarily due to limited capacity (GEF IEO 2015). An evaluation in the Caribbean had similar conclusions: while project-level M&E has improved over time and has clearly contributed to adaptive management, environmental monitoring and the assessment of impact-level results have been extraordinarily challenging (GEF IEO 2012a). The reasons for this include a lack of baseline data as well as systematic monitoring data for assessing environmental trends over time. Other evaluations confirm these findings. For instance, in Timor-Leste and Jamaica, M&E has played a very limited role in managing the GEF portfolio and in providing environmental data to aid decision making (GEF IEO 2012b, 2012c).

As is evident from the analysis that follows, a number of factors hamper institutionalizing M&E in SIDS. One is simply the small size of governments and their capacities, combined with the fact that M&E is often seen as an external requirement related to donor-funded projects, thus rendering it low on the list of priorities. There are also significant challenges pertaining to data availability and the capacity to collect relevant data.

This chapter draws upon experiences in monitoring and evaluating sustainable development in the Pacific and Caribbean SIDS. It focuses on systemic and capacity constraints that need to be addressed in order to make M&E a useful tool for governments. In particular, evaluation is often seen mostly as being imposed by donor-funded projects or by regional organizations. It is important to overcome this perception in order to increase the utility of evaluation in these regions.

[4] Target 13.b.

SUSTAINABLE DEVELOPMENT IN SIDS IN THE ERA OF CLIMATE CHANGE

Recognition of the specific vulnerabilities of SIDS is not particularly new. In 1994, UNESCO's Island Agenda asked rhetorically, "Is Paradise an island?" (UNESCO 1994, 8), and outlined the various challenges facing small islands, ranging from small and subsistence economies to cultural issues and the mixed blessings of tourism. The report also identified limited natural resources, such as fresh water and energy, and the need for conservation of coastal and marine systems and unique island biodiversity as significant constraints to development. Vulnerability to natural disasters, such as cyclones, earthquakes, and tsunamis, and sea-level rise were identified as significant risks to small islands. Global warming, ocean circulation patterns, and climate variability were also mentioned, although at that time they did not yet receive major attention (UNESCO 1994). Similarly, the United Nations University World Institute for Development Economics Research presented a report to the Barbados Global Conference (UNU/WIDER 1995) that focused on overlapping issues. A chapter on natural disasters detailed the impacts on economic and social infrastructure in SIDS (Obasi 1995). Based on contemporary knowledge, the chapter was careful to note that there was no evidence that there had been an increase in tropical cyclones or their intensity due to climate change, but it also noted the increased risks associated with sea-level rise and its potential impacts on freshwater resources and coral reefs (Obasi 1995).

As the body of evidence grew, the international scientific community became more confident in stating that global climate change posed a major threat to low-lying coastal countries everywhere, and SIDS in particular. It was noted that SIDS were the first to pay the price for a problem that they had hardly contributed to (Pelling and Uitto 2001). Today we read reports of the dramatic effects of rising sea levels in island nations. An analysis of aerial and satellite images between 1947 and 2014 shows that in the archipelago of the Solomon Islands, five islands ranging in size from 1 to 5 hectares have already disappeared under rising seas, and another six islands have shrunk by 20–62 percent (Albert et al. 2016). Research appears to indicate that tropical cyclone frequency is decreasing, while the intensity of the storms appears to be increasing in ways that may create geomorphological change in the islands (Kelman 2016).

The 2015 United Nations report on SIDS focused exclusively on climate change, pulling together data from publicly available sources concerning the impacts on island nations (UN 2015a). The report identified serious threats to economic sectors that are essential for many islands, notably fisheries and marine resources, and tourism. It also highlighted how climate change impacts affect the social sector, including public health, food security, migration and displacement, and natural and cultural heritage. The report noted that the average annual losses from climate change are proportionally highest in SIDS: it is estimated that annual climate-related losses in Vanuatu are about 6.5 percent of gross domestic product (GDP). The cost of inaction is also high. If governments decide against any action toward climate change

adaptation, it is projected that the annual losses in the Caribbean will rise to $22 billion—10 percent of the current size of the Caribbean economy—by 2050. In the Pacific, the total value of infrastructure, buildings, and cash crops at risk from climate change is estimated at $111 billion (UN 2015a).

To address the above challenges, reliable data, systematic monitoring, and credible evaluation of the effectiveness and efficiency of policies, as well as their impacts, will be crucial. The following sections discuss issues pertaining to evaluation capacities in the Pacific and Caribbean island nations.

PACIFIC SIDS: DEVELOPING NATIONAL EVALUATION CAPACITY IN THE CONTEXT OF DIVERSITY

The nature of Pacific SIDS presents a special case for achieving and evaluating sustainable development.[5] SIDS in the Pacific collectively comprise several hundred islands with remarkable geographic, socioeconomic, environmental, and cultural diversity, spread out over millions of square kilometers of the Pacific Ocean. Numerous threats throughout the region that have the potential to affect sustainable development include climate change; rapid urbanization that is spurring social inequality, deterioration of fragile biodiversity and ecosystem services, fresh water degradation, and resource depletion from growing populations (Duncan 2011; UNEP 2014). Furthermore, Pacific SIDS are often categorically described as being ill-equipped to face these challenges due to their remoteness, small size, and limited access to resources. However, the degree to which these and other factors affect sustainable development and the ability to react to change varies widely from island to island. When speaking of sustainable development in SIDS, it is important to be mindful that different SIDS face different challenges and opportunities. In the first of the following sections, the diversity of Pacific SIDS is emphasized, and the implications for evaluation policy are discussed.

Sustainable development requires sustainable evaluation: this is best achieved when countries have ownership and control of their own evaluation processes, as required by the SDG framework. The governments of Pacific SIDS face difficult challenges in owning evaluation: understanding their governance structures and processes is an important early step in planning how to fit an evaluation system into their specific contexts. The second of the following sections outlines the case for local evaluation of sustainable development, and provides a brief look at governance structures in Pacific SIDS.

Diversity of Pacific SIDS

A quick glance at geographic and demographic statistics of Pacific SIDS reveals stark differences between countries and territories. For example, the "small" in "small island developing states" has an indefinite meaning:

[5] The Pacific SIDS are American Samoa, Cook Islands, Fiji, French Polynesia, Guam, Kiribati, Marshall Islands, Federated States of Micronesia, Nauru, New Caledonia, Niue, Northern Marianas, Palau, Papua New Guinea, Samoa, Solomon Islands, Timor-Leste, Tonga, Tuvalu, and Vanuatu.

land areas range from more than 450,000 square kilometers in Papua New Guinea to just 21 square kilometers in Nauru. The distribution of human settlements also varies widely, with the majority of Palauans and Marshallese living in central urban areas, whereas more than 70 percent of Ni-Vanuatu and Solomon Islanders live in rural areas spread across dozens of islands. People living on large volcanic islands often depend heavily on surface water sources and their associated ecosystems, while these do not exist on small atolls, where people depend instead on coastal and lagoon ecosystem services.

Levels of social and economic development also differ substantially. Five Pacific SIDS are classified by the UN as least developed countries—that is, countries that exhibit the lowest indicators of socioeconomic development. Meanwhile, some territories, such as New Caledonia and French Polynesia, are relatively developed, with a per capita GDP close to that of New Zealand.

Evaluation policy toward sustainable development in Pacific SIDS must be flexible enough to respond effectively to the variety of settings in which they exist. This requires representation from each of the Pacific SIDS so that they can explain their sustainable development objectives, and learn how evaluation resources can be made available to them, and used coherently. Fortunately, multiple sustainable development agendas for SIDS have already been put forward (e.g., BPOA, MSI, and the SAMOA Pathway), and attempts are being made to improve the platforms for regional cooperation in the Pacific, for example with the Pacific Plan and the Melanesian Spearhead Group (Pacific Plan Review 2013). These are useful entry points for the international evaluation community to promote and support the implementation of evaluation systems. These sustainable development agendas and regional platforms currently highlight the importance of evaluation, but there is room for more operationalization in the context of individual SIDS.

One evaluation challenge unique to some SIDS is the small size of governments. Pacific SIDS with very small population numbers often struggle to assess and report on all of the indicators set by external agencies. Human resources may be limited by the small number of government staff, or by a "brain drain" problem, as in the Cook Islands and Niue, whose inhabitants are automatically granted citizenship in developed countries that offer better economic opportunities. Careful thought must be given to prioritizing those SDG targets and indicators that can realistically be monitored.

Despite great diversity across Pacific SIDS, they do share important cultural characteristics that must be respected by external agencies that are looking to build their capacity for evaluation. Pacific SIDS have some of the highest levels of indigenous populations in the world, and among these populations, collectivist values are strongly maintained (Koshy, Mataki, and Lal 2011). As a result, many Pacific islanders strive toward achieving harmony, respect, risk avoidance, and loyalty within institutions even at the expense of efficiency and assertiveness as valued by Western institutions. There is a need to build capacity for evaluation in Pacific SIDS, but external facilitators must be sensitive to the way indigenous values affect the willingness and preferences of the islanders to develop this capacity. (For in-depth discussion of this topic, see Rhodes 2014).

Strengthening Government Capacity for Evaluation

Evaluation is more likely to be sustainable if the governments of SIDS are put in control of the evaluation processes and systems. Country- or government-led, as opposed to donor-led, evaluation is believed to help create a culture of evidence-based decision making, and to better reflect the information needs and values of the country stakeholders (Segone 2009). Establishing a sense of ownership over the systems can also help motivate government authorities to maintain evaluation processes, and to apply the results.

With respect to sustainable development, there are additional reasons to encourage government-led evaluation. First, not all sustainable development efforts are driven and funded by donors. For example, Pacific SIDS generally take the threat of climate change very seriously, and have their own sustainable adaptation initiatives and the associated information needs. Governments of Pacific SIDS should be supported in developing evaluation systems that they can use for their own endogenous purposes. Next, the social, economic, and environmental dimensions that make up sustainable development are adaptive and constantly changing. Unlike time-bound bilateral and multilateral programs, government is a permanent fixture, and it is in a better position to assess levels of sustainability over long periods of time if monitoring mechanisms are institutionalized.

Most sustainable development agendas influencing Pacific SIDS are conceptualized at the national or international levels, but governments will also need to think locally. Understanding interactions between human and environmental systems is key to assessing and evaluating sustainability (Rowe 2012), but these interactions are often highly context-specific. Therefore, in addition to top-down evaluation strategies, bottom-up strategies that can capture local nuances of sustainability are also needed. It will be beneficial to consider how local government in Pacific SIDS can be included and strengthened to fill this need.

Many Pacific SIDS governments have adopted decentralization policies, and several have constitutional provisions for local government. The scale of local government that exists varies widely (Hassall and Tipu 2008). In the Solomon Islands the local government is only decentralized to the level of provincial and municipal councils each of which currently oversee only tens of thousands of people. Meanwhile, individual islands and island groups in Kiribati have their own councils, some of which may oversee as little as a few hundred people. Some of the larger Melanesian SIDS have tiered national/regional/local government structures, while all government is maintained centrally in microstates like Niue and Nauru. Consideration of potential M&E responsibilities, and the mapping of the potential flow of information through disparate government structures is an important early step in assessing the potential for a comprehensive evaluation system.

Another interesting aspect of Pacific SIDS is that traditional or customary governance systems that date back to precolonial times are often blended with democratic governance systems and empowered through legislation (Hassall and Tipu 2008; Hassall et al. 2011). For example,

village councils composed of the heads of extended families are granted administrative power by the state in Samoa. In Tuvalu, elected local officials are accountable to a traditional assembly of elders who are given power by the state to oversee local affairs. These cases offer interesting opportunities for evaluation at the local level because the methods of inquiry could coalesce with traditional forms of engagement to produce rich, useful data that otherwise might be missed by top-down national approaches.

In order for evaluation systems to be effective within Pacific SIDS governments, they will need to be crafted appropriately to fit a variety of contexts, while working within on-the-ground realities. Sustainable development is a complex issue that plays out at not just at the global and national levels, but also at local levels, where traditional forms of governance are still common. Evaluation approaches need to be adaptive enough to handle both the complexity and variety found in Pacific SIDS, and should respect cultural heritage (Hoey 2015). Furthermore, local governments in Pacific SIDS generally are underresourced financially, and for the smallest islands this will likely continue for the foreseeable future (Hassall and Tipu 2008). The challenge going forward is not just to demand more evaluation and to offer training, but to work with Pacific SIDS stakeholders to identify which forms of evaluation make sense in each of their given contexts.

EVALUATING THE EFFECTS OF CLIMATE CHANGE ON SUSTAINABLE DEVELOPMENT IN CARIBBEAN SIDS

The SIDS of the Caribbean are among the world's most vulnerable countries when it comes to the effects of climate change (Todd 2011, 2013; UNFCCC 2005). This could become critical to their social and economic development, as well as to their terrestrial, coastal, and marine environments if no appropriate action is taken. Many islands are threatened by increases in the number and severity of extreme weather events, rising sea levels and coastal erosion, coral bleaching, and damage to biodiversity. At its worst, climate change could result in substantial loss of life and the damage to property and infrastructure that can easily cripple small and fragile economies. The Caribbean SIDS comprise a substantial part of the membership of the Caribbean Community (CARICOM), a regional political and economic grouping of some 20 member and associate member states. The World Bank has estimated that about 11 percent of the total GDP of all 20 CARICOM countries could be adversely impacted annually by climate change (Toba 2009; see also IDB 2014; World Bank 1997).

Although the Caribbean SIDS are not high greenhouse gas emitters on the world stage, they do have opportunities for climate change mitigation, and many countries are pursuing them through such activities as improved energy efficiency of buildings; increased production of solar power; the use of household equipment such as solar water heaters; and the scaled-up adoption of electric vehicles.

A Common Approach for Increasing Climate Change Resilience, Adaptation, and Mitigation among Caribbean SIDS

In 2009, through the Liliendaal Declaration,[6] CARICOM leaders presented their vision of a common regional approach that would enhance resilience and adaptation by addressing the threats and challenges of climate change on Caribbean society and economy, as well as by providing support for mitigation-related policies and measures. This approach will be delivered through an overarching regional framework that comprises a set of strategic elements, each with defined goals and indicators, which should contribute to the achievement of the framework's overall objectives. Responsibility for designing and managing a common M&E system for the regional framework and its elements is vested in the Caribbean Community Centre for Climate Change (CCCCC), based in Belize.

The common approach is made up of five strategic elements that embody the key objectives contributing to the longer-term goal of building regional resilience and capacity to adapt to climate change:

- **Strategic Element 1.** *Mainstream climate change adaptation* strategies into the sustainable development agendas of the CARICOM member states
- **Strategic Element 2.** Promote the implementation of *specific adaptation measures* to address key vulnerabilities in the region
- **Strategic Element 3.** Promote actions to *reduce greenhouse gas emissions* through fossil fuel reduction and conservation, and switching to renewable and cleaner energy sources
- **Strategic Element 4.** Promote actions to *reduce the vulnerability of natural and human systems* in CARICOM countries to the impacts of a changing climate
- **Strategic Element 5.** Promote actions to derive social, economic, and environmental benefits from the prudent *management of standing forests* in CARICOM countries

This complex array of strategic elements has generated an associated set of 21 goals. It is intended that these should be monitored and evaluated in one M&E framework applied across the 20 member and associated states (Groupe-Conseil Baastel ltée 2013).

Challenges for Evaluation of the Regional Framework

Establishing, populating, and analyzing such an M&E framework would be a major undertaking for any region, and will pose particularly substantial challenges for many Caribbean SIDS, as outlined below.

[6] https://caribbeanclimateblog.com/tag/liliendaal-declaration-on-climate-change-and-development/.

Challenge 1: Targets, indicators, and data. Progress toward each of the 21 goals should be measured through sets of baselines, targets, and indicators. The prevailing situation with respect to climate change targets is consistent across the CARICOM countries: that is, targets are generally only established for donor-funded projects. In many other cases, particularly for government-sponsored policies and activities, progress indicators have been defined, but are not associated with specific targets. Baseline data are generally available for national and sectoral-level frameworks, as well as for donor-funded projects, and are therefore relatively strong compared with targets and indicators.

Challenge 2: Variable resources for monitoring and evaluation. The data landscape is highly uneven across CARICOM member states. Countries with relatively major economies, notably Jamaica, have more resources to invest in national-level data collection and management than much smaller economies, where public administration has limited human and financial resources. Similarly, countries with many externally supported projects, such as St. Lucia, have greater access to external M&E advice, and the funds to support this function, than countries with fewer projects.

In recognition of this variation, it is important to note that M&E models that have been found effective in relatively well-resourced CARICOM countries cannot simply be transferred to other countries in the region, which have fewer resources. It is therefore important to ensure that evaluation activities and systems assess whether specific countries have made acceptable progress toward climate change targets according to their own unique situations, priorities, and resources, rather than against region-wide standards derived from countries with greater human and financial resources.

Challenge 3: Ability to respond to challenges of the regional framework. Underlying the concept of a unified CARICOM-wide M&E framework is the expectation that each country will have some indicators and targets that reflect the goals of its national priorities and policies. However, a preparatory study for the M&E framework found that many key stakeholders in government ministries dealing with climate change, or in other national bodies with a climate-related mandate, were not strongly aware of the specificities of the regional framework, and still less aware of how it might be evaluated. Furthermore, there is often no clear separation between monitoring activities and evaluation activities, which would only be feasible in much more developed and well-resourced systems. This means that even the limited data collected by monitoring systems are often not suitable for evaluation purposes.

Challenge 4: Low status of monitoring and evaluation. An additional important challenge to overcome is the low status and limited capacity of M&E activities in most of the countries in the region. Monitoring and evaluation are largely conducted in internationally funded interventions, and are often of a regional or subregional nature. Associated with such interventions,

several regional or subregional M&E frameworks have been proposed or established in various sectors (notably disaster management), usually with limited take up at the country level.

Underlying weak national implementation is the substantial inequality and perceived disconnect between the regional bodies proposing M&E frameworks and the national government departments or units that must conduct the detailed work of designing and implementing data collection and analysis at the country level. Characteristically, the regional bodies have sufficient technical capacity, human resources, and funding to participate in complex M&E exercises. Government departments, on the other hand, are often understaffed and poorly resourced, but have substantial implementation and reporting obligations, including some that are mandatory under the requirements of internationally funded activities or conventions. At the country level, therefore, evaluation is currently rarely recognized as useful, and has a correspondingly low functional status.

In the context of the constraints outlined above, the all-embracing M&E framework for climate change is often viewed as a largely unwelcome addition to existing tasks for the relevant government offices at the country level and may, in view of insufficient human and financial resources, be undeliverable. The requirements for national-level capacity building need to be carefully incorporated into the development of the overall regional M&E framework: this would include providing substantial financial resources over a period of time, given the low starting point in many countries.

Advancing Evaluation Capacity

The Caribbean SIDS have limited numbers of experienced evaluators, including those with climate change expertise. The demand for evaluators comes mainly from international bodies, including the countries and institutions that provide funding support to governments and regional institutions. To convert the desire for comprehensive evaluation of the effects of CARICOM policies and strategies to reality will therefore require substantial evaluation capacity development.

However, this capacity development needs to be placed within a fundamental upgrading of the role and implementation of evaluation. Underlying the possibility of such change are the following fundamental questions:

- How can the value and status of evaluation be raised, so that rather than being seen as a function that is of interest only to donors, it is seen as a useful resource for governments?
- How can evaluation support from various external funders be brought together into a coherent package, from its current state of fragmented bits and pieces?
- How can a more equitable and effective balance between the evaluation capacity of well-resourced regional institutions, and those of "shoestring" national government departments and offices be created?

If evaluation is to have any chance of becoming a viable component of the CARICOM regional approach to climate change adaptation, resilience, and mitigation, it needs to be very carefully focused on a few critical issues, and realistically scaled, so that it both appears to be, and is, implementable at the country level.

Evaluation capacity development needs to be part of a comprehensive package that will include in-person engagement of national stakeholders in its development through regional planning meetings of operational staff, and capacity building and financial support for national M&E functions. This process should place increased emphasis on evaluation, rather than focusing exclusively on indicators for results-based management.

FUTURE MONITORING AND EVALUATION IN THE SIDS

Experience in both the Pacific and Caribbean SIDS suggests that monitoring and evaluation as analytical tools are currently underutilized. However, we propose that M&E can be very powerful in promoting sustainable development in the island states, if it is institutionalized at the governmental level and integrated into government policies, strategies, and programs. Monitoring is needed to ensure that the various interventions are on track, and are completing their stated activities on time and in a cost-effective manner. Evaluation is needed in order to ascertain that the policies, strategies, and other interventions are reaching their goals and contributing to sustainable development without causing unanticipated negative consequences.

There are a number of prerequisites in order for this to happen. The first pertains to the fact that currently M&E is seen mostly as a requirement from donors and/or regional organizations. It consequently receives low priority at the level of national government ministries and departments. It is important to change this perception: this will require M&E functions and activities to demonstrate their added value. Building the capacity of national stakeholders to appraise and use evidence will contribute to creating demand for M&E. National stakeholders must be engaged in the development of these capacity-building efforts in order to promote country ownership, and a special focus on the utility of evaluation beyond monitoring indicators is needed.

There is also significant scope for cross-learning between SIDS in other regions: that is, not only in the Pacific and the Caribbean, but also in the Atlantic and Indian oceans, the Mediterranean, and the South China Sea. Existing mechanisms could be used for this purpose, including the Alliance of Small Island States (AOSIS), which has a membership of 44 states and observers covering all oceans and regions. Similarly, the UN manages the Inter-Agency Consultative Group (IACG) on SIDS, which brings together regional SIDS organizations[7]

[7] Caribbean Community Secretariat (CARICOM), Commonwealth Secretariat, Global Island Partnership (GLISPA), Indian Ocean Commission (IOC), Organisation of Eastern Caribbean States (OECS), Pacific Islands Development Forum Secretariat, Pacific Islands Forum Secretariat (PIFS), Secretariat of the Pacific Community (SPC), and Secretariat of the Pacific Regional Environment Programme (SPREP).

as well as international, intergovernmental, and nongovernmental partners, and could ensure that all M&E activities pertaining to the agency work are coordinated. Another existing platform that could be used more effectively to share M&E experiences and lessons learned between the regions is the GEF International Waters Learning Exchange and Resource Network (IW:LEARN),[8] which already provides an established forum for learning shared among a large number of environmental programs and projects across regions.

It is important to tailor M&E systems to the specific situations at hand. One size does not fit all, and top-down approaches that impose uniform frameworks on countries and departments regardless of their needs and capacities are counterproductive. It will be essential to choose the targets of monitoring and the subjects of evaluation carefully, based on the utility of these actions in helping to meet national priorities. It will also be important to adjust and design M&E systems to the particular institutional and cultural systems that are prevalent in each country.

Promoting country-led monitoring and evaluation will best serve the need for countries to manage sustainable development. Sustainable development issues involve balancing social, environmental, and economic costs and benefits: this means making value judgments about what exactly is most important to sustain in each country. Ultimately, it is the countries involved that should be making these judgments, informed by robust evidence and in recognition of the fact that sustainable development looks different at different scales and means different things to different people. Furthermore, climate change, which is one of the biggest threats to sustainable development in SIDS, is insidious because of the uncertainty about how society and nature will react to it in the coming decades. This makes consistent implementation and use of M&E all the more important: SIDS will need to be active learners as they adapt to constantly changing environments.

REFERENCES

Albert, S., J.X. Leon, A.R. Grinham, J.A. Church, B.R. Gibbes, and C.D. Woodroffe. 2016. "Interactions between Sea-level Rise and Wave Exposure on Reef Island Dynamics in the Solomon Islands." *Environmental Research Letters* 11: 1–9.

Duncan, D. 2011. *Freshwater Under Threat: Pacific Islands.* Bangkok: United Nations Environment Programme.

GEF IEO (Global Environment Facility Independent Evaluation Office). 2012a. *Cluster Country Portfolio Evaluation: GEF Beneficiary Countries of the OECS (1992–2011).* Evaluation Report No. 72. Washington, DC: GEF IEO.

———. 2012b. *GEF Country Portfolio Study: Timor-Leste (2004–11).* Evaluation Report No. 77. Washington, DC: GEF IEO.

———. 2012c. *GEF Country Portfolio Study: Jamaica (1994–2010).* Evaluation Report No. 66. Washington, DC: GEF IEO.

———. 2015. *GEF Country Portfolio Evaluation: Vanuatu and SPREP (1991–2012).* Evaluation Report No. 98. Washington, DC: GEF IEO.

[8] http://www.iwlearn.net.

Groupe-Conseil Baastel ltée. 2013. "Monitoring and Evaluation Instrument and Reporting Framework in the Caribbean." Prepared for the Caribbean Community Centre for Climate Change.

Hassall, G., and F. Tipu. 2008. "Local Government in the South Pacific Islands." *Commonwealth Journal of Local Governance* 1: 6–29.

Hassall, G., M. Kaitani, F. Tipu, and L. Wainiskesa. 2011. *Comparative Study on Local Government and Traditional Governance Systems Interaction in Pacific Island Countries*. Suva: University of the South Pacific.

Hoey, L. 2015. "Show Me the Numbers: Examining the Dynamics between Evaluation and Government Performance in Developing Countries." *World Development* 70: 1–12.

IDB (Inter-American Development Bank). 2014. *Building Resilience and Reducing Emissions*. Office of Evaluation and Oversight. Washington, DC: IDB.

Kelman, I. 2016. "Science Shows Climate Change Impact in the Pacific." *Australian Journal of Emergency Management* 32 (1): 15–16.

Koshy, K., M. Mataki, and M. Lal. 2011. *Sustainable Development—A Pacific Islands Perspective*. Apia: UNESCO.

Obasi, G.O.P. 1995. "Natural Disasters and Sustainable Development of Small Developing Islands." In *Small Islands, Big Issues: Crucial Issues in the Sustainable Development of Small Developing Islands*, United Nations University World Institute for Development Economics Research (UN/WIDER), 63–83. World Development Studies 1. Helsinki: UN/WIDER.

Pacific Plan Review. 2013. "Report to Pacific Leaders." Pacific Islands Forum Secretariat, Suva.

Pelling, M., and J.I. Uitto. 2001. "Small Island Developing States: Natural Disaster Vulnerability and Global Change." *Environmental Hazards: Human and Policy Dimensions* 3 (2): 49–62.

Rhodes, D. 2014. *Capacity across Cultures: Global Lessons from Pacific Experiences*. Fairfield: Inkshed.

Rowe, A. 2012. "Evaluation of Natural Resource Interventions." *American Journal of Evaluation* 33 (3): 382–92.

Segone, Marco. 2009. "Enhancing Evidence-based Policy Making through Country-led M&E Systems. In *Country-led Monitoring and Evaluation Systems*, Marco Segone, ed. New York: UNICEF.

Toba, N. 2009. "Potential Economic Impacts of Climate Change in the Caribbean Community." LCR Sustainable Development Working Paper No. 32. World Bank, Washington, DC.

Todd, H. 2011. "Children and Climate Change: Action for Prevention, Mitigation and Response to Fulfil the Rights of Children to Education in Emergencies." Final report submitted to UNICEF, Barbados.

———. 2013. "Disaster Risk Management, Climate Change and Food Security in Barbados." Final Report to the Food and Agriculture Organization of the United Nations.

UN (United Nations). 1994. "Report of the Global Conference on the Sustainable Development of Small Island Developing States." A/CONF.167/9.

———. 2012. "The Future We Want." A/RES/66/288.

———. 2015a. *Small Island Developing States in Numbers: Climate Change Edition*. New York: Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and Small Island Developing States (UN-OHRLLS).

———. 2015b. "Transforming Our World: The 2030 Agenda for Sustainable Development." A/RES/70/1. UN General Assembly, New York.

UNEP (United Nations Environment Programme). 2014. *Emerging Issues for Small Island Developing States*. Nairobi: UNEP.

UNESCO (United Nations Educational, Scientific and Cultural Organization). 1994. *Island Agenda: An Overview of UNESCO's Work on Island Environments, Territories and Societies*. Paris: UNESCO.

UNFCCC (United Nations Framework Convention on Climate Change). 2005. *Climate Change, Small Island Developing States*. Bonn: UNFCCC Secretariat.

UNU/WIDER (United Nations University World Institute for Development Economics Research). 1995. *Small Islands, Big Issues: Crucial Issues in the Sustainable Development of Small Developing Islands*. World Development Studies 1. Helsinki: UNU/WIDER.

World Bank. 1997. "Caribbean Planning for Adaptation to Global Climate Change Project." Project document. World Bank, Washington, DC.

Chapter 9

Evaluation Standards for Latin America and the Caribbean – Experimentation and Evolution

Ana Luisa Guzmán and Warren Crowther

Abstract. *This chapter reviews an excellent contribution to the development of evaluation standards for Latin America and the Caribbean developed by the Latin American and Caribbean Network of Monitoring, Evaluation and Systematization (ReLAC). Lessons learned from experimentation in applying a code of ethics in the development of seven diverse and ambitious projects over a 10-year period in Costa Rica are discussed. The basic principles covered by the standards are posited, and the practical relevance, viability, and trade-offs, or relative emphasis in the evaluation and decision-making aspects of the projects, are analyzed. The proposed standards for Latin America and the Caribbean constitute an excellent guideline for applying and implementing the improvement of evaluation processes. This observation and analysis suggests an agenda for further debate, including the inclusion of additional evaluation standards.*

Ana Luisa Guzmán, Latin American and Caribbean Network of Monitoring, Evaluation and Systematization (ReLAC) and University of Costa Rica, guzmaneste@hotmail.com; Warren Crowther, Independent Consultant, warren13@racsa.co.cr.

The Latin American and Caribbean Network of Monitoring, Evaluation and Systematization (ReLAC) has made an important contribution to the experimentation and evolution of standards for program evaluation in the region. In the first stage, many evaluators were consulted and a literature review was carried out regarding competencies and standards, including the evaluation standards used in United Nations (UN) agencies, internationally recognized partnerships, and other such initiatives In 2015, a draft proposal of standards was circulated among and commented upon by interested members of the network, and in 2016 a published edition of the Evaluation Standards for Latin America and the Caribbean (EEALC; Rodríguez et al. 2016) was circulated and commented on by interested organizations and professionals.

During this period of elaborating and sharing criteria for the standards, a review of elements of ethical guidance—based on the code of ethics published by UNICEF (2002) for research, monitoring, and evaluation of programs and projects taking into account the rights and interests of children and adolescents—took place in Costa Rica.[1] Its basic principles can be applied to the target populations in the design and evaluation of the programs and projects catering to their rights and interests, and the use of action research methodologies. These same principles were adapted for other target populations, including low-income families needing assistance with initiatives in entrepreneurship, exceptional patients with serious illnesses, older adults, and communities and users affected by transportation and communications investments.

The experience with these projects is instructive regarding the need and conditions for an adequate set of standards, adapted to Latin American and Caribbean realities, while recognizing that although Costa Rica has a mixed population when it comes to ethnicity, urban and rural culture, economic conditions, and environment, it still represents only a fraction of the diversity that needs to be considered in the entire region.

The lessons learned from the demonstration projects are summarized in table 9.1. They highlight implications for the evolution of program evaluation standards for the region.

The most interesting of these experiences illustrate the ample options and major obstacles to some of the critical factors of quality monitoring and evaluation (M&E) of policies, programs, and projects (PPP), which not circumstantially have received the most intensive attention and development in the discussions regarding the EEALC.

Between the initial proposal and the latest version of the EEALC, there has been an impressive and promising advancement in the definition of these standards, particularly related to four major themes: relevance to decision making; influence on decision making; validity of the evidence; and the fulfillment of preconditions. Table 9.1 illustrates the importance and value of these themes, as well as some nuances or complications that have not yet been

[1] In addition to UNICEF (2002), attention was also given to two other norms: UNICEF Costa Rica (2000) and ICAP and UNDP (2003).

broached in the standards, and are in constant evolution as their application is being monitored.

Some of the projects are reviewed in table 9.1, regarding what was intended, what resulted, and what was learned. These are all demonstration projects, initiated by the ProHumana21 Foundation between 2004 and 2014. They were undertaken with intensive grassroots participation and attempts to create prototypes or models of better practices, including more adequate impact assessment.

USING EVALUATION FOR DECISION MAKING

This is mentioned in the EEALC: it is an important starting point, but it needs to be reviewed in order to consider programs and projects as subject to constant experimentation, and not just implementation and resolution of predetermined terms of reference and premises. There are always both positive and negative surprises that affect the engagement, dimensions, and parameters of the M&E and decision-making processes of projects and their impacts. The projects reviewed in table 9.1 emphasize this.

It became apparent that in the implementation of these projects, there was no relationship between the relative gravity or threat of the situation discovered with innovative evaluation methods, and the amount of local and/or national interest that existed to deal with those facets of that situation, and in defining the projects´ terms of reference. Thus a focus on the unexpected turned out to be more revealing—in either an instructive or a disturbing way—than a focus on the achievement of pre-established goals.

For example, it was observed that the resolution of many socio-environmental conflicts and medical pathologies led to unanticipated delayed effects in the form of new conflicts or pathologies. The apparent resolution of one socio-environmental conflict often leads to the unraveling of other conflicts, and the remission of one illness is often the preface to a later downturn in health. Thus, the adoption of a solution for one challenge can often create another challenge.

Another insight from analyzing these projects resulted in fostering comparative research and evaluation, which involved ambiguity in choosing the models and indicators that were sensitive to particular cultural and local realities, and those that were emphasized when comparisons were made. Yet, ironically, comparisons can help rather than hurt in this situation. The greatest "relevance" of an evaluation is not only in appraising how well preconceived challenges and the consequent predefined goals are being met, but also in identifying or uncovering the locally or presently unseen or unexpected factors: in this regard, comparative evaluation can actually be very helpful in making those factors more visible or well anticipated.

The periodic Latin American Development Administration Congresses (CLAD) are very instructive in comparing experiences: however, in the most recent congress, held in Chile in 2016, there was a lack of communication between two main streams of work. There were several sessions dedicated to reviewing advances in PPP evaluation, including discussion of the EEALC. Other sessions were focused on program and project "innovation and

government laboratory" advancements: yet it was noticeable that very few people who attended the evaluation sessions also attended the innovation sessions. Unfortunately, there was no discussion about the importance of assimilating the benefits of these approaches. We will readdress this perspective when we consider the importance of evaluating impact and not just results.

HOW EVALUATION AFFECTS DECISION MAKING

The projects in Costa Rica have used action research methodologies to integrate, at the local level, the evaluation and decision-making processes. These projects have also invited—or pressured and educated—the target populations and other present stakeholders who sometimes met with resistance, to broaden their perspective on the relationship between evaluation and decision making, and how to most positively engage in them. Attention was given to questions of confidentiality and sensitivity of the evidence, and judgments in the evaluation, which prior to the evaluation should be agreed upon by key stakeholders.

In the projects reviewed in table 9.1, a general benefit was enhancement of the knowledge management capacities of the most active among the target populations, and at times of other stakeholders as well. However, this meant focusing more attention on the sensitivities—and sometimes uncomfortable ethical implications—of this knowledge management sharing and decision-making responsibility and privilege, and of the evidence itself. A major difficulty arises where a "culture of shame" prevails, meaning a general avoidance of sharing what could be sensitive or embarrassing information.

The common withholding or suppression of evaluation results was even more tense and awkward: this included evidence considered by one or more parties to be offensive to their agenda or interests, or potentially damaging to other parties.

LIMITATIONS OF EVIDENCE VALIDITY

In the projects reviewed in table 9.1, it became evident that the more intimate and closely connected the evaluators were to the local details and consequences of PPP implementation, the greater the difficulties they faced in identifying evaluation indicators that were both suggestive of validity and of applicable practicality in the evaluation. This was also illustrated by empirical results highlighted throughout the analyzed literature review. That is, the indicators were more geared to their relatively easier application and to acceptability by data sources than to the pertinence of the evidence for the variables that were posited in each case. This review depicted inconsistencies between the connotations of the variables outlined in the conceptual or modeling configuration of what was being studied or evaluated, and what was evidenced through the practical indicators.

Pertinent to this are proposals regarding quality criteria for the selection of indicators. Somewhat different examples of such proposals have been

offered by Stockmann and Crowther, both of whom have shown that "ample criteria" included some that in practice are not compatible with each other (Crowther 1999; Stockmann 2011).

This means that inevitably, in all evaluations, the predispositions of the evaluator (or decision makers) play a role regarding the overall criteria used for selecting indicators.

THE REQUIRED PRECONDITIONS FOR A PARTICIPATORY EVALUATION

For almost any reputable and useful evaluation, considerable work has to be done just to create the preconditions. This has been the case for decades in Latin America, where the insistence on a precipitated or institutionalized evaluation procedure without these preconditions has come to denigrate the credibility of evaluation. Often those most affected—that is, either those being most closely evaluated or the major sources of evidence—might see program evaluation as, for example, lying, "tattling," an intrusion. Or they might simply resist it altogether.

OTHER CONSIDERATIONS AND ISSUES

Different initiatives for developing evaluation standards have taken various positions regarding the four major themes, and constructive attention has been given to them in EEALC.

It is notable that with the exception of the requisite conditions for the evaluation, the United Nations Evaluation Group (UNEG), in its 13 norms, tends to assign less attention to these four standards, and instead highlights the ones that EEALC has identified as the most difficult ones to resolve. For example, UNEG stresses impartiality, and many in UNEG would see impartiality as supporting and strengthening validity. However, UNEG is not explicit about the inherent limitations of evidence and judgment validity for example, regarding issues of sensitivity and confidentiality. On this last point, the American Evaluation Association's Program Evaluation Standards more realistically assigns importance to "technically adequate" information.

In applying these standards, a general lesson from the projects shown in table 9.1 is that their compliance is fostered with the application of action research principles and methodology. In addition to complex planning and implementation of the evaluation process, more than the usual amount of human and economic resource dedication is also required. The target populations and other stakeholders are not readily prepared for the protagonist role they are invited to, or that they need to, assume: much preparation for this role is needed.

On the other hand, there are five more themes with which the projects faced more troublesome practical challenges, and for which, optimistically, the evolution of EEALC may provide solutions. These themes are: involvement of the principal stakeholders in decisions about the evaluations; examination of the underlying values influencing the evaluations; impact; evaluator competencies; and the role of context.

DECISION MAKING: WHO DECIDES WHAT?

The EEALC indicates that the main target population of the PPP are "groups affected by" the evaluation. Regarding the very diverse sets of responsibilities and/or privileges they should have in the evaluation and decision-making process, 16 different categories or segments of actors or stakeholders are designated, with some overlapping regarding both groupings and functions.

For the projects shown in table 9.1, the most interested segments of the population could be readily identified. However, some of the interested stakeholders demonstrated some degree of resistance to sharing or conceding their personal incidence in the decision making or real, informed participation in the design and evaluation of PPP, to those most affected by them. This resistance was less than expected, although it varied among localities and subcultures. With PPPs directed at the rights and interests of adolescents and youth, there was less adult centralism than anticipated, but there was much more than expected regarding PPPs directed at older adults, including among the professionals attending this generation.

UNDERLYING VALUES, OR CRITERIA, INFLUENCE WHAT IS BEING EVALUATED

In the projects reviewed in table 9.1, the agendas of key stakeholders often did not coincide with the ethical standards the projects were trying to adopt: or they were not compatible with each other regarding the underlying values or criteria to be applied.

In these projects, there were numerous cases of unpreparedness for new challenges, and others of institutional units adhering to their agendas and general priorities, and seeking support for more coverage and technological upgrading rather than pertinence and realism in what they were offering. In such cases, substantive evaluation could seem very threatening.

The inherent or inevitable incidence of the evaluator's (or decision maker's) predisposition regarding the selection of indicators is presented in all areas of evaluations, since an external referent to the situation being evaluated is always being introduced either implicitly or explicitly -reporting that things go well or badly, better or worse, optimistically or fearfully, and so on. (Perhaps the exception to this is the response "accepted or tolerated" or "not accepted or tolerated.") One can adopt a baseline or benchmark to show apparent progress or regression, but there is always that important question, which is usually unanswered—"In comparison to what?"—which is needed in order to conclude whether a "reasonable expectation" is being met. Where does this "reasonable expectation" come from? Not only is the lack of reference to a defined reasonable expectation a common deficit in evaluation: it is also lacking in almost all social, educational, and administrative research.

TOWARD EVALUATING IMPACT

When comparing proposed evaluation standards, it is important to consider the connotations and meanings in different languages of the terminology

being adopted. In fact, the same term can have different connotations even in different countries that speak the same language. That said, in the evolution of regional standards in Latin America and the Caribbean, the most consistently affirmed values, or criteria, have been utility, precision, pertinence, and timeliness.

A most concerning tendency in the region is placing much more emphasis on evaluating *results* than on evaluating *impact*, although increasing attention is being given to the latter. The positive side of the emphasis on results is that it gives more attention to the democratically elected government's compliance to the promises they have posited, or the expectations they have raised. But a downside is that it can lead to efforts to improve a PPP, or do a better job of managing it, when a real impact evaluation may indicate that it shouldn't be taking place at all, or that it needs major reform.

Impact as such is not mentioned in the EEALC, but there are interesting hints of it in the insistence that there should be a positive contribution to decision making, and the "quality of life" of the "beneficial public" and "other interested parties."

The UNEG norms make seven references to the evaluation of impact as an alternative to evaluating outcomes: output, relevance, efficiency, effectiveness, sustainability, value-for-money, and client satisfaction. But those norms don't really consider what impact as such entails.

For the projects summarized in table 9.1, a strong definition of impact has been adopted, which includes the added value of the target population regarding the most fundamental transformations of its quality of life; the target population having good knowledge of and insight into what *constitutes* and influences the quality of their lives; and consideration of the opportunity costs of all interested or affected stakeholders. This definition is not very operative, except as a point of reference to discuss with the key stakeholders, including the target populations. It means not just comparing yesterday or today with tomorrow, which is the usual manifestation adopted for a baseline: this kind of benchmarking needs to project *what was, or is, likely to occur* in a given period of time *without* the intervention that is being evaluated.

WHAT COMPETENCIES SHOULD THE EVALUATOR HAVE?

This is a delicate topic: many people who have engaged in evaluation diligently and continuously do not begin to have the qualifications elaborated on in texts such as that of Rodriguez et al. (2016). Furthermore, some of the standard qualifications are extremely difficult to evaluate, such as whether an evaluator is being "objective" and "impartial." The emphasis on credibility puts this up for grabs: who will be most convincing to those who will be hiring the services of an evaluator? Comparing the terms of reference of such qualifications—including for other projects with the same target populations as those in table 9.1—the differences as to expectations of competencies are monumental.

TAKING CONTEXT INTO ACCOUNT

The importance assigned to "context" is often associated with recognition and respect for local and cultural differences. What is seldom discussed is how much subjectivity is inevitable in determining what "context" to consider, as almost all of the phenomena being evaluated are interrelated with infinite conditions or variables, and there is always a presumption as to where and how to draw the lines.

For Costa Rica and the other 40 countries categorized as high middle income, there is an additional limitation: not counting on think tanks dedicated to analyzing and comparing their commonalities, such as those that exist for the more developed and the less developed countries. There is little systematization of the common problems at this stage of development, which could help contextualize studies in the individual countries.

CONCLUSIONS

It is extremely challenging to create a consensus about more precise standards especially regarding the last five mentioned above—decision making, the influence of underlying values or criteria, evaluating impact, evaluator competencies, and context—including an operative definition of impact.

Nonetheless, UNEG's 13 norms and the 30 standards of the American Evaluation Association's Program Evaluation Standards focus on these very expectations, in some cases mentioning them without really defining them.

These international norms and standards also give major attention to two issues that can be closely interrelated, and which the Latin American and Caribbean standards do not deal with directly: that is, human rights in general, and the well-being of those participating in the evaluation, not just those affected by its results. Human rights, and especially its byproducts of equity or equality, involve constant and inherent trade-offs. Fourteen operative definitions of "equity" that are commonly adopted in practice can be identified; all have substantial legitimacy, but they can also be quite contradictory with each other.

It is necessary to recognize the importance of the learning experiences presented in this work, both in terms of the results and the impacts obtained in the projects developed, and the constant search for high quality in their evaluation. This is especially true in the case of Latin America, where it will help to refine the evaluation standards with greater precision regarding the contexts and characteristics of the programs and projects of the region.

The standards for Latin America are also a guide, a way forward in the design and execution of evaluations that will provide greater knowledge, clarity, and understanding of the contexts in which policies, programs, and projects are being developed, and will thus improve evaluation processes.

TABLE 9.1 **Demonstration projects using research action and participatory evaluation: relevant contributions to evaluation standards**

1. Education and guidance counseling that intends to be more ample, realistic, and personalized, with the development and validation of a website, and didactic modules	
Background	The deficiency in guidance counseling appears to be the biggest generalized abuse of adolescents in the country, leading to much frustration and unfortunate decisions by youth, and contributing to the greatly increasing nonworking youth population and violence. Making it worse has been civic education that has encouraged students to cheat on their civics graduation exams to misconstrue the need for improvements and thus for their positive contributions. It has been especially hard for today's young adults (aged 20–35) who are of a baby boom generation for which the country did not prepare. They are undereducated (only 40% have a high school diploma) and have collided with a depressed employment market
Intent for target population	▪ Emphasis was on the student being able to intelligently evaluate and choose his or her best prospects. ▪ Development of a website on how to choose which university or technical educational majors/specialties are most needed in the labor market and which are accessible given admissions quotas. ▪ Experimentation with guidance discourses and didactic modules to help in choosing high school and university majors and specialties, and to understand the complementary value of virtual education. ▪ Focus was on those entering their last year of high school before choosing university majors when registering for the admissions exam, those in 9th grade who often can change their category of high school, and those who must choose a category on entering 7th grade. ▪ Before giving group and individual feedback to the high school students, survey results were shared with the educators. ▪ Promotion of more forthright civic education that is honest about where youth is needed. ▪ Experimentation was undertaken largely by university graduates and professors of guidance counseling
Results	▪ The survey was undertaken in 38 of Costa Rica's 81 local governments (municipalities), with 19,000 current or potential university and high school students, and the general response of the great majority was gratitude; it was undertaken in 100 mostly high schools, plus a few graduating 6th grade classes of elementary schools in low-income neighborhoods. ▪ In numerous cases, the presentation of student survey results surprised the high school educators. ▪ Some high school directors were resentful, because with this information, many of their students changed their schools to a different category; the project did, however, succeed in greatly increasing the enrollment of elementary school graduates as a result of the counseling

Results	▪ In a follow-up study of youth who had received this guidance five years earlier, most had continued with their university studies and judged the guidance as having been helpful; surprisingly, they had been very accurate in anticipating the practical problems they would encounter, such as socioeconomic background and family dynamics in pursuing their preferred prospects. ▪ The elementary school graduates of very low-income neighborhoods were registered in high schools generally not of the student´s preference, but rather responding to the parent´s dominant criteria of security and transportation; therefore, the discourse was altered on this, telling students they could choose a different category when finishing the 9th grade. ▪ Initiation of a comparative study between countries regarding civics education pinpointed the influence of the legitimization or not of militarization (given that Costa Rica has no military)
Lessons learned	▪ The target population was mostly helpful in monitoring the effects and impacts of attempts at better counseling practices. ▪ This has not led to a generalized application of the better practices; attempts were made to identify the major sources of resistance to generalizing their adoption, or even to evaluation of standing procedures, particularly as this resistance was justified by referral to intellectual and organizational autonomy and/or authority that educational institutions or units insist on for themselves. ▪ This resistance was very strong by universities, which are influenced by professional associations and are very slow to create new majors, specialties, and courses, and to resist the notion that guidance counseling should be more realistic—which is needed to break the tendency that the most-solicited majors are the ones leading to severe underemployment, and those majors highlighted for meeting international accreditation standards rather than for their national relevance, while there are great needs of determined specialists often in the same professions. ▪ The surge of new categories of high schools caught the existing elementary and high schools unprepared, and the latter often put up obstacles against more free choice for the students. ▪ The Public Education Ministry was adamant in not addressing or even wanting to hear about conflicts within its own units that are affecting student decision making, such as the conflict between civics education, apparent entrepreneurship programs, and realistic counseling; however, with a change of government, the ministry has indicated interest in taking lessons from this experience. ▪ The project allowed the students a role of censorship: it was important to recognize that many students—more of middle-class than low-income neighborhoods—were emphatic about not wanting their parents involved in the evaluation and selection of a category of high school or of a university major; where possible, counseling was offered to parents, but insisting that the students could choose whether to extend the invitation to their parents

2. Identification and creation of windows of opportunity and necessity for professional and technical human resources in the zone of residence, especially of youth, to reverse their tendency to identify their best future prospects as being distant from that zone of their country, and the adaptation of formal and parallel education to prepare them for what is most relevant to where they live	
Background	Local opportunities and the needs for such competencies, more as entrepreneurship than as employment, are generally ignored in community development and education plans. Most rural and many suburban communities export the resources they most need for their development, i.e., their more educated youth, even though their vocational interests may coincide with needs or potential opportunities in the local zone. Meanwhile, university and high school students undertake their required field projects, including those of community service, with very little (and at times negative) consequences for the community or target population; the discipline; or their own professional portfolios.
Intent for target population	▪ The identification of the opportunities and need for professional and technical human resources in each zone, using field research and community service projects of university and high school students and community participation workshops to do this. ▪ Incidence in the curriculum to respond to those spaces and prospects of the students to respond to or develop them, and promotion of the use of virtual education. ▪ Field practice for students, e.g., regarding program evaluation, local social environmental conflicts, ecotourism opportunities, and architecture for schools or community services, all promoting the idea of conforming to a niche. ▪ Workshops and follow-up in entrepreneurship, with help in getting the required financial, technical, and institutional support. ▪ Attempts to influence governmental agencies, NGOs, community inter-institutional coordination commissions, and international cooperation to recognize the need for this, and to take lessons learned from the project. ▪ Seek ways to enable the community to have what it lacks now, which is intelligent planning, design, and evaluation of its underlying and most fundamental transformations.
Results	▪ 2,000 students (fewer than expected) receiving guidance counseling in many local governments also participated in field practices related to their vocational interests; many were made more conscious of zonal needs and opportunities for their intended professional or vocational prospects, or those they are interested in. In some cases there has been immediate interest by prospective employers, such as in updated accounting and informatics for commerce of artisanal products. ▪ A few technical high schools experimented with some curricular additions, like educational ecotourism. In one local government, the lowest rated in the UN Life Quality Index, the project helped bring a new technical high school with recently created (and long needed) specialties for such schools.

Results	▪ University students of 18 majors in 6 universities participated, annexing and adapting their field projects to support the investigation and development of development needs and opportunities in the participating communities. Also, many high school students contributed to this with their community service projects. ▪ In the review of Project 3 below, the subject of entrepreneurship is elaborated upon somewhat. ▪ Extensive bibliographies, citing as many as 400 research reports, have been prepared in each of 4 local governments, with copies including many of the cited documents, to local authorities, and to public and school libraries.
Lessons learned	▪ The potential for much more relevant student field practice seemed quite evident, except that it is so very difficult to get the established governmental and NGO institutions to readdress their agendas and methodologies: some are set in their ways, and some are influenced by the agendas of international cooperation. Even some that highlighted youth participation in their discourse were not receptive to receiving and using the interest profiles of local youth that were offered to them. ▪ There were numerous university students interested in undertaking their obligatory field activities (thesis or other graduation projects, internships, or community service) in these subjects, and the project urged them to do this in their family's residential zone. However, the universities in general are very poorly organized for such projects to be relevant in the terms the project has promoted (for the student´s professional portfolio, for the community, and for the discipline). Therefore a number of initiatives were curtailed. ▪ Some organizations have been influenced to take lessons from this experience, and to adapt their agendas and methodologies, but the evident need for this is being met only fragmentally. ▪ The use of our bibliographies of research on local realities (a large proportion being university theses and field reports), highlighting substantial documented local empirical intelligence, has been very sporadic. Both external and local organizations prefer to embark with their preconceptions, immediate experiences, and agendas, and not be distracted by substantiated evidence of local realities.
3. Social and business entrepreneurship, especially of community and agro-ecotourism projects, small businesses for low-income families, and youth initiatives	
Background	Numerous studies of social entrepreneurship show that most projects have less impact than expected, such as in rural community tourism, and family business projects, including those that had received training in entrepreneurship and project development. Most are very deficient in key ways, and are not adequately prepared to deal with critical factors: most family businesses do not survive more than a few years.

Intent for target population	▪ Workshops and monitoring or follow-up (including incubation), with a focus on an operative plan resolving the critical factors of the efficacy of such entrepreneurship, including segmentation of the market or beneficiaries; adjustment of the products' characteristics; realistic financial analysis; the use of information technology in marketing and transaction with the users, complementation among projects instead of overly zealous competitiveness, and in many cases, the advantages of emphasizing the social and ecological benefits of the initiative. ▪ Visiting the location of each project before the workshop, in order to assure better advice from the instructors and better understanding and preparation of the entrepreneurs for what a focus on efficacy entails. ▪ Periodic updating of a database with numerous (more than 60) sources for financing of these projects, with emphasis on donations. What are updated with regard to the financial sources are the themes of interest, conditions for soliciting the funds, and their approximate availability. ▪ Forming of facilitators of entrepreneurship.
Results	▪ Some community entrepreneurship has been undertaken with our help, especially in cultural activity, and much family entrepreneurship as well, especially in tourism, clothing, culinary, beautician, informatics, the value-added chain of agro-products, and health service endeavors. ▪ Numerous workshops have been undertaken, with a huge drop-off of those indicating interest to those with regular assistance. A big factor in the drop-off was our warning that the curriculum was demanding, requiring perseverance and patience (especially with the bureaucracy), and particularly with regard to financial analysis (although many with little formal education were able to process it with our help). ▪ It appears that by far most that have regular assistance in the workshop do continue with their projects, and obtain the needed support (over 90% of the low-income women who solicited funds from a special government fund with our help, compared to less than 20% of requesters in general). ▪ Very few have taken up the offer to form and give diplomas to facilitators of entrepreneurship, although demand for this is very high.
Lessons learned	▪ The project's curriculum has been criticized for being very demanding by various institutions and instructors who offer workshops and incubation in entrepreneurship. However, this exacting curriculum responds to studies indicating how few entrepreneurial projects succeed even somewhat near expectations, and the need to focus on critical factors of efficacy. Thus, an attempt has been made to emphasize more practical help than academic assistance. ▪ There have been very variable differences of grades of commitment and follow-through among the communities and the families that have participated with entrepreneurial projects. The reasons for this are being analyzed by monitoring the perspectives and decisions of the entrepreneurs. There seems to be a complex set of personal, social, economic, and political factors involved.

Lessons learned	▪ A greatly increasing number of families who have serious health problems have participated. They are desperately seeking income to pay for private care, given the huge waiting lists in public health care services. This of course requires great attention to how to address this combination of health and financial crises. ▪ Much depends on the local counterpart organizations that solicited this project, and those that either enjoin or distance themselves from entrepreneurial initiatives.
4. Attention to the increased violence in communities and schools, along with its effects; and of intrafamily violence on the psychosocial, vocational, and civic prospects for youth; and to the need to conciliate the emphasis on containing human violence with adequate preparation for disasters and emergencies	
Background	The major burden of responsibility in the Law of Protection of Children and Adolescents goes to educators in schools, who receive very little help with carrying it out. The programs of prevention of suicide, femicide, and abuse in general do not respond to underlying factors or statistical tendencies. The Manual of the National Emergency Commission with the Ministry of Education on how to prepare for disasters and emergencies, including brigades and simulation, is horribly unrealistic if one takes into consideration lessons learned from the earthquakes in Haiti and Mexico, as well as incidents in Costa Rica.
Intent for target population	▪ Workshops for youth facilitators of peace and security in their schools and communities, in combination with the Ministry of Justice and Peace, Municipalities, and other agencies dedicated to the subject. ▪ Advice to school student commissions dedicated to the theme. ▪ Workshops for educators on how to assist adolescents traumatized by violence, indicating that there is a government agency to help them in case of reprisals. ▪ To create a commission on how to mitigate the huge arrival of so many arms, drugs, fugitives (including pedophiles) into the country, and the impact of the romanticizing of violence, most of this from the United States. ▪ On-site evaluation of how to conciliate the great attention to containing human violence, by enclosing the students and educators in barbwire fencing, small gates, and guards (who are often off on errands), and lack of the open spaces not under roofs and ceilings, with realistic preparation for the mass movement of students, educators, rescuers, hysterical parents, and water, sediments, and lack of electricity in case of a disaster such as an earthquake, fire or huge storm.
Results	▪ Only two local governments responded to the concern about the contradiction between measures of containing human violence and preparation for disasters and emergencies, facilitating an investigation, which made recommendations to 17 schools. However, technical solutions for key problems, such as economical remote locks that the guards can handle at a distance and that are not dependent on the electrical system have not been found.

Results	▪ In two other local governments, there was a surprising regularity of attendance at our workshops for 125 youth facilitators of peace and security in their schools and communities (including students from 25 high schools), given the schedule of 6 all-day Saturday sessions, and the distances to travel that were required. Through gaining confidence with the students, and interviewing them, it was understood that the great majority were at least partially motivated because they have violent internal family situations, and most had not shared that information with anyone. ▪ Naturally, they contributed considerably to penetrating the world of violence in which they are living, and are observing. ▪ The students gained confidence by their preparing 25 videos of their own artistic creations, including themes not presented in the conferences of experts and authorities. Their artistic renditions have been circulated to many schools and local child protection committees. ▪ Advice was given to municipal commissions, educators, and student commissions on their respective work plans regarding violence, and help in articulating with local agencies. ▪ Several high schools enabled the training of educators on how to attend to adolescents traumatized by violence. ▪ The government agency in charge of helping endangered victims and witnesses did not show an interest in an offer to create a video about their services to circulate among educators. ▪ This project was curtailed abruptly, as the Ministry of Education created new protocols, insisting that they were to be "supreme and exclusive," although they were not adapted to the realities discovered in the demonstration projects.
Lessons learned	▪ Organized commissions in two municipalities facilitated the project´s activity of the youth in favor of more peace and security in the community, and officials in two other municipalities facilitated inspections on how to prepare schools better, with periodic evaluation of what was undertaken, until the project was curtailed abruptly. The Public Education Ministry is now, several years later, reviewing what happened in this regard. ▪ The last few years we have attended 11 forums on dealing with violence in communities and schools in Costa Rica, with more than 65 conferences by international experts and national counterparts of executing agencies, of which very few shared lessons of the realities of their field experiences. By far most made propaganda for what was and will be intended. Several key themes or factors in the game were never mentioned in the conferences, and only in a few cases was there opportunity for discussion. ▪ The governmental and university organizations dedicated to preparation for disasters showed no interest in the problem of conflict of measures regarding violence that have put so many students and educators in jeopardy. This is now being undertaken, but the criteria for selecting model schools are very deficient and impractical; and the models have serious defects. Hopefully, this new program will take lessons from what was undertaken by the project.

5. Taking lessons from exceptional patients, with due attention to critical factors that are receiving inadequate attention	
Back-ground	Many recent panel studies of clinical trials of oncological patients in different countries show that 5–10% of patients survive more than three times the average life span of all patients receiving treatment. There is a great need to draw lessons from the experience of these and other exceptional patients.
Intent for target popula-tion	▪ Anthologies of biographies of exceptional patients to see what antecedents, conditions, coping, and complementary treatments they have or have adopted. ▪ Special attention to the huge challenges for the family caregiver, internal familial tensions, impact on family children and adolescents, medical and neuropsychological services, support groups, and implicit suicide chosen by so many ill people who are suffering more financially, or for other reasons, than physically or emotionally. ▪ Wide circulation of these lessons, to enhance medical and neuropsychological attention beyond just following epidemiological tendencies, to help transcend those tendencies; this can take the form of textbooks, manuals for patients and family caregivers, protocols for medical personnel and neuropsychologists, indicators for researchers, and brochures for children and adolescents. ▪ Helping with comparative (and hopefully also international) studies of R&D on these patients, with an emphasis on M&E of the impact of agency programs for patients and family caregivers.
Results	▪ Intensive research is in process in very extraordinary cases to serve as prototypes for more extensive studies. ▪ Preliminary findings have been shared with medical personnel and neuropsychologists, family caregivers of dementia and palliative care patients, and with support groups—e.g., for the 115 Costa Rican cancer patients who suffered overradiation in their cobalt treatments. ▪ Some exceptional university programs and NGOs in public health do show interest in what is being discovered, and in seeing how to adjust their agendas and procedures accordingly, especially with family caregivers. ▪ Key challenges continue to be treated very marginally or exceptionally for the seriously ill, including exceptional patients who are also being evaluated: e.g., children in families with a medical crisis or premature death of a very ill sibling or parent; in general, intrafamily relationships that are so often tense and conflictive, affecting both the patient and the key family caregiver. ▪ Social health agencies and programs interested in impact M&E of their programs are being helped to engage university students undertaking their fieldwork or internships, with the intention that when feasible they can elaborate on and test badly needed protocols.

Lessons learned	▪ The experience of 115 cancer patients who in 1996 suffered overradiation is a prelude to what is evident in Costa Rica and even more elsewhere, such as the United States: that legal, paradigmatic, and ideological currents can greatly impede due attention to factors that can explain the prolonging and quality of life for exceptional patients. ▪ The emphasis has been on indemnification for malpractice palliative care, ideological debates regarding euthanasia, and the usual benchmarks in clinical trials, at the expense of due attention to the possibility of prolonging the length and quality of patients' lives. ▪ A key problem for evaluation can be the chaos or "snarl" in handling patient medical records, including legal handling. ▪ Family caregivers are more receptive to taking lessons from exceptional patients, but they are imbued with very haphazard support for what they have to resolve, and ironically the meager attention they do receive focuses almost exclusively on some of the onerous tasks before them, and not on the personal benefits they often derive from caregiving; this is evident in the most frequently used questionnaires and protocols for caregivers. ▪ The most positive impact has been with patients, but the impact is limited by the enclosed perspectives of the general discourse on this subject, and that of professional and informal caregivers. ▪ There is attention to an incipient movement, especially in Europe, to alter the perspective of medical personnel and neuropsychologists on this matter, e.g., that palliative care can be given early on, and may assure more prolongation as well as quality of life, not just preparation for a more serene death.
6. Knowledge management for older adults	
Background	Costa Rica´s legislation is rather unique in recognizing the importance of two particular interests with regard to older adults: to enhance, adapt, and take advantage of their accumulated empirical intelligence; and to have a direct role in the design and evaluation of programs and projects for their generation. However, we see very little application of these principles.
Intent for target population	▪ Guidance on recapitulation of his or her accumulated intelligence, and how to transmit the lessons and benefits of that knowledge to succeeding generations. ▪ Voice and vote in the design and evaluation of the programs and projects intended to enhance the older adult´s quality of life.
Results	▪ There has been spotty acceptance among some public and private agencies in helping foster and organize this initiative, but overall there have been many obstacles. ▪ There has been much less progress than hoped for with these intentions, and thus with the evaluation of the experience. ▪ In one suburban zone, an older adult association is strongly committed to pursuing these intentions; which offers some hope.

Lessons learned	▪ Older adults, including the more educated, believe that they are to be retired, respected, entertained, and perhaps educated in some subjects, but not with the type of activity intended in this project. ▪ To our surprise, much more than with professionals attending to adolescents and youth, where adult-centrism was expected (in projects 1–4 above), those attending to older adults generally were adamantly against these objectives, possibly fearing a conflict between the need for caregiving and program management. ▪ The older adult is beset with stigmas held by others, and even with self-adoption of such stigma.
7. Effects on communities and users of transportation and communications public service investments in buffer, forest, plantation, and mining zones	
Background	There is a strong sense that these investments enhance transitory exploitation and not sustainable development of these zones, contributing linking to socio-environmental conflicts; great distortions in the added value chain of local products; and higher user costs affecting all sectors. This is an example of a problem or conflict resolution apparently being a predictable precedent of others.
Intent for target population	▪ To know the role of transportation and communications in the evolution of community and family economic sustenance and survival strategies. ▪ To study the options and the predominant criteria of investments in these sectors. ▪ To know the role of the predominant models for these investments and protocols by academics and international cooperation.
Results	▪ Initial studies of Central Appalachia in the United States, and central and southern Costa Rica. ▪ Probing of other case studies and a historical overview of the application of international cooperation models in Latin America. ▪ Preliminary results indicate historical adoption of similar investment models of earlier private investors and later public agencies. ▪ The highly influential evolution (really, several huge evolutions) of the globalization strategy of transnational corporations has not been well addressed in economic development studies of Latin America.
Lessons learned	▪ The significance of this for the community and families who live in it has been largely overlooked by economic historians and anthropologists. ▪ Interviews with older residents of these zones give inconsistent versions of community and family history; there is a need for more creative methods, such as theatrical representations, of that reality. ▪ This appears to vouch for the great need in Central America of R&D-focused university majors in transportation economics.

REFERENCES

ICAP and UNDP (Central American Public Administration Institute and United Nations Development Programme). 2003. "Perfil de Cooperante y del Recipiente Responsables." Vol. 3. Guía de Gerencia de la Cooperación Internacional.

Crowther, Warren. 1999. *Manual de Investigación-Acción para la Evaluación Científica en el Ámbito Administrativo.* San José, Costa Rica. Editorial Universidad Estatal a Distancia.

Rodriguez Bilella, Pablo D., Sergio Martinic Valencia, Soberón Alvarez, Sarah D. Kilier, Ana Luisa Guzmán Hernández, and Esteban Tapella. 2016. *Evaluation Standards for Latin America and the Caribbean.* 1st ed. Buenos Aires: Akian Graphic Publishing.

Stockmann, Reinhard. 2011. *Manual de Evaluación: una guía práctica de procedimientos.* Universidad de Costa Rica, Editorial de la Universidad de Costa Rica, Costa Rica.

UNICEF Evaluation Office. 2002. "Children Participating in Research, Monitoring and Evaluation (M&E) Ethics and Your Responsibilities as a Manager." *Evaluation Technical Notes* No. 2. (Spanish translation by Priscilla Hurtado and María Isabel Araya Tristán.)

UNICEF Costa Rica. 2000. "Código de Ética para las ONGs que participan en proyectos." Internal document.

Chapter 10

The Role of Monitoring and Evaluation in the MENA Region, with a Focus on the Arab Uprising Countries

Sonia Ben Jaafar and Awny Amer

Abstract. *The demands of monitoring and evaluation (M&E) for the Sustainable Development Goals in the Middle East and North Africa region are particularly challenging. The diversity of the 22 countries; a large and youthful population; unprecedented political transitions; and a variety of conflicts and humanitarian crises have created a cluster of complex development needs. Development aid activity in the region is now operating in a landscape of high accountability demands and traditionally poor governance practices. This chapter explores the way these regional transitions are cultivating a new paradigm that promotes national capacities and country M&E systems in which local M&E professionals are valued and supported in the development of their own countries and region.*

Sonia Ben Jaafar, EduEval Consultancy, s.benjaafar@edueval.com; Awny Amer, Independent Consultant, awnyamer22@gmail.com.

The Arab region is comprised of 22 countries in Northern Africa, the Levant, and the Arabian Peninsula. This region, which is known as the Middle East and North Africa (MENA) region, has a population of approximately 350 million, the majority of which is under the age of 25.[1] The current challenges affecting this region are partially tied to an unprecedented wave of political transition that has swept the region since 2010, coupled with an increasing number of conflicts. According to the United Nations Development Programme (UNDP):[2]

> Weak social, political and administrative accountability mechanisms and politically oriented socioeconomic planning models have resulted in the neglect of large parts of the population. These nations face the challenge of forming new, accountable governments that reflect popular aspirations.

THE ARAB UPRISING

The Arab Uprising refers to a series of antigovernment protests, uprisings, and armed rebellions across the Middle East that surfaced in 2010. By 2012, the rulers in Tunisia, Egypt, Libya, and Yemen had been pushed out of power; civil uprisings had erupted in Bahrain and Syria; there were major protests in Algeria, Iraq, Jordan, Kuwait, Morocco, and Sudan; and minor protests had occurred in Djibouti, Mauritania, Oman, Palestine, Saudi Arabia, and Western Sahara (Smith 2016). This revolutionary atmosphere in the Arab region was tempered in the Gulf states, where a more reformist approach took root (Abdalla 2012). The Gulf monarchies demonstrated the ability to adapt to regional shifts and to address internal issues with policy measures, by using their "oil wealth, historical legitimacy, Bedouin culture, demographic scarcity, extensive security services, patriarchal regimes, and the absence of an opposition and political parties" (Abdalla 2012, 30).

Moreover, according to UNDP's 2015 *Human Development Report*, Bahrain, Kuwait, Qatar, Saudi Arabia, and the United Arab Emirates are ranked among the top countries in the Arab region in terms of average income per capita (UNDP 2015) and the economic competitiveness index,[3] which creates a certain level of comfort for their people. For example, the United Arab Emirates leads the Arab states, and is considered globally as one of the happiest countries: in the 2015 *World Happiness Report*, it is given special mention as an example of a country in which well-being has been made a central tenet of the design and delivery of the national agenda (Heliwell, Layard, and Sacha 2015).

[1] Source: United Nations Development Programme, http://www.arabstates.undp.org/content/rbas/en/home/regioninfo.html#Introduction.

[2] Ibid.

[3] World Economic Forum, "Competitiveness Rankings," http://reports.weforum.org/global-competitiveness-report-2015-2016/competitiveness-rankings/.

However, for the purposes of this chapter, we will discuss the challenges with an emphasis on the majority of the region, where there has been revolutionary upheaval and a large impact on development. We focus on the impact of more than 11 million people forced from their homes in Syria;[4] an estimated 7 million internally displaced people within Syria; and more than 4 million who have fled Syria as refugees, a large portion of them landing in Jordan, Lebanon, and Turkey. We focus on the impact of an initially peaceful uprising in Libya that quickly became an armed conflict with Western military intervention. According to Amnesty International,[5] once Libya had become deeply divided, the internal conflict has caused civilians to live in a constant state of threat, with nearly 2.5 million people in need of humanitarian assistance, clean water, sanitation, and food since 2014. In this chapter, we focus on the areas in crisis, which has drawn international attention and aid, because the potential to meet the Sustainable Development Goal (SDG) targets in the region is increasingly bleak.

DEMANDING ACCOUNTABILITY

The key driver in this unprecedented regional call for change is the call for greater accountability from the government to the people. This root demand has thrust the role of monitoring and evaluation (M&E) to a new level of significance, giving it greater value for all stakeholders. As M&E and accountability have been becoming more important to the local populations, donors are more interested in the *effectiveness* of aid in the region. The SDGs that are key priorities in MENA countries focus on ongoing and intersecting issues of poverty, hunger, health, education, energy, equity, economic downturns, climate, peace, and stability in the region.

As aid funding increased in a progressively complex and challenging region during the aftermath of the Iraq War, so did attention to developing specific mechanisms to ensure aid effectiveness. This promoted solutions like the application of results-based management to program, thematic, and sectoral evaluations, rather than simple project evaluations (UNDG 2011). Solutions such as country-level evaluations that consider coordinating the efforts of multiple donors with joint criteria will contribute to ensuring the effectiveness of the evaluation processes (Baradei, Abdelhamid, and Wally 2014). Although this shift is not restricted to the Middle East and North Africa, the region has grave challenges in adopting these solutions because of traditional social, political, and economic practices that are unaccustomed to the three key pillars of results-based management: accountability, national ownership, and inclusiveness (UNEG 2011), pillars that happen to align with the reverberations of the Arab Uprising.

[4] Source: Syrian Refugees website, "A Snapshot of the Crisis in the Middle East and Europe," http://syrianrefugees.eu/.

[5] Source: Amnesty International website, "'The Arab Spring' Five Years On," https://www.amnesty.org/en/latest/campaigns/2016/01/arab-spring-five-years-on/.

Given this reality, the M&E community in the region has been increasingly focused on accountability, national ownership, and inclusiveness. Some of the issues of core importance to the M&E community follow:

- **Accountability** as it pertains to the extent of the relationship between the impact of aid effectiveness and the level of effort from implementing and donor organizations; and the value added of M&E processes in terms of corrective program actions and policy changes
- **Inclusiveness** as it pertains to the role of the beneficiaries of aid efforts in relation to implementing partners; the traditional North-South paradigm, in which development interventions and M&E are conducted primarily for donors; and the lack of shared accrued knowledge derived from M&E that adds to the future independence of local actors and organizations
- **National ownership** as it pertains to the extent of local M&E capacity and professionalism; and the extent of involvement of local leadership in assessing the effectiveness of the aid received

These areas of focus align tightly with evaluation of the SDG plans and programs that necessitate the involvement of national efforts to cultivate appropriate evaluation capabilities at all levels and across all stakeholders (El-Saddik et al. 2016). It is noteworthy that although accountability is a shared value across the globe, with good governance as a core element, the Arab region lagged behind other regions in most governance indicators in 2015 (UN 2013). Given that good governance is not only essential for accountability, but is also a gateway to inclusion and national ownership, M&E efforts in the region are arduous and highly political. Local M&E professionals work within very challenging parameters in their attempts to change the landscape on the path to meeting the SDG targets.

An illustrative example of the limited role of M&E in the region is the need for the Taqeem Initiative, which was established in 2009 by a partnership between the International Labour Organization, Silatech, the World Bank, the Jacobs Foundation, and the Swedish International Development Foundation. Taqeem was specifically created to support youth employment policy makers and practitioners in enhancing the M&E of their programs (ILO 2009).

The lack of evaluation information on youth employment programs in the region is telling, given that it is mostly populated by youth,[6] and that the youth employment challenge is a critical issue that threatens the already fragile economic and political state of many countries in the region. Creating opportunities for Arab youth has long been a leading policy priority for international organizations, governments, nongovernmental organizations, and social entrepreneurs. Traditional and nontraditional programs, projects,

[6] Source: UNDP, http://www.arabstates.undp.org/content/rbas/en/home/regioninfo.html#Introduction.

initiatives, and partnerships have been thrust upon the region by both local and foreign organizations. The Taqeem Initiative targets these particular solutions and helps local organizations drive quality M&E while creating a regional database of evaluation information that will help policy decisions concerned with solving the youth employment challenge.

Although institutionalization of the Taqeem Initiative is laudable, it is regrettable that the culture of the region did not already have the demand and supply of M&E for these solutions to meet a vital regional challenge. The reality of the situation is well summarized in a recent study of the state of M&E in post-revolutionary Egypt, in which the authors highlight the fact that the government urgently needs effective support "to demonstrate and measure the results of each policy" (Baradei, Abdelhamid, and Wally 2014).

THE CHALLENGES OF EVALUATING RESULTS FOR DEVELOPMENT

The regional M&E community has been sharing knowledge to identify key issues that contribute to promoting the demand for the evaluation of development results in response to a rising public demand. A series of strategic discussions and debates across MENA countries have taken place to better appreciate the current situation of M&E processes and practices in the region (El Kabbag 2011). The following key challenges have been identified as barriers to normalizing supportive M&E in the region.

A deterring culture of evaluation (El Kabbag 2011). In most MENA countries evaluation is tightly coupled with undesirable beliefs about the impact of the results. In this context, evaluation has either no influence at all, or grave consequences for local stakeholders and projects. Evaluation is believed to be restricted to audits and financial reviews that demonstrate inefficiencies or corruption, or public reviews that expose fraud. It is also believed that it is only relevant to donors for the purpose of satisfying bureaucratic requirements; that it is conducted apart from the beneficiary communities; and that it is tied solely to outputs, with no regard for outcomes and impact, and all of the other traditional high-stakes and potentially harmful ways evaluation has been used in the past.

The status quo is the norm. Most of the organizations in MENA have a culture of following evaluation processes that are stagnant within the organization. This static reality is in contrast to an evaluation approach that focuses on a dynamic and holistic perspective designed to transform the organization for the better.

National capacities are not ready for change. Given that the MENA region has not had a strong historical integration of M&E into their public governance, it is not surprising that there is a less than adequate national capacity in many of the countries, at both the governmental and civil society levels. In particular, evaluators in the region generally do not have ready access to accredited M&E training, and have gaps in some of their knowledge and

skills. For example, there are gaps in adopting minimum standards of ethics during the evaluation process. There is also a gap in knowledge on how to use system-level data, because many governments have incomplete, inaccurate, insufficient or dated data.

Lack of good governance affecting good evaluation. As previously stated, the Arab region has one of the poorest governance records globally. When we consider the indicators of good governance, such as voice, accountability, and governmental effectiveness, it is evident that poor governance will have a significant influence on evaluation practices (Kaufmann, Kraay, and Mastruzzi 2010). This situation is part of the overall socio-political context that is being called into question with the Arab Uprising.

A non-enabling environment for developing evidence-based policy. Evidence-based policy is rooted in responsible government, which has become a mantra in the post-Arab Uprising MENA region. Evidence-based policy is expected to reduce wasteful spending, expand innovative programs, and strengthen accountability (Kaufmann, Kraay, and Mastruzzi 2010). Part of the framework for this approach, which enables governments to make better choices, is program assessment, outcome monitoring, and targeted evaluation. Regrettably, a lack of good governance is often tied to a lack of evidence-based policy. The situation in the Arab region is such that the tracking of progress at the policy level for SDGs and Millennium Development Goals has been, and continues to be, a key challenge. This means that in addition to a low supply, there is a low demand for M&E.

PROFESSIONALIZATION OF M&E IN THE MENA REGION

The challenges presented in the previous section are directly tied to the systematically poor governance that created the conditions for the Arab Uprising. M&E supply and demand have been less significant in the past, but the uprising, and the subsequent attention to SDGs have exposed an urgent need for M&E. Local governments and the international community are now promoting accountability and unveiling a substantive demand. At the same time, the M&E communities of professionals have been working to create an enabling environment for M&E. The regional professionalization of M&E specifically promotes national ownership and inclusion, and activates an evolved accountability that goes beyond the conducting of external audits to appease donor distrust (Segone 2009).

The increasingly high demand for M&E professionals in the MENA region is an issue, especially since quality M&E requires an investment of time and resources in order to plan, collect data, and report appropriately. Increasing efforts to professionalize the field are leading to improved evaluations and better evidence. This increases the demand for more and better evidence, which in turn stimulates further improvements in professionalism. However, program and operational funds in organizations in this region often neglect to budget for appropriate M&E, for various reasons. The key issue, a lack of data culture, which makes for a situation in which it is difficult to

cultivate M&E professionals, has been addressed with multiple efforts by international and local organizations.

In particular, the launch of the Middle East and North Africa Evaluators Network (EvalMENA) in 2012 galvanized M&E professionals at the regional level. EvalMENA is an informal network of stakeholders who are dedicated to professionalizing M&E in the region. It aims to promote and strengthen M&E culture and practices in all the MENA countries by encouraging country-level professional M&E associations, and by offering a networking platform for thoughtful debate about local issues and solutions.

The work of EvalMENA is practical and addresses local issues. The success of the organization is largely due to the fact that inclusiveness was—and continues to be—a fundamental building block of the organization and all the practices of EvalMENA. This success includes lobbying for M&E, networking events, capacity building, creating new evaluation associations, and professionalization efforts. The inception of the organization was a research and development project championed by the Environment and Sustainable Development unit of the American University of Beirut, with technical and financial support provided by the International Development Research Centre in Canada (IDRC). However, even with donor funding and appropriate supportive guidance from IDRC, the leadership and working teams have always been local Arab professionals. Since its inception in 2008, EvalMENA has managed to achieve the following (Moussa 2015).

Bringing MENA evaluators together, in the region and around the world. In 2013, the membership was made up of 146 evaluators from 23 countries. By 2015, the membership had grown to 360 members from 38 countries. Most of the members were from Egypt, Jordan, Lebanon, Morocco, Palestine, and Tunisia, with a small minority from North America. Currently, the membership is over 500 members, indicating a growing interest and support for the vision of EvalMENA.

Promoting and supporting seven new national evaluation associations. National ownership and local inclusion has been a top priority of EvalMENA since the inception of the network. The formalization and maturation of the Moroccan Evaluation Association (MEA) has been an important catalyst in the region. MEA has now matured to the point where its work with policy makers in Morocco is allowing them to take an important seat at the same table with the highest level of government. The success of MEA has proved to be a catalyst in the formalization of six more new networks: the Egyptian Research and Education Network (EREN) in 2012; the Palestinian Evaluation Association (PEA) in 2013; the Jordan Evaluation Association (EvalJordan) in 2014; the Tunisian Evaluation Association (RTE) in 2014; the Lebanese Evaluation Association (LebEval) in 2014; and the Egyptian Development Evaluation Network (EgDEval) in 2015.

All of these new associations are now actively contributing to the leadership and management of EvalMENA, and are promoting its vision at the national level.

Organizing five annual regional conferences between 2012-16. The Arab culture places a high value on face-to-face networking. Given the critical importance of bringing M&E professionals together to share and create knowledge that is localized for the region, it was—and is—important to hold local conferences. As the membership and national ownership has grown, EvalMENA has shared and distributed leadership and responsibilities among the participating countries, while offering them ongoing support. As table 10.1 indicates, the national organizations have taken ownership of the regional vision. This is especially impressive after the initial IDRC funding ended, as the national and regional organizations collaborated to find other funding sources to continue the work.

Promoting MENA to an international audience through active leadership in the global evaluation movement. The context within the MENA region is becoming increasingly challenging, with humanitarian crises becoming the norm. The stunted progress in many Arab countries is cultivating high-risk zones. This creates a two-pronged problem with respect to local capacities. First, local M&E professionals are not plentiful in the region, and donors are more comfortable with M&E professionals who have more experience and formal training. Second, local M&E professionals are in greater demand, because foreign M&E officers are less and less able to access the zones where information is most needed.

EvalMENA has been working strategically to achieve recognition for the region's M&E professionals from the international community. Given that most donors are part of the international community, this recognition is important for donor trust of local capacity.

As regional annual conferences have gained traction in the region, M&E professionals have gained increasing access, awareness, and encouragement

TABLE 10.1 **EvalMENA conferences**

Year	Location	Key donors
2012	Lebanon	International Development Research Center (IDRC)
2013	Lebanon	International Development Research Center (IDRC)
2014	Jordan	International Development Research Center (IDRC)
2015	Egypt	UNICEF Regional Office for MENA (MENARO) IDRC IOCE/EvalPartners International Initiative for Impact Evaluation (3ie) Islamic Corporation for the Development of the Private Sector (ICD) Environment and Sustainable Development Unit of the American University of Beirut (AUB-ESDU)
2016	Tunis	Tunisian Evaluation Network (TEN) International Initiative for Impact Evaluation (3ie) UN Women-Regional office IOCE/EvalPartners

to be actively involved in the global evaluation movement. For example, active EvalMENA members have become executive board members of the International Development Evaluation Association (IDEAS), the International Organization for Cooperation in Evaluation (IOCE), the Africa Gender and Development Evaluators Network (AGDEN), and the African Evaluation Association (AfrEA), including the AfrEA presidency in 2012 and the IOCE presidency and EvalPartners cochair for 2015–16. These leadership roles have demonstrated to many members of the international community that high-level M&E professionals are present, and ready to contribute to the work in the MENA region. It has also shown Arab M&E officers that there are opportunities to grow professionally, and to engage with global networks.

Launching of the first online training course on development evaluation in Arabic. The first online training course on development evaluation in Arabic was launched in 2014 on the "My M&E" portal. This is a significant accomplishment because it is an Arabic course that is listed on a high-profile site along with courses from UNICEF and UN Women. This course boasts a credibility that is valued by donors and international agencies. Since it was launched, 276 Arab-speaking individuals with an interest in learning about M&E have successfully completed the course. It is noteworthy that many of the participants are from Iraq, Sudan, Syria, and Yemen, countries where there is no national network for evaluators. This online course helps Arabs who have limited local access to opportunities learn and share knowledge about development evaluation overcome an important barrier to joining the M&E community. This is particularly important as online learning, as an information and communications technology (ICT)–based solution, becomes increasingly vital for promoting access to learning in conflict and poverty-affected areas. The course continues to be accessed by Arabs who are hopeful about further developing their M&E knowledge, skills, and professional network.[7]

Facilitating South-South collaboration on evaluation within and beyond the MENA region. As we move into a new paradigm that values and promotes inclusion and national ownership in M&E, it is essential that there is a shift from the traditional North-South donor-recipient mentality to a shared values and joint work mentality. South-South collaborations promote the notion that all parties to the partnership are equally valuable, and that they aim to achieve a shared goal. In strategically bringing together and supporting so many actors in the region, EvalMENA has cultivated a culture for knowledge sharing and creation. In addition to the EvalMENA-sponsored national workshops and events organized in Egypt, Jordan, Lebanon, and Morocco, three projects won the 2014 Peer-to-Peer small grants from IOCE:

[7] The course is available at http://www.mymande.org/elearning/course-details/6.

- **Media Promoting Evaluation Culture in MENA.** A joint partnership between EREN, the Faculty of Communication at Cairo University, and Environment and Sustainable Development Unit /EvalMENA;
- **Integrating Evaluation in Legislative Bodies** A joint partnership between EREN, the Secretariat of the Egyptian Parliament, and the Parliamentarian Forum for Development Evaluation in South Asia; and
- **The Evaluation Database Enhancement Project.** A joint partnership between the Palestinian Evaluation Association and the Jordan Development Evaluation Association.

All of these accomplishments in professionalizing M&E in the region would be notable at any time. But it is particularly impressive and relevant given that all of these activities took place in the aftermath of the Arab Uprising, one of the most turbulent periods in the recent history of the region. The instability of the region necessitates rethinking M&E approaches in general, but especially in areas where there are humanitarian emergencies. This is particularly true where there is an unprecedented humanitarian crisis with predictions for worse to come, as is the case in this region. According to the United Nations High Commission for Refugees, the conflict and violence in Iraq and Yemen have displaced 4.5 and 2.18 million people, respectively; and the conflict in Syria has displaced 4.8 million people who are seeking safety in Jordan, Lebanon, Iraq, and beyond.[8]

The commitment of the national M&E communities to develop national evaluation capacities for the SDGs was exemplified by Egypt's willingness to conduct a voluntary review of the SDGs at the United Nations High Level Political Forum on Sustainable Development in 2016. More than 60 representatives from the private sector, civil society organizations, foundations, academia, youth, special interest groups (women, environment, startups, etc.), and development partners were invited by the government of Egypt to discuss the road map for implementing and monitoring the SDGs. The meeting was hosted by the Ministry of International Cooperation, and co-organized by UNDP and the World Bank, as part of their joint effort to raise awareness about the SDGs in Egypt. This group focused on Egypt Vision 2030, the national sustainable development strategy.

In a spirit of accountability, Egypt volunteered to report on their efforts to achieve the agenda: they reported that engaging with local interest groups is pivotal to building ownership of the SDGs; capitalizing on local efforts for knowledge and resources; and increasing mutual accountability of national development results.[9] The practical outcome of the report is a series of elements that need to be addressed through collective action. These are as follows:

[8] Source: United Nations High Commission for Refugees website, http://www.unhcr.org/en-us/syria-emergency.html.

[9] Source: UNDP website, "Building Ownership of the Sustainable Development Goals in Egypt," http://www.eg.undp.org/content/egypt/en/home/presscenter/articles/2016/june/building-ownership-of-the-sustainable-development-goals-in-egypt.html.

- **Tackle the data gap, and map national status.** The local groups expressed their readiness to support the monitoring and follow-up of the SDGs. They identified the need to set baselines for all SDG goals and targets.
- **Integrate the implementation efforts of local groups led by national authorities.** A big challenge is to integrate the efforts of government bodies and ministries for the implementation of the SDGs with other nonstate actors. Terms of reference will be developed for different goals where a multidisciplinary working group composed of government and key group representatives will support the implementation of the SDGs.
- **Build partnerships with nonstate actors, especially the civil society sector.** In parallel with the review work, Egypt Vision 2030 was formulated through an extensive nationwide consultation process. The 2030 agenda provides a platform for government and nonstate actors to initiate discussions around common areas of work, and to complement their efforts to achieve national priorities.

As in Egypt, all of the countries in the region are promoting greater local capacity and ownership of M&E as a key element in strengthening M&E practices in connection with the SDGs. The notion that local M&E talent should be used is obvious within a modern paradigm where inclusion and national ownership are core values. For those organizations that have yet to shift into this approach, there is a more practical reason to leverage local M&E capacities. According to a study released in 2015 on M&E practices during humanitarian emergencies (Jansury et al. 2015), M&E is a means for international organizations to use in addressing issues of access and security in complex emergencies. Organizations can choose to either work through local partners (nongovernmental organizations), or to employ local staff.

Although there are challenges to including local M&E professionals, the benefits have been shown to be important. For example, in 2010 the United States Agency for International Development (USAID) mission implemented the Yemen Monitoring and Evaluation Project. This project used third-party local partners to provide on-the-ground performance monitoring, verification, and evaluation of USAID activities. They were able to successfully identify problems in the quality of some of the Community Livelihood Project's rehabilitation activities, and of goods delivered. This was especially important as security deteriorated after the Arab Uprising, and U.S. employees could no longer access some regions of the country to monitor and identify problems with project activities (Office of Inspector General 2015).

These are the kinds of success stories that demonstrate that local M&E professionals can, and should, be involved in the work of foreign aid projects. It is unfortunate that this is not a normative practice despite the discourse concerning local engagement among many donors and foreign aid organizations. For example, the USAID Office of Inspector General conducted a survey to identify the challenges USAID faced during the early transition period (Q4 2010–Q3 2014) in Egypt, Libya, Tunisia, and Yemen. One of their conclusions

was a recommendation to employ third-party monitors in transitional situations because they "can help an office gain access to regions of a country that become inaccessible to U.S. direct hires or when U.S. personnel are ordered to evacuate" (Office of Inspector General 2015).

Moving beyond the "absolute necessity" argument of international organizations delivering foreign aid, local M&E professional communities are shifting the role of M&E within their local governance systems. For example, PEA has offered training on the evaluation of humanitarian programs that is focused on evaluation design and methodology; facilitated national roundtable discussions on the evaluation of SDGs; and collaborated with members of the Palestinian Legislative Council to systematize M&E activities within government initiatives. Another example is EREN, which has co-offered seminars on evaluation for improving governance practices with Plan International, and workshops with government officials on results-based management with UNICEF. They have also been working with local governments' policy briefs, and through a national conference on country-led M&E have focused on promoting transparency and efficiency with policy makers.

These efforts have been matched with those of other local organizations committed to including M&E professionals on staff despite the supply challenge. For example, a recent study in Egypt found that most organizations involved with development M&E provide training internally, in which "much of the training conducted is done through on-the-job learning" (Baradei, Abdelhamid, and Wally 2014). In addition, local organizations face a high turnover in the M&E role, in part because qualified M&E officers are in short supply and thus often highly sought after by other organizations (ILO 2009). This often means that when local organizations do dedicate the time and budget to training program officers to become M&E officers, other organizations—often international ones with greater resources than local ones—poach talented and trained program officers once they have the experience and background to conduct M&E work in the region (Boitnott 2015).

CONCLUSION

Well-intentioned external parties often treat the MENA region as a monolithic entity. But the countries within the MENA region are very diverse. UNDP identifies four distinct groups of countries within MENA: the Mashreq and Maghreb countries, the countries of the Gulf Cooperation Council, and the least developed countries (UN 2013). International organizations that do not differentiate the needs within the region learn their lesson after implementation issues have been identified. For example, in the evaluation of the Media Cooperation under the Danish Arab Partnership Programme (2005–12), which was commissioned by the Danish Ministry of Foreign Affairs, it was found that the cultural context differs in Jordan, Lebanon, Morocco, Syria, and Tunisia, creating conditions in which a uniform regional approach across all Media Cooperation Programme themes is not conducive to achieving the best results (Ministry of Foreign Affairs of Denmark 2013). In this regard, it is worth saying that commissioners of evaluations tend to lay the responsibility for the findings with the evaluators, not with themselves.

The inclusion and national ownership movement is essential for finding sustainable solutions for one of the most volatile youthful parts of the world. The paradigm shift in terms of development in the MENA region is tied to the current state of development in general. Cultivating accountability, inclusiveness, and national ownership is a challenge in a region where there is a continuing war, and a growing humanitarian crisis. (More than 5 million people have fled Syria since 2011, seeking safety in neighboring countries, and millions more are displaced inside Syria.[10]) The unprecedented challenges in the region are drawing international attention and funding, with various solutions to help minimize the loss of life, hope, and health, and to reduce the conditions of indignity for many people. These solutions are being constructed as aid provided with a traditional approach, which has yet to move beyond external accountability, and to value inclusiveness and national ownership. At the same time, measuring aid effectiveness in hazardous areas is opening new opportunities for local M&E professionals to gain experience, trust, and training from international agencies simply because these are the M&E people who are on the ground with access to the conflict areas (Jansury et al. 2015).

The evaluation efforts tied to the SDGs in MENA are increasingly aligned with the key principles of accountability, inclusiveness, and national ownership. This triad cultivates accountability for sustainable development when public policies are subject to local evaluations whose purpose is to ensure the best solutions for local issues. This approach means explicitly addressing the data gap and mapping national status along the SDG indicators; integrating the efforts of major groups involved in implementation, led by national authorities; and building partnerships with nonstate actors, especially with civil society. This can only be achieved when local M&E systems and professionals are involved in the process.

The country and regional efforts that have been contributed by local M&E professionals have been impressive. These initiatives are directly addressing the challenge of a low enabling environment coupled with restricted M&E skills and trustworthiness, which places Arab countries in an opinion-based as opposed to an evidence-based system (Segone 2009). The region's M&E systems, capacities, demands, and professionalism are maturing. Dedicated local actors, such as the advocates and participants of EvalMENA, EREN, PEA, EvalJordan, RTE, LebEval, and EgDEval, are cultivating the culture for national evaluation processes to be aligned with other planning, budgeting, and statistics processes to drive the 2030 Agenda (El-Saddik et al. 2016). The demands of this era call for more international organizations and foreign donors to adopt the perspective of the IDRC and the Taqeem sponsors, in which local actors are regarded as partners with the capacity to support M&E and learning, and are appreciated as valuable assets to the development of their own countries and region.

[10] Source: United Nations High Commission for Refugees website, http://www.unhcr.org/en-us/syria-emergency.html.

REFERENCES

Abdallah, Abdulkhaleg. 2012. "Repercussions of the Arab Spring on GCC States." Research Paper, Arab Center for Research and Policy Studies, Doha, Qatar.

Baradei, L., D. Abdelhamid, and N. Wally. 2014. "Institutionalising and Streamlining Development Monitoring and Evaluation in Post-Revolutionary Egypt: A Readiness Primer." *African Evaluation Journal* 2 (1).

Boitnott, John. 2015. "How to Poach Talent Politely." *Entrepreneur*.

Heliwell, John, Richard Layard, and Joffrey Sacha. 2015. *World Happiness Report 2015*. New York: The Earth Institute, Columbia University.

El Kabbag, Nivine. 2011. "MENA Evaluation Experiences and Networks." Paper presented at IDEAS Global Assembly, April 2011.

El-Saddik, Kassem, Dorothy Lucks, Stefano D'Errico, Thomas Schwandt, and Zenda Ofir. 2016. "Developing National Evaluation Capacities in the Sustainable Development Era: Four Key Challenges." IIED Briefing, International Institute for Environment and Development, London.

ILO (International Labour Organization). 2009. "Making Learning Count: Effective Monitoring and Evaluation of Youth Employment Programs in the Arab World." Taqeem Initiative. Geneva: ILO.

Jansury, Laurel, Jennifer Moore, Juliana Peña, and Abigail Price. 2015. "Findings in Monitoring and Evaluations Practices During Humanitarian Emergencies." Paper prepared for International Business & Technical Consultants, Inc., Washington, DC.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010. "The Worldwide Governance Indicators: Methodology and Analytical Issues." Draft Policy Research Working Paper. World Bank, Washington, DC.

Ministry of Foreign Affairs of Denmark. 2013. *Evaluation of Media Cooperation under the Danish Partnership Programme (2005–2012)*. Copenhagen.

Moussa, Ziad. 2015. "Mainstreaming Evaluation Theory and Practice in the MENA Region." Final Report. International Development Research Centre.

Office of Inspector General. 2015. "Survey of USAID's Arab Spring Challenges in Eypt, Tunisia, Libya, and Yemen." Survey Report No. 8-000-15-001-S. United States Agency for International Development (USAID), Frankfurt.

Segone, Marco. 2009. "Enhancing Evidence-based Policy Making through Country-led M&E Systems. In *Country-led Monitoring and Evaluation Systems*, Marco Segone, ed. New York: UNICEF.

Smith, Lydia. 2016. "Arab Spring 5 Years On: Timeline of the Major Events and Uprisings in the Middle East." *International Business Times* January 23.

UN (United Nations). 2013. *The Arab Millennium Development Goals Report*. Beirut: United Nations Economic and Social Commission for West Asia (ESCWA).

UNDP (United Nations Development Programme). 2015. *Human Development Report 2015: Work for Human Development*. New York: UNDP.

UNDG (United Nations Development Group). 2011. *Results-Based Management Handbook*. UNDG.

Chapter 11

Leveraging Monitoring and Evaluation Systems for Good Governance

Rashmi Agrawal, Asela Kalugampitiya, Jigmi Rinzin, and Kabir Hashim

Abstract. *Successful pursuit of the recently adopted Sustainable Development Goals by the world community depends on appropriate national development policies and actions. These should be guided by considerations of social equity, gender equality, and respect for environmental stability, and must be supported by good governance in order to contain corruption. This chapter discusses the concept of good governance, with an emphasis on the mutually dependent and mutually reinforcing relationship between good governance and sound monitoring and evaluation. Recent efforts and initiatives to bring about improved governance and monitoring and evaluation (M&E) effectiveness in three South Asian countries—Bhutan, India, and Sri Lanka—are discussed. Initiatives in India and Sri Lanka designed to contain corruption in public service delivery and make it more transparent, and the significance of M&E in promoting good*

Rashmi Agrawal, International Development Evaluation Association, rashmi_agrawal56@rediffmail.com; Asela Kalugampitiya, EvalPartners, asela.kalugampitiya@gmail.com; Jigmi Rinzin, National Council of Bhutan, jigmi@hotmail.com; Kabir Hashim, Sri Lankan Ministry of Public Enterprise Development.

governance and enhancing Gross National Happiness in Bhutan are described. These initiatives suggest some of the options available for leveraging M&E systems for good governance.

The adoption of the 2030 Agenda for Sustainable Development on September 25, 2015, at the United Nations Summit, which formalized the Sustainable Development Goals (SDGs) to be achieved by the world community of nations by 2030, has placed the SDGs center stage in the global development process. Attaining the 17 complex, interrelated SDGs in various socioeconomic sectors, and achieving the multitude of targets corresponding to each goal within the set time frame, has spurred action by the international community as well as by national governments to deliberate on what steps should be taken in order to make, measure, and assess progress toward these goals. While international action concerns sharpening of the goals and targets, and strengthening national capacities in measuring progress toward these goals, national governments are reviewing the capacities of their statistical and nonstatistical systems to respond to the requirements of such measurement. Attaining these goals calls for a systems approach, because the individual goals are not "in silos": instead, they are mutually impacting, at times reinforcing, and on occasion, retarding other related goals. Multiple stakeholders—states, industries, civil society organizations (CSOs), and the community at large—must join hands and work together for this cause. While well-conceived policies and programs for equitable social and economic development through good governance can lead to sustainable development, sound monitoring and evaluation (M&E) systems would accelerate that progress. Policies and programs can lead to far better results under a scenario of good governance than when governance is poor. Good governance and effective M&E systems, which are mutually dependent and mutually strengthening, can contribute to optimizing returns on investment. While good governance assumes willingness on the part of policy makers and program planners and implementers to be subjected to a critical review of their actions, and to learn from such assessment, a robust M&E system can operate with a good measure of success only within an enabling environment. The imperative need for these two elements to work together has never been as obvious as it will be in the pursuit of SDGs by nations over the next decade and half. A robust M&E system that integrates within itself the requirements of sustainable development, mainstreamed within the national development agenda, is essential.

This chapter focuses on the concept of good governance and the interrelationships between good governance and sound M&E systems for sustainable development. It presents recent experiments and experiences in good governance and M&E in three South Asian countries—Bhutan, India, and Sri Lanka—and argues that there is more than one viable option available. The approaches followed in various countries can provide good learning that can be replicated in other places.

THE CONCEPT OF GOOD GOVERNANCE

What is "governance," and what is implied by "good governance"? In Kautilya's *Arthasashtra,* good governance by a ruling king is described in these terms: "… in the happiness of his subjects lies his happiness, in their welfare his welfare."[1]

The United Nations Economic and Social Commission for Asia and the Pacific (ESCAP) defines governance as "the process of decision-making and the process by which decisions are implemented (or not implemented)" in corporate, local, national or international contexts (ESCAP n.d.). Governance, therefore, is the result of the collective interaction of the decision making and implementing actions of the various actors and institutions concerned including those in both the public and private spheres of action: government at the national and local levels, industry, trade unions, CSOs, influential individual players, and even various organs of the media.

The World Bank, in its studies in more than 200 countries, has developed several indicators to measure the quality of governance using six dimensions: voice and accountability; political stability and absence of violence; government effectiveness; regulatory quality; rule of law; and control of corruption.[2] The World Bank views good governance as a necessary precondition for development, and the *Human Development Report* has defined good governance "as a democratic exigency [that], in order to [rid] societies of corruption, [gives] people the rights, the means, and the capacity to participate in the decisions that affect their lives and to hold their governments accountable for what they do" (Nzongola-Ntalaja, as quoted in UN DESA 2007, 4).

The United Nations has identified transparency, accountability, responsibility, participation, and responsiveness as the core attributes of good governance. Good governance is increasingly viewed as an essential element of any well-functioning society: when resources are allocated and used effectively, and the delivery of services to citizens in an equitable manner is ensured, the government gains a good measure of social legitimacy.[3] The essential characteristics of good governance can be summarized as shown in figure 11.1.

Whatever the definition, good governance refers to a situation in which a set of institutions and actors combine to lead to sound processes of decision making, and the implementation of policies, programs, and projects that contribute to economic and social development, with "no one left behind." "No one left behind" is the theme of the SDG agenda, and it is directly linked

[1] *Arthasashtra*, a treatise in Sanskrit on governance and statecraft, is traditionally believed to have been authored by Kautilya (also known as Chanakya and Vishnugupta) c. 350–283 BCE.

[2] The Worldwide Governance Indicators project (http://info.worldbank.org/governance/wgi/#home) reports aggregate and individual governance indicators on these six dimensions of governance for more than 200 countries and territories over the period 1996–2016.

[3] See United Nations Human Rights, http://www.ohchr.org/EN/Issues/Development/GoodGovernance/Pages/GoodGovernanceIndex.aspx.

FIGURE 11.1 **Components of good governance: a summary**

to good governance. The concept includes an element of flexibility in relation to its application in different country contexts: it may mean different things in different countries, depending on the cultures, traditions, political structures, economies, and levels of development. In a broad sense, good governance is an umbrella concept that covers respect for human rights, rule of law, an efficient and effective public sector, and processes of accountability and transparency of actions in public sphere.

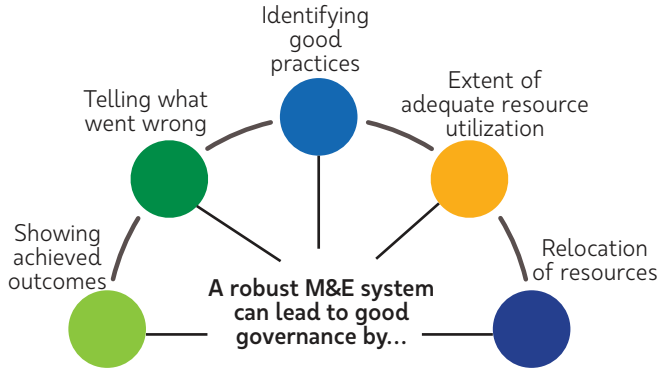
THE RELATIONSHIP BETWEEN GOOD GOVERNANCE AND M&E

There is a symbiotic relationship between good governance and M&E. Good governance creates an enabling environment for M&E, and M&E contributes to good governance. The former includes a keen desire on the part of the system to assess its own performance from time to time, to learn from experience, and to improve the outputs, outcomes, and impacts of the policies and programs it is pursuing. This is achieved through better planning, management, and implementation (PM&E): and this in turn creates a demand and an enabling environment for M&E.

Participation and responsiveness are two of the core attributes of good governance that also tend to increase the demand for M&E from civil society and other stakeholders. An enabling environment for M&E would also include the adoption of a clear national policy that promotes M&E in all its aspects, including the development of evaluation capacity, socially equitable

and gender-responsive elements, and the sustainability of evaluation. In turn, M&E works as a "reality check" tool for what authorities are saying about development, and what is really happening on the ground, and generates evidence-based lessons for the future, thereby contributing to knowledge, and suggesting policy and program modifications for enhanced future outcomes. Good governance provides a way for good M&E systems to exist and evolve, while M&E systems provide governments with evidence and learning that helps in need-based policy planning, and the improvement of ongoing programs and learning. The relationship of good governance and M&E can be better understood from figure 11.2.

FIGURE 11.2 **Relationship between good governance and M&E**



GOVERNANCE AND M&E IN INDIA: AN OVERVIEW

India's constitution provides for a republic with a democratic, secular, and socialistic form of society: it places the principles of universal equality and social justice on a high pedestal. Appropriate affirmative action by the administration aims to raise the standard of living for the less socially and economically advantaged so that the gap between them and the rest of the society tapers off, and they can join the mainstream of development. The government aims to translate this intention into reality through a series of social and economic development policies, plans, and programs, through massive investment in the social sector. But huge investment does not necessarily produce the desired outcomes, unless it is operated in a scenario of sound governance. Good governance is critical to ensure that these investments lead to significant outcomes and impacts on the ground, through the efficient use of allocated resources; optimal management of public service delivery; and effective management of natural resources.

In a democracy, rising income levels also bring with them rising expectations among citizens, and a demand for good governance at all levels of the government: national, state, and local. Experience over the past over six decades indicates that while significant progress has been made in India

on most of the economic and social parameters, the impacts are not commensurate with the resources utilized, and could have been vastly superior with better governance. A significant step-up is required, through systemic improvements in implementation; increased efficiency of public agencies in the delivery of services to consumers; and tackling the menace of corruption, which has siphoned off huge chunks of public investments.

The Indian system of governance has two types of actors—formal and informal. **Formal actors** include the national and state legislative bodies, the judiciary, government functionaries, and constitutional bodies such as the Election Commission, the Comptroller and Auditor General, the Central Vigilance Commissioner, and organized industries and services in both the public and private sectors. **Informal actors** are the multitude of civil society organizations, academicians, the media, and the community. In the democratic set-up in India, both of these types of actors play an important role in planning and implementation. Together they contribute to the governance of the land as policy makers, enforcers of the enacted policies and laws, program planners, providers and receivers of various goods and services, and guardians of citizens' rights. In terms of economic and social development, there is broad consensus concerning the goals of inclusive and sustainable growth and gender equality, in line with global objectives and standards.

The National Institution for Transforming India (NITI Aayog), an organization that replaced the 60-year-old Planning Commission, is currently the main think tank for Indian policy makers on developmental issues.[4] NITI is looking forward to maintaining a state-of-the-art resource center, to be a repository of research on good governance and best practices in sustainable and equitable development, as well as to help them disseminate results of such research to stakeholders. NITI is also responsible for actively monitoring and evaluating the implementation of programs, and for identifying the resources needed to strengthen the probability of success and the scope of delivery. This is the main arm for M&E in the country: thus it is aimed at both strategic policy and program frameworks, and is monitoring both their progress and their efficacy.

Recent Initiatives for Good Governance in India

India has a long history of policies and programs directed toward economic and social development, tuned to the principles of inclusiveness (five-year plans; for instance, see Government of India 2013). However, despite lofty ideals and goals, performance has often fallen short of expectations. The present government came to power with a motto of "minimum government and maximum governance." In its objective of providing good governance, it has been seeking to simplify the delivery of services to citizens, and make the process as transparent and corruption-free as possible, primarily through the application of information technology as an interface between the

[4] http://niti.gov.in/content/overview.

government and citizens. Some recent initiatives aimed at these goals can provide learning.

Digital India. Digital India is a flagship program of the government of India, with a vision of transforming India into a digitally empowered society and knowledge economy.

The program builds on various e-governance initiatives for wider sectoral applications, with an emphasis on citizen-centric services. The main pillars of the program include universal access to mobile connectivity, public Internet access, e-governance, electronic delivery of services (*e-Kranti*), and information for all.

Jan Dhan Yojana.[5] A vast majority of the Indian population have traditionally depended on informal financial services and remained outside of the formal banking system, making it difficult for service providers to reach them through formal channels. This informal system also gave birth to corruption. Jan Dhan Yojana is an important step taken by the government to mainstream that part of the population that has previously not been covered by banking services into the formal system by enabling and encouraging them to do so through incentives. Regular monitoring has indicated that until December 21, 2016, about 260.3 million savings bank accounts were opened under the scheme, 158.6 million of them in rural areas, and 101.7 million in urban areas, accounting for a total deposit balance of Rs 7,155.7 billion (or about $110 billion). This is a significant step toward financial inclusion that will facilitate seamless and direct transfer of subsidies and other benefits into beneficiaries' accounts, reducing the number of opportunities for funds being pilfered en route.

De-monetization and Digi Dhan Yojana. Over 80 percent of the country's economic activities are carried out in the informal sector, through informal financial and other transactions, with no accounts kept. This state of affairs leaves enormous room for underreporting or nonreporting of incomes, thus undercutting the potential for tax collections; allows for corrupt practices of paying in cash for irregular services to take place; and enables some people to accumulate enormous amounts of black money.[6] This money, in large quantities and in fake currency, enables the funding of drug-related activities, terrorism, and other antisocial activities. A recent action by the government aimed at dealing with corruption, black money, and tax evasion has been de-monetizing high-denomination currency of India (notes of value Rs 1,000

[5] Jan Dhan Yojana (literally meaning "people's money program") is aimed at ensuring access to various financial services including availability of basic savings bank accounts, access to need-based credit, remittances facility, insurance, and pension to excluded populations (e.g., vulnerable and low-income groups).

[6] Black money points to incomes and wealth from undisclosed and often illegal transactions, and on which taxes are not paid.

and Rs 500), which together accounted for about 85 percent of all cash in circulation, and issuing fresh currency to replace the deposits of old currency in the banks. Though the process has resulted in some temporary inconvenience to the people, long-term gains are expected, through increased tax revenue and a reduction in corrupt practices. There has been a visible decline in terrorist activities.[7]

Simultaneously, the government has launched a massive program to enable and encourage both consumers and traders to learn and to use noncash (digital) means of money transfers. It is expected that this changeover to an economy that is less dependent on cash transactions would make transactions more transparent, boost tax income, and make it easier to curb corrupt practices. A continuous monitoring process has helped to assess the problems coming out of these new initiatives, and to take corrective actions immediately. For example, to ensure that the lack of Internet connectivity and education do not hamper the practice of digital money transactions, a new app, Bharat Interface for Money (BHIM), which does not require Internet connectivity and can be used even by people without education, has been launched.[8]

Clean India Mission. This is another major initiative of the government aimed at making India open defecation–free by 2019 and making people aware of the importance of keeping their environs neat and clean, through an aggressive awareness-generation campaign involving prominent people from all walks of life, and providing financial assistance for building household and institutional sanitation facilities. M&E is a regular part of this program. Cities, towns, and villages are being ranked according to their level of cleanliness. Success stories are being disseminated across the country, and the work is going on with active community participation.

Make-in-India and Skill India. Other major developmental initiatives with an equity focus include the Skill India and the Make-in-India programs. The main goal of Skill India is to create opportunities, space, and scope for the development of the talents of Indian youth, and to further develop those sectors that have already been playing a role in skill development for the last so many years, and also to identify new sectors for skill development. This new program aims to provide training and skill development for 500 million youth by 2020, and to cover each and every village.

[7] See, for instance, *India News* October 8, 2017, referring to the finance minister's observations on the subject, www.dnaindia.com.

[8] The BHIM app enables people to make simple, easy, and quick payment transactions using a unified payments interface (UPI) with just a mobile number or UPI ID. Pioneered and developed by the National Payments Corporation of India, BHIM was launched by the prime minister, Narendra Modi, on December 30, 2016, to usher in a financially inclusive nation and a digitally empowered society.

The purpose of Make-in-India is to encourage local and international manufacturers to set up production facilities within the country to boost production and employment.

All of these initiatives include M&E components: thus, M&E is being mainstreamed into the development process.

Strengths and Weaknesses of M&E in India

India, with its history of more than six decades of developmental planning supported by a network of M&E institutions and activities, is not new to this sphere of activity. But the growing importance of SDG-oriented M&E demands a stronger and more focused approach to M&E, supported by an M&E-enabled atmosphere, capacity building, and evaluation knowledge-sharing mechanisms.

Institutional mechanisms. The Development Monitoring and Evaluation Organization, a component of NITI Aayog, aided by its 15 regional offices across the country, keeps track of the developmental agenda in the context of SDGs, with evaluation as a priority. Apart from this centralized institution, most of the ministries and departments in the national and state governments have their own M&E systems.

Capacity building in M&E. Multipronged efforts are being made to develop national M&E capacities. A number of national institutions, such as the National Institute of Labour Economics Research and Development (NILERD), and international institutions such as the International Initiative for Impact Evaluation (3ie) and the Centers for Learning on Evaluation and Results (CLEAR) are organizing workshops, sensitization programs, and both short and long-term training programs for this purpose. Various states have approached NILERD, asking them to organize short-term programs for their officials. This indicates a growing awareness of the need for M&E capacities.

Toward a national evaluation policy. In spite of a long history of development evaluation, and the amount of emphasis currently being placed on M&E, India still does not have an explicit national evaluation policy. The strong need for a national evaluation policy that will provide a framework that defines the principles governing the role of M&E in development; the approach, quality, methods, and ethics to be ensured in the practice of development evaluation; utilization of the evidence-based results of such evaluations; and, importantly, the human and material resources to be optimally allocated for this purpose, is increasingly being recognized.

The Evaluation Community of India. Due to the growing demand for evaluations, and for capacity building from various corners, it was felt there was a need for a platform where planners, implementers, evaluators, and communities could come together to discuss various issues relating to M&E. Such a platform was launched in 2015, through the Evaluation Community of India (ECOI), a voluntary organization for professional evaluation (VOPE) with

the motto "to share and learn." ECOI has action groups working on various aspects of evaluation, such as preparing a draft national evaluation policy, capacity development in evaluation, and so on. ECOI is looking forward to networking and interactions with various partners to collaborate in further developing evaluation culture.

RECENT DEVELOPMENTS IN M&E AND GOVERNANCE IN SRI LANKA

Over the years the important role of M&E has been well recognized by the government of Sri Lanka.[9] Sri Lanka was one of the first countries to establish a web-based project monitoring system, e-PMS, to track projects implemented across all ministries. A homegrown electronic system was a significant aspect of that set-up. It was established in the then Ministry of Plan Implementation, to track financial and physical progress in implementation, and the results of all development projects and programs. The system could generate project information donor-wise, sector-wise, and ministry-wise. The Department of Project Management and Monitoring (DPMM), which has the mandate for M&E, has now replaced this system with a new Integrated National Development Information System.

However, the system is more or less confined to output-based progress monitoring of various development programs. In the context of SDGs, and the growing demand for effective monitoring and higher-level impact evaluations, the need for a comprehensive national M&E system is being increasingly felt. The Global Evaluation Agenda that evolved at Kathmandu in 2015 (EvalAgenda 2020) aims to strengthen the enabling environment for evaluation; develop institutional capacities; build individual capacities for evaluation; and support links among these first three dimensions by all stakeholders—governments, parliamentarians, VOPEs, the United Nations, foundations, civil society, the private sector, academia, and other interested groups—working together. In response to these developments, the Parliamentarians' Forum for Development Evaluation (PFDE) South Asia project team organized a series of events to facilitate the establishment of a national M&E system in Sri Lanka. Groundwork for such a system has recently been laid by the initiatives of VOPEs, the Sri Lanka Evaluation Association, the South Africa Monitoring and Evaluation Association, and the Malaysia Evaluation Association, together with their respective government agencies, and with the support of the EvalPartners Peer-to-Peer Small Grants Programme.

Sri Lanka's National Evaluation Policy

Although Sri Lanka commenced evaluations in the 1990s, the National Evaluation Policy (NEP) process had not been continued consistently until recently: one reason for the delay was the lack of its endorsement by the government. The DPMM is the department within the Ministry of National Policies and

[9] Paper submitted by Priroshini Trikawalagoda to the APEA conference 2016.

Economic Affairs that has the mandate for M&E, under the leadership of the Prime Minister. DPMM was stimulated to take a lead in the NEP process after the implementation of EvalPartners Peer-to-Peer Small Grants Programme 2015: one of outputs of this was the draft preliminary action plan formulated to roll out the NEP. A draft of the policy initially developed by Sri Lanka Evaluation Association is being refined through a series of consultative processes between DPMM and other stakeholders that has been organized by PFDE and supported by EvalPartners, the EvalGender+ initiative, and UNICEF. A road map—another output from the stakeholder consultation—is expected to help guide the process for obtaining Cabinet approval for the NEP.

Motions in the Sri Lankan Parliament

Two significant and highly encouraging events have recently taken place in the Sri Lankan Parliament. Two adjournment motions were moved in the Parliament: one to formulate a NEP, and the other to allocate funds for evaluation. The first motion was made in August 2016 by a member of parliament who proposed formulating a NEP, and an evaluation system for the country to assess whether the anticipated results from development programs have been achieved. Making the motion, the honorable member stated that Sri Lanka had pledged to achieve the SDGs by 2030, and emphasized the importance of the role of evaluation in assessing whether the anticipated results from development interventions have been achieved. He cited examples from several countries where evaluation systems have been established, and stated that around 20 countries have already formulated NEPs, demonstrating his interest in and commitment to the cause. The same member also called for a separate allocation of funds from the national budget to be set aside for evaluation. All of the parliamentarians who participated in the debate were in favor of both motions. It seems likely, therefore, that the Sri Lankan Parliament is keen to adopt a NEP.

The parliamentarians who joined the debate agreed in unison that the present national government, with the president and the prime minister representing the two leading parties, presents a good opportunity for implementing a NEP. PFDE–South Asia is closely supporting reactivation of the NEP process, to be backed by evaluation capacity development. Being the heads of the ministries concerned with the subject, the prime minister and his deputy are encouraging the initiatives.

During the debate, Malaysia, Nepal, and South Africa were cited as examples of countries that had commenced the process ahead of Sri Lanka. It is noteworthy that globally this may be the first-ever motion on evaluation moved in a parliament. Nevertheless, if the evaluation process is to function independently, a systematic framework that includes ministries, departments, boards, and so on, would have to be developed. For this purpose, a sizable allocation of funds is needed. These funds are worthy investments, since evaluation will facilitate the achievement of the objectives of the projects. Furthermore, the lessons learned from the process will help in effective decision making in the short, medium, and long term for the projects being evaluated, as well as when implementing future projects of a similar nature.

An Online Web-Based Project Monitoring System

During the stakeholder consultations, the need to integrate areas of interrelated work, evaluation, data, and indicators with a strong information system that would be backed by relevant, high-quality, disaggregated, comparable, and timely data was emphasized. The existing system needs to be modified and upgraded so that it can cater to current needs. However, because it has no local partner the cost of maintenance is exorbitant. The system has been handed over to the Information Communication and Technology Agency, the government's information and communications technology arm, which is in the process of entering into a memorandum of understanding to modify and upgrade to a more user-friendly system. Some of the other issues currently faced by DPMM are inadequately trained staff in line ministries due to transfers to other departments with no succession planning, and delays in updating the system by the line ministries. Furthermore, it is difficult to verify the accuracy of data when the observations of senior management are not entered into the system. More stakeholders would be encouraged to use the system if it could be made more user-friendly, and if project progress could be tracked and used to contribute to informed decision making. PFDE–South Asia has also recognized the need to build awareness and strengthen the Department of Census and Statistics, and has invited them to participate in evaluation capacity development programs.

Capacity Building of Public Sector Officials

In Sri Lanka, several capacity-building initiatives have taken place, including training workshops for government officials, and study tours to Malaysia and South Africa. The training workshops included evaluation, the Logical Framework Approach in evaluation and designing and managing evaluations.

A four-member delegation comprised of representatives of DPMM, and two representatives from PFDE–South Asia, visited South Africa with the objective of developing a preliminary action plan to implement a national evaluation system in Sri Lanka. The delegation had the opportunity to learn about the South African system of government and the application of planning, monitoring, and evaluation by the South African Department of Performance Monitoring and Evaluation. Discussions included the timelines used in developing a macro PM&E system, the rationale for PM&E, the various roles and responsibilities of different organs of the government, the outline of the department's work in the context of the central government, the National Development Plan 2030, the National Evaluation Plan, and the problems PM&E aims to address. The web-based system of frontline monitoring of the presidential hotline for community problem solving was also demonstrated.

The delegation also visited a participatory workshop on preparing terms of reference for an integrated development plan to ensure safety in the Republic of South Africa where they learned about the Management Performance Monitoring Tool (MPAT) and how performance is measured in departments using scorecards. The study tour was an important step in the long-term evaluation capacity-building program in Sri Lanka, supported by

the EvalGender+ network and the UNICEF country office. Such international experiences are very important if countries are going to work together to achieve the SDGs within the scheduled time frame.

In short, substantial efforts have been made to develop an evaluation culture in Sri Lanka, but there is a still need for the following:

- Strengthen parliamentarians' desire to use and demand evaluation
- Strengthen the evaluation capacity of district development committees where a decentralized budget is used
- Government endorsement of the national evaluation policy

GOOD GOVERNANCE AND M&E IN BHUTAN: OVERVIEW

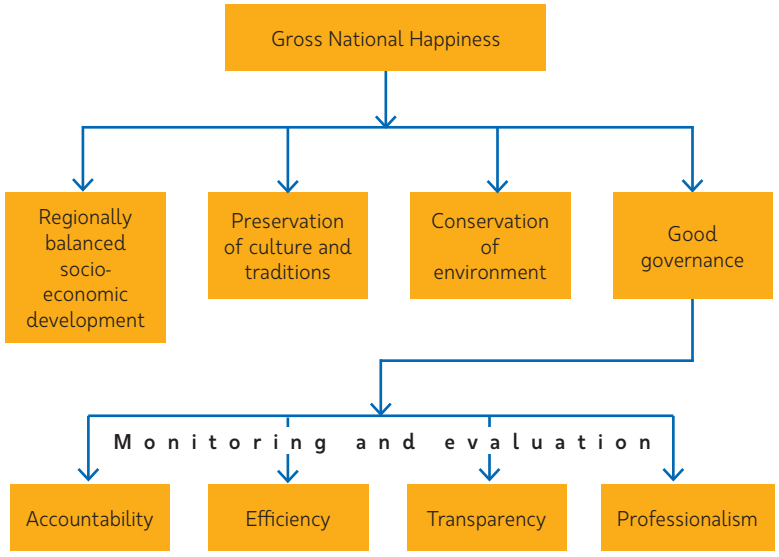
Bhutan's economic plans and programs are guided by an overarching development philosophy of Gross National Happiness (GNH). The concept of GNH attempts to ensure that an economic activity not only reaps material benefits but also positively impacts issues like equity, sustainability, preservation of the environment, and culture. GNH is supported with four pillars: good governance is the fourth pillar, which is the underpinning for the success of the other three pillars.

Good governance in Bhutan is characterized by four features: accountability, transparency, efficiency, and professionalism. While the concept of M&E is a cross-cutting phenomenon transcending all four pillars of GNH, the practice of M&E is more obvious and apparent with respect to good governance. Evidence-based practice of good governance is desirable for ensuring success of the other three pillars. Figure 11.3 presents a flowchart depiction of GNH, good governance, and M&E in the Bhutanese context.

Because of its success in achieving the Millennium Development Goals, under the development strategies of the GNH philosophy, Bhutan involuntarily has already begun the implementation of development activities for the SDGs. This is due to the fact that under the GNH development strategies, activities desired under the SDGs have inherent links with existing development plans and programs. For instance, the GNH pillar concerning "regionally balanced equitable socioeconomic development" has inherent links with SDGs 1, 2, 3, 4, 5, 7, 8, 9, 10, and 11. Similarly, the pillar of "preservation and promotion of culture and tradition" has links with SDGs 11 and 12. The third pillar, "conservation of the environment" can conveniently absorb SDGs 6, 7, 12, 13, 14, and 15, while the "good governance" pillar is related to SDGs 16 and 17.

Bhutan's Planning Commission, known as the GNH Commission (GNHC), is the central coordinating agency for development PM&E. In a bid to streamline and institute an effective system for M&E of development plans in the country, in 2006 the GNH Commission developed a national M&E system, as a standard system for monitoring and evaluating the development plans and programs administered by ministries and agencies. A dedicated unit, known as the Research and Evaluation Division, is in place at the GNH Commission.

FIGURE 11.3 **GNH, good governance, and M&E in Bhutan**



The Status of M&E in Bhutan

A strengths, weaknesses, opportunities, and threats (SWOT) analysis of Bhutan's M&E system was undertaken during one seminar on evaluation in Bhutan in March 2013, conducted by the GNHC and UNICEF-Bhutan. It was found that the evaluation system was weak; the technical capacity to conduct, commission, and manage evaluations was lacking; and the demand for evaluation was low. It was also noted that evaluations in Bhutan were generally donor-driven. These factors posed challenges for strengthening the evaluation culture in Bhutan. It was also recommended that a nonprofit association and a network of evaluators be established, to provide the much-needed platform to promote evaluation in Bhutan.

Therefore, evaluation is still at the nascent stage to this day. Evaluation is less understood and appreciated. Neither is there a demand for evaluation nor the supply. To this extent, there is a lack of capacity in evaluation.

On the contrary, there has been a considerable progress on the monitoring aspect. The sense of monitoring is not only becoming firmer in the system, but it is easier and more straightforward for agencies for implementation and oversight purposes.

In order to streamline, strengthen, and institutionalize the evaluation system, the national evaluation policy, and the national evaluation protocol and guidelines have already been formulated, and are awaiting the government's approval. Several evaluations of development policies and programs have also been conducted since 2013, through the Research and Evaluation Division of the GNHC, and in collaboration with government ministries. These

initiatives are geared toward promotion of the demand and use of evaluation by governments and parliaments so as to inform policy development and increased social accountability to citizens through evaluation.

Bhutan's Planning and Monitoring Process

Bhutan's development plans and programs are based on overall five-year plans (FYPs); currently, the country is in its 11th FYP (2013–18). FYPs are prepared through government consultations with implementation agencies, both at the national and grassroots levels. Plan consultations are preceded by issuance of planning guidelines to agencies. As part of results-based planning and management practices, plans are corroborated with identified result areas known as national key result areas, sectoral key result areas, key performance indicators, and specific key interventions. Individual FYPs are prepared by the respective ministries, autonomous agencies, and local government agencies, and their plans must be aligned with the national priorities. Based on the approved FYPs, annual work plans (AWPs) and budgeting are prepared and executed during the year.

Appropriate information technology (IT) systems are employed at various stages in the planning and monitoring continuum. The formulation of overall FYPs is based on information provided by planning and monitoring systems (PLaMS). PLaMS also provide support during the preparation of AWPs and their implementation.

Annual budgets are prepared using a multiyear rolling budget system. Budget releases are made on a quarterly basis, upon receiving plan monitoring and progress reports, which are essentially both physical and financial progress reports. While these reports will be made by implementing agencies through the PLaMs, the Ministry of Finance will release periodic budgets through another IT system known as the public expenditure management system (PEMS). Every government transaction is conducted online through PEMS, upon submission of progress reports. Thus, the AWPs and the budget, during the implementation phase only, will have periodic monitoring and reporting requirements that must be followed by the agencies.

It is customary for the government to conduct midterm reviews of FYPs in the middle of the Five-Year period, typically when the plan has progressed two and a half years into the FYP. In addition, the present government has initiated the drawing of annual performance agreements between the prime minister and individual agencies to ensure successful implementation of the AWPs. The activities identified in the AWPs would be determined by the respective agencies and duly agreed upon with the prime minister. There will be an annual review of annual performance agreements with the implementing agencies.

Status of the Evaluation Profession in Bhutan

Reaffirming the importance of evaluation within the system, UNICEF-Bhutan has been engaged along with the GNH Commission since the first day of the dialogue on promoting evaluation in Bhutan. Equally, parliament, through its

various standing committees, has always been on the frontline to promote evaluation culture in the country.

In an attempt to promote evaluation culture in Bhutan, with the technical and financial support of UNICEF-Bhutan, and the administrative support of the GNH Commission, the Evaluation Association of Bhutan (EAB) was formed in 2013, with a multibackground membership. Its registration for formal recognition of CSO status is already in process. In collaboration with UNICEF, the Community of Evaluators-South Asia, and the GNH Commission, the EAB regularly conducts training and experience-sharing events for its members and other stakeholders. Thus, the EAB is working on creating a network of high-quality evaluators and linking them with other evaluation communities.

From 2009 until mid-2015, about 16 officials from the government had attended the International Program for Development Evaluation Training (IPDET) funded by various sources of training, a major portion of which was supported by Danida. This training contributed immensely to enhancing awareness and capacity in the Royal Government of Bhutan. About 24 Bhutanese with an interest in evaluation are members of the International Organization for Collaborative Outcome Management, which was established in 2010. In 2013, a few Bhutanese evaluators joined the Community of Evaluators as individual members: this has strengthened the discourse on development evaluation, and the evaluation culture in the country.

However, despite these steps forward, the evaluation profession has not progressed very much. Except for donor-funded programs, no evaluation by independent practitioners has been conducted for government programs. And even for donor-led evaluations of programs and projects, the evaluations are mostly carried out without established processes and standards. Nonetheless, there is a silver lining, with the government's relentless effort toward the promotion of both demand and use of evaluation, and the disclosing of policy development through increased social accountability mechanisms.

The more plausible solutions, at this stage, are to:

- Expedite formal government approval of the national evaluation policy, guidelines, and protocol;
- Recognize the EAB as a legal CSO;
- Upgrade the capacity of evaluation practitioners; and
- Enhance the utilization of evaluation reports by agencies.

CONCLUSION

Bhutan, India, and Sri Lanka are all South Asian countries, but their approaches to governance and M&E vary. Analysis indicates that Sri Lanka is still confined to an output-based monitoring system, although recent events suggest a growing realization of the importance of, and gradual transition to, higher levels of evaluation. The emphasis in India, with its long-entrenched M&E systems, has been on outcome monitoring and impact evaluations. However, it has no evaluation policy. This often results in inadequacy or a multiplicity of efforts in M&E, a lack of standardized practices of evaluations, and more

importantly, their inadequate utilization. Sri Lanka on the other hand, has initiated the process of developing an NEP, and outcomes have found a place in the national parliament. Sri Lanka has adopted a long and robust process of dialogue and consultation with various stakeholders to make their evaluation policy a reality, and a motion for specific fund allotment for evaluation has been raised in their parliament. In India, there is a strong need for evaluation policy, and the allotment of dedicated funds for evaluations to make them utility-oriented, as highlighted in recent stakeholder discussions. In recent times, India has witnessed initiatives toward good governance and promotion of the use of IT to make citizen services efficient and financial transactions transparent, thereby reducing corruption. Bhutan, meanwhile, has a totally different philosophy of good governance, and measures it by public happiness. Bhutan considers the sole purpose of development is making people happy. High incomes may lead to material benefits, but general happiness is more important than these material benefits. Therefore, in Bhutan M&E is a cross-cutting issue looking to assess impacts in terms of GNH.

All three of these countries are committed to achieving the SDG targets; and M&E, as well as good governance are essential tools in that process. All three countries share the view that capacity building in M&E is extremely important in order to achieve the SDGs. While in India evaluations are getting mainstreamed into the development agenda and there is a demand for evaluation from various stakeholders, in Sri Lanka and Bhutan evaluations are generally donor-driven. Now is the time for various countries to come together, share their experiences, and learn from each other for future action.

REFERENCES

ESCAP (United Nations Economic and Social Commission for Asia and the Pacific). n.d. "What Is Good Governance?" Fact sheet. ESCAP, Bangkok.

Government of India. 2013. "Twelfth Five-Year Plan (2012–2017), Vol. I: Faster, More Inclusive and Sustainable Growth." New Delhi.

UN DESA (United Nations Department of Economic and Social Affairs). 2007. "Public Governance Indicators: A Literature Review." ST/ESA/PAD/SER.E/100. United Nations, New York.

Chapter 12

Integrating Feminist Approaches to Evaluation – Lessons Learned from an Indian Experience

Rajib Nandi and Rituu B. Nanda

Abstract. *Feminist approaches to evaluation seek to unpack the nature of gender and social inequalities; treat evaluation as a political activity, not as a value-free assessment; and use it as part of the change process. In response to the increased attention being given to evaluation, and the possible role of feminist evaluation in influencing policy that could lead to gender equality along different dimensions, the Institute of Social Studies Trust, based in New Delhi, designed a four-year program. The aim of program was to enhance capacity and understanding of feminist evaluations of various stakeholders, with a conviction that it would influence gender-transformative policy making. This chapter discusses the objectives and activities of this program, and critically analyzes the lessons learned. It also highlights the important takeaways that can make evaluations transformative as far as gender-based inequalities and power dynamics are concerned.*

Rajib Nandi, Institute of Social Studies Trust, nandi.rajib@gmail.com; Rituu B. Nanda, Institute of Social Studies Trust and Constellation, rituubnanda@feministevaluation.org.

The Indian economy and society have been struggling with persistent gender gaps and inequalities. Despite substantial progress over the years, gender gaps continue to persist in education, health, participation in the workforce, and decision making (Hay et al. 2012). Poverty, early marriage, malnutrition, and lack of health care during pregnancy are associated with high levels of maternal and infant mortality. Data show that in India almost 60 percent of girls are married before the age of 18 (UNICEF 2014), and nearly 60 percent of them bear children before they are 19 (Young Lives 2016). In addition, almost one-third of all babies are born with low birth weight. Although gender parity in school enrollment has largely been achieved, there are gender differences in the reasons for dropping out, for irregular school attendance, and for the pathways that open up through education. Against a male workforce participation rate of 53.0 percent in rural areas and 53.8 percent in urban areas, the female workforce participation rate was 30.0 percent and 15.4 percent in 2011 (GOI 2016). Women continue to be employed mainly as "marginal" workers, in home-based, informal economy work, and as unpaid family labor. They remain underrepresented in decision-making positions, even though quotas have enabled more than a million women to enter local governance institutions.

Because gender norms affect all aspects of work and life, and gender is seen as a cross-cutting issue, it often becomes invisible in policy and planning documents. For example, neither India's midterm appraisal of the Eleventh Plan, nor the issues for approach to the Twelfth Plan, mention "gender equality and empowerment" as a separate tangible goal, and discussion with planners reveals that it is seen as a "cross-cutting" factor (Planning Commission 2011b, 2011c). The reluctance to make gender concerns more explicit, which is derived from ignorance of the facts about gender gaps, suggests that there is still significant doubt as to whether any alternative policy or program design would significantly alter gender-related outcomes or not. However, evidence from the work of feminist and gender-sensitive social scientists shows that identifying and addressing the factors leading to gender inequalities can indeed help develop strategies for both policy advocacy and implementation for better outcomes from a gender perspective.

The present chapter highlights the discussions and lessons learned from a four-year program of capacity building on feminist evaluation that was prompted by an apprehension that evaluations with a feminist lens are capable of offering evidence-based policy advocacy that is oriented to gender equity and social justice.

The authors consulted different program documents including the program proposal, narrative reports, workshop reports, and the evaluation reports drafted by the external evaluators in order to write this chapter.

ENGENDERING POLICY THROUGH EVALUATION: BACKGROUND AND OBJECTIVES

Gender in Indian Policy and Planning

The ways in which a gender and equity lens in evaluation can help improve policies and programs is an important message for policy makers to receive.

On the other hand, it is widely recognized that evaluation has been emerging as a critical space through which gender and equity questions can be brought back into policy discourse.[1] The Indian government's Eleventh Plan adopted a gendered lens to initiate a process of systemic improvement in the lives of women and children. But the plan's midterm appraisal shows that while certain sectors have shown remarkable improvement, others are lagging behind (Planning Commission 2011c). The approach to the Twelfth Plan shows that only 35.8 percent of the Eleventh Plan outlay has been allocated during the first three years of the plan (Planning Commission 2011b). Schemes for single and internally displaced women, domestic workers, and minority women, to name a few, have not found a voice in the first half of the Eleventh Plan (Planning Commission 2011a).

During the formulation process of the Twelfth Plan, the coalition of civil society organizations and the planning commission was strengthened. The planning commission invited civil society groups to contribute to the preparation of the approach paper for the Twelfth Five-Year Plan, which is aimed at achieving faster, more sustainable, and more inclusive growth. The United Nations Development Programme (UNDP) India supported a facilitation process that provided women and men from marginalized communities living in remote corners of the country with an opportunity to voice their opinions on key development issues, and in doing so, marked a significant step in making the planning participatory. Planners interacted with 16 population groups comprised of *dalits*,[2] migrants, the urban poor, ethnic minorities, Muslims, people living with HIV/AIDS, and transgender persons, among others (WNTA 2011).

Consequently, the Twelfth Plan approach paper acknowledged that the plan must break the vicious cycle of multiple deprivations faced by girls and women because of gender discrimination and undernutrition. The paper ensures that ending gender-based inequities, discrimination, and violence faced by girls and women must be accorded the highest priority, and that this needs to be done in several ways. The midterm appraisal report of the Eleventh Plan acknowledges a need to undertake impact evaluations of intended outcomes. To undertake such evaluation, it has been decided to establish an independent evaluation organization linked to, but distinct from, the Planning Commission (Planning Commission 2011c).

Within this environment, there is some evidence that greater attention is now being paid to government-commissioned independent evaluations, as well as to the methods used and the findings. Acknowledgment by the government has created a tacit space for discussing gender-responsive evaluations in India. A meeting of the evaluation agencies and implementing agencies of the Support to Training and Employment Programme (STEP) for Women of the Indian government's Ministry of Women and Child Development was

[1] *Transform: The Magazine for Gender-responsive Evaluation*, Issue 1, June 2015. UN Women, Independent Evaluation Office.

[2] *Dalit*, meaning "oppressed," is a term for the members of lower castes of India.

convened in May 2011 for the first time since the program had started in 1986, to reflect on the design and evaluation processes of the program. The participants provided their thoughts about a systemic development of the evaluation process for the program. The increasing demand for greater transparency and openness is further reflected in the recent citizen agitation demanding an anti-corruption bill,[3] and in the use of the Right to Information Act to ensure public accountability. The act is becoming more crucial for the vulnerable sections of our society as they battle for social and economic justice. This is particularly true for women's issues. Some recent reports show that groups of women have also been able to fight gender-based discriminations through the Right to Information Act (Bakshi and Bhattacharya 2010).

A meta-evaluation of the STEP was conducted in 2012 by the Institute of Social Studies Trust (ISST), using a feminist lens. The study points out that care responsibilities and constraints on mobility can influence outcomes, and need to be factored into assessments of both design and outcomes. This study was an attempt at a formative meta-evaluation using a synthesis method, and with the purpose of using completed evaluation reports to inform and strengthen future evaluations ("Introduction," Sudarshan, Murthy, and Chigateri 2015).

The meta-evaluation of the STEP has provided critical insights, as follows:

> ...the importance of assessing whether the trainings were contextualized to the sector and catered specifically to the needs of the women was also emphasized. The argument made by the evaluating agency was that the trainings would not be effective otherwise, and that women would just remain recipients of STEP. In order to empower the women, the trainings had to be linked to the lives of the women. Furthermore, the trainings had to cater to the functional requirements of the women. For instance, it was recommended that the legal training not be limited to awareness of legal rights but should also address functional and transactional legal literacy in the particular context of the sector, for instance to know what happens in the case of non repayment of loans, or to any other specific issue from a particular sector like farming or weaving etc. Therefore, in order to assess the effectiveness of the training component, it is also important for the evaluating agency to assess whether the training was contextualized and catered specifically to the needs of women. (Chigateri et al. 2015, 65)

Importance of Building Evaluation Capacity in India

There is a strong presence of international evaluators, both organizations and individuals, in India today. This has helped to generate a more visible discourse about evaluation. Questions are being raised regarding things such

[3] See, e.g., news reports at http://www.thehindu.com/news/national/article1607789.ece.

as the extent to which greater professionalization of the evaluation function is needed, and how important the role of contextual understanding and domain knowledge is.[4]

Whether there is a growing demand for the evaluation of programs and projects, or whether there is simply greater visibility of this area of work, there is a sense among many commissioners of evaluations that local evaluation capacity is weak: this means that even if more evaluations are being locally commissioned, there may be a sense of discomfort among the commissioners regarding their credibility. Shiva Kumar, for example, has commented that "Professionals carrying out evaluations in South Asia tend to be social science researchers, not trained evaluators. Many evaluators of development interventions and commissioners of evaluation have not fully realized that the competencies needed to become an evaluator are different from, though complementary to, those needed for conducting social science research" (Shiva Kumar 2010).

Gender-transformative approaches to evaluation seek to unpack the nature of gender and social inequalities: further, they see evaluation as a political activity, not as a value-free assessment, and use it as part of the change process. Responding to the increased attention being given to evaluation, and the possible role of feminist evaluation in engendering policy and supporting changes that lead to gender equality along different dimensions, a consultation on gender and participatory evaluation was organized by ISST in August 2010 in New Delhi. There is, so far, little writing on gender-transformative evaluation tools and frameworks, and little on the difference made to evaluation findings by using a feminist lens. This workshop brought together a group of feminist and gender-sensitive equality advocates who, as social science researchers, have carried out evaluations and not had the opportunity to reflect on the role of these evaluations within their larger research agendas. Examples were given of the ways in which evaluation has been able to shift policy perspectives through the redesign of programs, and more importantly through systems change.

The participants of the above consultation might have lacked familiarity with mainstream evaluation theories and theorists, and the tools and frameworks associated with them, but the presentations by evaluators at the workshop showed that they have a good knowledge of the concerned sector. Evaluation findings can contribute to developing a future research agenda, leading to evidence-based policy recommendations, and drawing on the findings of a whole body of knowledge *including* evaluations. This positive contribution could be enhanced by strengthening the evaluation skills of social science researchers through exposure to evaluation theories and tools. At the same time, the immensely valuable set of knowledge and experience that has been gained by a number of feminist development practitioners will be unable to reach wider communities of evaluators if their evaluative

[4] For example, Abhijit Sen, in his keynote address on "The Role of Evaluation in Policy and Programming" at the Evaluation Conclave, Delhi, October 2010, made a strong case for strengthening evaluation as a discipline (Evaluation Conclave 2010).

writings are not framed and structured in the language that is recognized by the conventional evaluators.

Program Objectives

With this understanding, ISST in New Delhi designed a four-year program to address the increasing demand for transparency and accountability from program managers, including the government; the increasing interest in evaluation, but equally recognition of gaps in capacity; and the presence of a group of feminist social science researchers who have also engaged with evaluation.

ISST coordinated the program, which was entitled "Engendering Policy through Evaluation: Uncovering Exclusion, Challenging Inequities," from 2011 to 2015, in response to persistent gender inequalities in Indian society and the economy in various spheres.[5] In principle, gender-equitable outcome is regarded as a cross-cutting objective across all sectors of development in India. However, it is often observed that, at the policy level, there is a reluctance to make gender concerns explicit in program design and implementation. It seems policy makers are not convinced whether a gender- and equity-focused lens in program design and evaluation would significantly improve outcomes or not. On the other hand, there is sufficient evidence from existing research that alternative strategies can indeed help in reducing the gender gap. The motivation for the program then was "to try to change and improve things on the ground" by demonstrating evaluation approaches derived from feminist theories of social inequities.

The overall objective of the program was to strengthen the understanding of gendered implications of policies and programs, and to enable the formulation of gender-sensitive approaches. In particular, focus was placed on evaluating selected key issues related to education, health, governance, and livelihood. Specific objectives included building evaluation capacity using a feminist lens, and expanding research and inquiry into the benefits of doing so. The work would contribute to building the field of feminist evaluation, and building an active network of professionals who are engaged in advancing its theory and practice.

KEY ACTIVITIES AND ACHIEVEMENTS

Over a period of four years, this program has built a strong network of individuals and organizations in India that are interested in and working on feminist evaluation (table 12.1). For the purpose of capacity building, expanding a network of persons interested in evaluation with a gender and equity lens, and engaging policy makers, the following activities were undertaken.

[5] The program was jointly sponsored by the International Development Research Centre, Canada; and Ford Foundation, New Delhi.

TABLE 12.1 **Structure of the program**

Participants	Activities	Outputs
▪ Feminist evaluators ▪ Gender researchers ▪ Members from implementing organizations ▪ M&E personnel of different organizations ▪ Members from donor agencies/other development agencies	▪ Interactive training workshops ▪ Moderation of an online community of practice ▪ Participation in evaluation conferences ▪ Reflective writings on evaluation experiences with feminist lens ▪ Development of resources ▪ Dissemination seminars with NITI Aayog	▪ Resource materials on gender transformative evaluations (print and online) ▪ Edited collection of meta-evaluations ▪ Edited collection of feminist evaluations ▪ Short training videos ▪ Website and online community of practice

- The network-building activity was sustained through a feminist evaluation Google group; an interactive online community of practice on gender and evaluation; a Facebook page; and a Twitter account.
- The work of capacity building and building a base of knowledge on feminist evaluation was sustained through workshops, small research grants, and support for participating in international workshops and seminars.
- The knowledge produced through this process was disseminated through publications, and both print and online training videos.
- The program has tried to engage policy makers at different stages.

The program conducted six training workshops, which functioned both as capacity-building spaces as well as opportunities to share research and knowledge on feminist evaluation. The program also offered scholarships to attend international conferences.

The program offered 11 small research grants for reflective research on evaluations. Some of the sponsored studies were "A Study of Gender and Water, Sanitation and Hygiene (WASH) Evaluations in India" and "The Culture of Evaluations: Women Empowerment Programs under the Corporate Social Responsibility (CSR) Initiatives." There were also two studies on meta-evaluations: one in the field of education, the other on health programs of the government of India.

Four books were published as outputs of this program: a toolkit on gender-sensitive participatory evaluation methods; an edited volume, *Engendering Meta-evaluations: Towards Women's Empowerment*; an edited resource pack, *Resource Pack on Gender-Transformative Evaluations*; and an edited collection of feminist evaluations, *Voices and Values: The Politics of Feminist Evaluation*.

Since the inception of the program, knowledge sharing has been a key component. The base for this was provided by a feminist evaluation website,[6] various social media pages, and a Google group for those interested in feminist evaluation that was created in the first year of the program, as part of the knowledge sharing strategy. In addition, an online community of practice was set up in early 2013 to build a network of gender-responsive evaluation practitioners and researchers, including the core project participants, but also going beyond this group. The purpose was to provide an interactive, iterative platform that would build a knowledge base on gender and evaluation for the community of practitioners. This online community has more than 3,000 members from all over the world, and has generated new interest in gender and evaluation. It also provided an impetus for several organizations to seek collaborations with our feminist evaluation network. Since 2015, the online platform is also the online knowledge-sharing hub for EvalGender+.

There has been a widening of the networks beyond the initial base of project participants. While the project participants continue to form the core of our network, a broader network has been engendered by the online community, which has provided a much broader base for sustained conversations on gender and evaluation. Training videos on *What Are Gender-Transformative Evaluations?*, *Use of Evaluations in a Gender and Equity Context*, and *Principles, Values and Ethics of Gender-Transformative Evaluations* have been uploaded, which members have used, and through which they have also shared their inputs.

A key component of the program was also to engage policy makers, and to communicate the value of a feminist perspective in evaluations. Attempts to engage the policy makers have been underway since the beginning of the program. However, these efforts did not materialize until year 4, beginning with the International Year of Evaluation (EvalYear) events in India, for which ISST, the National Institution for Transforming India (NITI Aayog),[7] and the National Institute of Labour Economics Research and Development (NILERD) were co-organizers, along with other organizations.

Armed with some of the outputs of the program that directly address evaluation policies (for example meta-evaluations of government programs and state accountability mechanisms), the process of directly engaging policy makers proved to be more productive. Thus, in the fourth year, the program organized two policy workshops, in association with NITI Aayog and NILERD. And ISST's efforts to engage policymakers to use a feminist perspective in evaluations culminated in the launch of the Evaluation Community of India (ECOI), hosted by ISST.

The capacity to conduct evaluation and recognize the value of evaluations has been strengthened with each workshop, reflection session, and conversation on ISST's online platform. This group has contributed to building

[6] www.feministevaluation.org.

[7] NITI Aayog is a policy think tank established by the Indian government in 2015 to replace the erstwhile government institution called the Planning Commission.

the field in diverse ways, including the publication of papers and making presentations at national and international workshops and conferences, and at capacity-building workshops.

The group has been actively engaging with ongoing policy debates on reproductive and sexual health, particularly through the work on child and early marriages, and adolescent girls; on education (for instance, Sarva Shiksha Abhiyaan);[8] on the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), through a meta-evaluation study; and on livelihoods, through evaluating the governance of the irrigation sector through a gender lens.

Some of the participant organizations have incorporated their learning from the workshops into their own practice on monitoring and evaluation (M&E). For instance, Jabala, a community-based nongovernmental organization (NGO) that operates from Kolkata, conducted an internal evaluation on their economic rehabilitation program for survivors based on their learning from the workshops. The Child in Need Institute (CINI), an NGO based in eastern India, collaborated with ISST on using community-led participatory M&E tools in their project based in West Bengal. The Centre for Catalyzing Change (formerly known as CEDPA) is keen to develop a self-assessment tool for an end-line survey in one of their projects. And the Health Institute for Mother and Child (MAMTA) has started using participatory evaluation methods in their organization.

ANALYSIS

Given the wide variety of stakeholders that were targeted by the program, the amount and type of involvement, and the takeaways, also varied.

Practitioners were drawn from organizations that were conducting development programs, and were mid-to-senior-level professionals from different disciplinary backgrounds, including social science, management, legal, and others. For this group, evaluation knowledge was rudimentary, and their experience was a "third-party activity" to which they were subjected by donors. Against this background, participation in the project was found to be valuable in enabling ownership of evaluation within the organization: in assisting them to think evaluatively of their work; in unpacking gender dynamics that had been opaque in their program design and implementation; and in showing the way for creating an "evaluative culture" within their organizations.

The program has contributed to many key building blocks toward improving capacity building for organizations, and professionals engaged in development work at the grassroots level. Such organizations typically rely on third-party evaluations, and do not engage in viewing their own work using a feminist lens. This was the case even for organizations that were working

[8] Sarva Shiksha Abhiyan (SSA) is Government of India's flagship program for achievement of Universalization of Elementary Education in a time-bound manner, as mandated by the 86th Amendment to the Constitution of India, making free and compulsory education to children in the 6–14 age group a fundamental right.

on equity and women's rights issues. Such an engagement was found to be valuable in identifying dimensions for engagement, or for ongoing analysis of their work. Thus, feminist evaluation concepts of power and structures were becoming integrated into ongoing and new projects, and these opened up new ways of looking at empowerment. This capacity is crucial in order for an organization to actively engage in ongoing self-evaluation that can feed into the design of more formal periodic evaluations. The capacity-building workshops enabled development practitioners to better understand and measure change processes, and reflect on their own experiences.

Capacity building for conducting formal self-evaluations by these organizations was found to be more limiting, especially as there tended to be rotation among the staff who attended, and the workshop sessions were not geared to be evaluation training per se. With the evaluative thinking on feminist issues that was provided by the program, individual organizations embarked on their own evaluations, with mentoring support from experienced evaluators, and also with the support of ISST.

More senior practitioners, such as those who were involved with joint review missions for the government's flagship programs, found the feminist evaluation concepts very helpful in assisting them to formulate questions that helped to push the analysis beyond the data that was available for these programs. Integrating these ideas with their government counterparts was also well received.

Researchers and evaluators working on feminist evaluation found the dynamic interaction on research issues very helpful in furthering their work. Such opportunities for reflection, interaction, and feedback are valuable for conducting good research.

The researchers who participated in the program were at both the junior and senior levels. The junior-level researchers valued the ability to conduct work and the opportunity to attend and present results at international conferences provided by the research grants, and the reviews and discussions that enabled them to complete their publications. The senior researchers, who had taken on higher-profile meta-evaluations of national flagship programs, were supported for conducting and publishing the work in an edited volume, which subsequently has been released at a policy seminar cohosted by NITI Aayog. Being able to influence policy makers was enhanced by the involvement of other experienced evaluators who were also involved in this program. The series of workshops provided a space for reflection and introduction to innovative ideas that spurred research productivity. And the e-network provided an ongoing space for discussion of issues, concepts, and methods, and for researchers to get feedback on their work.

There was relatively little engagement with decision makers and policy makers from institutional settings where programs were being designed and/or implemented; where evaluation training was being carried out; or with government officials. The concept underlying this work was that results emanating from the capacity building and research output would subsequently be able to influence decision makers.

The program outputs have the potential to influence many of these groups of decision makers. For example, the training modules and resource

pack could contribute to the integration of these methods in standard evaluation trainings; and research outputs such as the compiled meta-evaluations of national flagship programs could influence a closer look at their implementation and modifications. The dissemination and outreach workshops and seminars conducted in the final year of the program highlighted the fact that such influence was indeed being generated.

Organizational leaders wanted more hands-on training in order to be able to conduct their evaluations. The program has introduced fresh perspectives to work being done by the NGOs that participated, and evaluative thinking was being integrated into their programs. However, they lacked the capacity and expertise to conduct their own evaluations, and would have liked to have an ongoing mentoring relationship established to facilitate that.

Evaluation practitioners, especially those who were active in mainstream program evaluations, had a mixed response. Some responded that the reflection and research conducted with their participation had enabled them to better interject feminist issues into program designing and implementation, though they felt that the terminology of gender-transformative evaluation was more acceptable within their own constituencies. Others felt constrained with their ability to apply feminist evaluation concepts in their practice.

Some of the reasons they cited for this included:

- Commissioners of evaluation, and most donors, pay only lip service to gender empowerment/transformative development;
- The need to demonstrate links to economic productivity with equity/gender empowerment programming; and
- Logistical issues in conducting fieldwork by female evaluators for getting women's perspectives.

The volume of high-quality research output and publications with international visibility has contributed to energizing this field of inquiry: a significant majority of participants said that this was a valuable contribution, and that a desire for ongoing engagement with the community of practice has been established.

PROGRAM OUTCOMES AS A WHOLE

Short Term

The regular workshops conducted over the course of the four years of program implementation created a vibrant space for interaction, reflection, and the sharing of innovative methods and approaches that engaged participants in a productive manner. Those who attended several of these events gained insights that they introduced to their organizations and into their work. The shared learning space and commonality of equity and feminist program and research interests of the participants was conducive to creating an effective community of practice.

The external visibility produced both within the participants' own organizations as well as in international forums, is likely to enable feedback loops that will contribute to benefits in the medium and longer term.

Medium Term

Publications of research on feminist evaluation include one edited collection of meta-evaluations of government programs (Sudarshan, Murthy, and Chigateri 2015), one toolkit on gender-sensitive participatory evaluation methods (Murthy 2015), one resource pack on gender-transformative evaluations (Chigateri and Saha 2016), and one edited volume on reflective writings on the practice of feminist evaluation in India (Sudarshan and Nandi 2018). These publications will provide an impetus for additional research and integration of a feminist lens in evaluations, program reviews, and their design.

The e-network, established and expanded internationally to more than a thousand members during the program period, is expected to be a major instrument for providing an open forum and long-term benefits for the continued development in the field of feminist evaluation that was spurred by this program.

Long Term

The yardstick that can be used to assess the longer-term outcomes of this program is the extent to which there is institutionalization of the key short and medium-term outcomes. The main evidence of this is from interviews with NGO practitioners who are integrating evaluative thinking into their programs, and who have acquired the capacity to apply a feminist lens to their programs. Those working in more mainstream programs did not face such a positive institutional environment.

The program was able to develop some integration of feminist or gender-transformative lenses in the evaluation training being offered at NILERD, which partnered in some of the workshops and outreach events. With the dissemination of publications that resulted from the program, there is potential for expanding such capacity building.

The impressive volume of high-quality research, and the publications that have been produced, will also contribute to longer-term outcomes.

SUSTAINABILITY OF THE PROGRAM

Human resource capacities built at the individual level by the program will be sustainable, given the depth and diversity of the discourse that has been provided. It is anticipated that these individuals will carry this capacity into their ongoing work, since they have been enabled to establish this type of discourse into any of the programs, evaluations, or research contexts in which they may find themselves.

Since the program was designed to influence capacity at the individual level only, at the organizational level influence is most likely to occur as the indirect result of participants who are able to integrate their learning on the

value and processes of introducing a feminist lens into the work of their organizations. As discussed earlier, this type of institutionalization was more apparent in the women's and human rights-based organizations, where the shift was related to introducing access to evaluative thinking, and a feminist lens enabled them to identify dimensions of their work that were not evident earlier.

However, the senior participants from mainstream development organizations felt confident that they would be able to make some inroads into the thinking at the institutions they worked with. At the same time, they also identified the constraints they faced. One of these is the challenge of using the terminology. Often the application of the term "feminist lens" has a tendency to address gender issues in a superficial manner, with little or no demand for systematic, gender-transformative programming and policy formulations, or their evaluation. And a number of the junior-level professionals did not feel competent enough to influence the work in their respective organizations.

The program has managed to have a small amount of direct engagement with national and state government officials concerning the value of using a feminist or gender lens in program evaluation, and in addressing program design for equity outcomes. This engagement only happened in the final year of the program, and was primarily the result of a few workshops and seminars that were cohosted with NITI Aayog. The previous chief executive officer of NITI Aayog and a number of senior-level officials participated, and engaged actively in those events. The program results presented were received very positively, and a favorable policy environment was evident. However, for this to be sustained and integrated into mainstream program evaluations and design will require concerted, ongoing work.

CONCLUSIONS AND RECOMMENDATIONS

Participants from rights-based organizations found value in learning from the program and the workshops. Many of them who are new to evaluation began to integrate evaluative thinking in their work.

The program provided a new way of looking at how to make evaluation gender responsive. But it was challenging at the organizational level, since new learning applications had to go through an internal process of explaining and convincing the leadership. Introducing a gender-transformative perspective in designing and implementing a program even at the organizational level would be slow and challenging.

Researchers opined that the program enabled them to connect with practitioners. As for using the learning in their work, a couple of them noted that the term "feminist lens" was not always acceptable to their colleagues, so they used the concepts but not the terminology. The online community was very helpful, and much was learned from the information sharing.[9] Overall, there has been a positive contribution to building the field for feminist/

[9] The online community of practice can be accessed at http://gendereval.ning.com.

gender-transformative evaluation and opinions on transforming policies and programs using gendered evaluation.

However, there are serious challenges in conceptualizing feminist evaluation. One major challenge has been confusion around the question of whether equity-focused and gender-responsive evaluation is different than feminist evaluation. The question was raised time and again, both inside the group and outside of it, whether feminist evaluation is a separate methodology or simply a frame of mind to use in looking at issues of inequality and gender-based power dynamics. There is also confusion as to whether feminist evaluation is an approach that could be used in any evaluation practice. This was discussed in some of the group workshops: whether the feminist approach should be integrated into the project design, or not. The question of whether to use a feminist lens in project monitoring was another area of concern.

During the midterm evaluation workshop, some of the participants raised the issue of wider acceptance of evaluation reports conducted with a feminist lens. One participant raised her concern that policy makers do not want complexity: they like to see a few clear results, stated in terms of numbers. This may go against the grain of feminist evaluation. The opinion was expressed in this workshop that the feminist evaluation approach cannot produce figures and numbers: rather, it captures social changes or program/project outcomes qualitatively. Perhaps this is the reason that policy makers and governments rely mostly on a set of institutes who practice and adopt more quantitative methods.

This brings up the second level of concerns for the group of feminist evaluators. The question was asked, how to bring feminist values into the evaluation findings, and how these values can be flagged so that policy makers will start noticing them. Other questions discussed were how to create an enabling environment for equity-focused and gender-responsive evaluation, and how to share lessons learned effectively in lessening gender inequalities. On this last question the group agreed that practicing feminist evaluation is part of a larger structure of feminist politics that ultimately tries to bring equality. This concern enabled the group to name feminist evaluation practice as gender-transformative evaluation in an Indian context.

During these interactive sessions, it was understood that there is a need for further dialogue and discussion among the members of the group and among both young and senior practitioners and evaluators. A number of participants in the program were open to sharing and exchanging. The recognition of the importance of converging various approaches and methods even increased greatly with the progress of the program. The group agreed on a crucial issue that feminist evaluation knowledge needs to go beyond feminist constituency and reach a wider audience. The group also discussed how to deepen the alignment between feminist researchers, mainstream evaluators, and formal and informal networks of evaluators, donors, university and research institutes, governments, and NGOs in order to build the field of feminist evaluation.

During the midterm external review workshop in 2014, the group agreed to strengthen the capacities of a larger group of development

practitioners and M&E personnel through developing an easy-to-read toolkit and the mentoring of young practitioners by senior feminist evaluators. They also suggested that the community of practice can facilitate a space where people will be able to interact, debate, discuss, and share knowledge.

A feminist evaluation curriculum for India and South Asia that is mainstreamed through the university system would help in developing capacity in feminist evaluation. To ensure that feminist evaluation is adopted at the organizational level, there is a need to orient staff, particularly the leadership. The group also felt a need to engage more with government officials, and with a focused and targeted approach.

A series of suggestions was rendered by the evaluators at the end of the four-year program. These suggestions came out through conversations with members of the group. One important suggestion was to continue organizing policy workshops, publishing policy briefs, and publishing in reputable journals and edited volumes in order to document the process of change that feminist evaluation has created. The second important suggestion was to find ways to increase visibility of the group of feminist evaluators in India. The third suggestion was to make efforts to integrate equity-focused, gender-responsive evaluation with mainstream evaluation methods such as utilization-focused evaluation, or impact evaluation.

In the longer run, organizational capacity building for integrating a feminist lens would be aimed at the foundation that has been built by this program. Attention is needed to support continued research on documenting the approaches and the value gained with feminist evaluation, and its integration into mainstream development research agendas.

Continuing the feminist evaluation e-network and community of practice is a priority. The recent development of EvalGender+ stepping in to support this is a big plus, and will help to maintain the momentum that has been generated. This platform has been valuable for encouraging exchange on theory and practice, creating a space in which to question and have a dialogue on issues, challenges, different methods, and their applications.

Outreach and dissemination of key messages and lessons learned from the knowledge products that have been produced for different audiences will help to expand the audience and understanding gained beyond those who are active participants in this work.

As a follow-up of the program, the core group of feminist evaluators might consider introducing courses in feminist/gender-transformative evaluation at evaluation training centers in India. Gender is becoming a standard cross-cutting objective in development programs, but not enough attention is being given to how the objectives can be attained in reality. At the same time, building of capacity at the organizational level for integrating a feminist lens in the design and monitoring of programs for producing gender-equitable results will be another key step. This can be linked to working with donors to promote building a culture of evaluation with a feminist lens.

The core group of gender evaluation network participants, along with ISST, are well positioned to expand collaboration with NITI Aayog and the state governments to review their evaluation guidelines and methodologies, and their M&E review processes. In some states, leaders of this feminist

evaluation network are already assisting with reform of the M&E processes, and are documenting their work so that it can be used to assist in cross-learning mobilizing support.

This would also be an interesting way to get the insights and involvement of the leaders in feminist evaluation, who otherwise do not have time to contribute to the online platforms or join the discussion groups.

Research funding for feminist evaluation is an ongoing constraint, and the online community of practice can take up this issue and find ways to establish a research fund for supporting ongoing research.

ACKNOWLEDGMENTS

The authors would like to acknowledge that the chapter is based on various project documents. A number of team members were involved in drafting those documents. The authors would like to mention particularly Katherine Hay and Ratna M. Sudarshan for conceptualizing the project along with a core group of feminist researchers and evaluators, and for developing the initial project documents and research papers. Two evaluation reports that were used as a source for this paper were drafted by Kalyani Menon Sen and Shubh Kumar-Range. The authors are grateful to all members of the feminist evaluation group for their contributions in the project and to the discussions. Finally the authors gratefully acknowledge the sponsoring organizations: the International Development Research Centre, Canada, and the Ford Foundation, New Delhi.

REFERENCES

Bakshi, M., and M. Bhattacharya. 2010. "Miles to Go: Effectiveness of RTI for Women." *Global Media Journal—India Edition* December.

Chigateri, Shraddha, Tanisha Jugran, Rituu B. Nanda, and Rajib Nandi. 2015. "Learning from STEP Evaluations." In *Engendering Meta-Evaluations: Towards Women's Empowerment*, Ratna M. Sudarshan, Ranjani K. Murthy, and Shraddha Chigateri, eds. New Delhi: Institute of Social Studies Trust.

Chigateri, Shraddha, and Shiny Saha. 2016. *A Resource Pack on Gender-Transformative Evaluations.* New Delhi: Institute of Social Studies Trust.

Evaluation Conclave. 2010. "The Evaluation Conclave 2010: Making Evaluation Matter. Communication Outreach Report." New Delhi.

GOI (Government of India). 2016. "Women and Men in India 2016." Ministry of Statistics and Programme Implementation, New Delhi.

Hay, Katherine, et al. 2012. "Why a Special Issue on Evaluating Gender and Equity?" *Indian Journal of Gender Studies* 19 (2): 179–86.

Murthy, Ranjani K. 2015. *Toolkit on Gender-Sensitive Participatory Evaluation Methods.* New Delhi: Institute of Social Studies Trust.

Planning Commission. 2011a. "Faster, Sustainable and More Inclusive Growth: An Approach to the Twelfth Five Year Plan." Government of India, New Delhi.

———. 2011b. "Issues for Approach to the 12th Five Year Plan." Presentation, UNDAF Orientation Workshop, New Delhi, May 12, 2011.

———. 2011c. *Mid-Term Appraisal: Eleventh Five Year Plan 2007–2012*. New Delhi: Government of India and Oxford University Press.

Shiva Kumar, A.K. 2010. "A Comment on 'Evaluation Field Building in South Asia: Reflections, Anecdotes and Questions.'" *American Journal of Evaluation* 31 (2): 238–40.

Sudarshan, Ratna M., Ranjani K. Murthy, and Shraddha Chigateri, eds. 2015. *Engendering Meta-Evaluations: Towards Women's Empowerment*. New Delhi: Institute of Social Studies Trust.

Sudarshan, Ratna M., and Rajib Nandi, eds. 2018. *Voices and Values: The Politics of Feminist Evaluation*. New Delhi: Zubaan.

UNICEF (United Nations Children's Fund). 2014. *Ending Child Marriage: Progress and Prospects*. New York: UNICEF.

WNTA (Wada Na Todo Abhiyan). 2011. *Approaching Equity: Civil Society Inputs for the Approach Paper—12th Five Year Plan*. New Delhi: WNTA and United Nations Development Programme, India.

Young Lives. 2016. "Tackling Child Marriage and Early Childbearing in India: Lessons from Young Lives." Policy Brief 29. Oxford Department of International Development, Oxford, UK.

Chapter 13

Evaluation Cooperation in West Africa

Abdoulaye Gounou

Abstract. *Evaluation is an increasing concern for the francophone countries of West Africa: however, for the most part impact assessment does not yet have the interest it should in these countries. To reverse this trend, under Benin's leadership, a capacity-building program has been initiated to promote impact evaluation as a tool for public policy analysis, support impact evaluation studies in these countries, and gain consideration for the results of impact evaluation within the framework of public management. This innovative project involves South-South cooperation between countries that share roughly similar economic, political, and social contexts, as well as a legal and economic framework that is converging. The growing interest of West African states in evaluation has led several countries to develop their own evaluation systems. In 2012, eight countries with similar interests met for a workshop on monitoring and evaluation. As a result of the workshop, three countries—Benin, South Africa, and Uganda—are continuing to cooperate by developing the Twende Mbele ("Let's All Move Forward") program to strengthen performance and evaluation monitoring. This program aims to strengthen national evaluation systems, and to gradually extend its interventions to other African countries through the development of appropriate tools for monitoring and evaluation, the capitalization of knowledge resulting from evaluations, and the sharing of national experiences in evaluation.*

Abdoulaye Gounou, Benin Directorate General of Evaluation, agounou0@gmail.com.

Francophone West Africa is identified as an area where there is a need for capacity building in evaluation. The practice of evaluation is itself very uneven and, depending on the country, the institutionalization of evaluation is still rudimentary. This observation was made during the francophone dialogue on evaluation capacity development held in Cotonou on July 7, 2015. More specifically, in the countries of the West African Economic and Monetary Union (WAEMU), the production and use of evidence through impact evaluations to inform public decision making is still very weak.[1]

Several initiatives at the national and regional levels have been undertaken to provide answers to this unpleasant picture as it has been observed in francophone West Africa. These initiatives include national evaluation days as well as regional programs such as Twende Mbele ("Let's All Move Forward") and the Capacity Building and Impact Assessment Program in West Africa (WACIE).

This chapter will show how such initiatives can help to effectively strengthen the institutionalization of evaluation, and improve evaluation practice in the francophone countries of West Africa. It will also show that the development of evaluation in these countries has also brought about the design and implementation of regional programs for the sharing of best practices, and support for sustainable capacity building.

The chapter is structured in four parts, plus a conclusion. The first part describes Benin's experience, in particular progress made in the field of evaluation and how cooperation has influenced this evolution. The second and third parts, respectively, present the WACIE and Twende Mbele programs. The fourth part demonstrates how effective implementation of these two programs and their interrelationship offers clear opportunities for the development of evaluation at the regional level.

BENIN'S EXPERIENCE IN REGIONAL COOPERATION

In 2007, Benin initiated the process of developing evaluative work by assigning this task to a ministry. Within the framework of the operational management of this function, the Office of Public Action Evaluation; the Bureau of Public Policy Evaluation; the Directorate General of Evaluation; and finally (and currently), the Bureau for the Evaluation of Public Policies and Analysis of Governmental Action were established.

The strategy adopted by this institution from its establishment in 2007 until the present basically involves benchmarking and integrating Benin into international evaluation networks. By participating in international evaluation conferences, Benin realized that South Africa and Uganda are in the lead in establishing national monitoring and evaluation (M&E) systems. In 2012, the government of Benin endowed the country with a national evaluation policy (NEP 2012–2021), and set up an institutional framework for public policy evaluation, which includes the organizations in charge of M&E and development planning.

[1] For more information, see Mendiratta (2011).

These advances at the national level have resulted in closer cooperation between Benin, South Africa, and Uganda, and as they appreciated the country's efforts in this area and its lead over other francophone West African counterparts.

Determination from its national stakeholders, supported by strong political will, has enabled Benin to join its peers from South Africa and Uganda in the evaluation cooperation program called Twende Mbele.

As members of the International Initiative for Impact Evaluation (3ie), South Africa and Uganda encouraged Benin to join this initiative for help in developing the practice of impact evaluation. In 2014, Benin signed a memorandum with 3ie that enables it to receive annual grants for the commissioning of impact assessments, and for capacity-building activities in impact evaluation.

When participating in 3ie's Evidence Weeks, Benin observed the absence of almost all the francophone countries of the WAEMU area. The initiative for a regional development program of impact assessment in the WAEMU area was designed, and eventually evolved into WACIE.

3ie's traditional field of intervention is impact evaluation, but the institution has broadened its scope to include the area of monitoring. Impact evaluation remains the organization's main focus, but a window is now open for providing support for the strengthening of national M&E systems, and the promotion of researchers from Global South countries, in an attempt to help reduce the gap between North and South in the field of evaluation.

Especially in West Africa, national evaluation systems are weak, the culture of evaluation is still in an early stage, and the practice of impact evaluation is almost nonexistent. This weak culture is caused by a series of difficulties that the promotion and development of evaluation faces, in particular low demand for evaluation in the region; weakness of national statistical systems; and the minimal degree of ownership of evaluation issues by states that have left the demand for and funding of evaluations to donors.

WACIE offers opportunities in the West Africa region for collaboration with the Twende Mbele program, which is a continental initiative. Benin's leadership in promoting evaluation has been demonstrated through its official support for WACIE, and its support for French-speaking countries in West Africa to look to Twende Mbele to help them make progress on governance and accountability issues. Benin promotes the Twende Mbele program in West Africa. And WACIE, which is a regional initiative, has become an opportunity for Twende Mbele to reach many more audiences.

THE CAPACITY BUILDING AND IMPACT EVALUATION PROGRAM IN WEST AFRICA (WACIE)

It is now recognized that rigorous impact evaluations can help quantify the socioeconomic impacts of projects and programs in several areas, and can provide a solid basis for consistent policy-making decisions. Although there is a growing global interest in impact evaluations, a report on trends in Sub-Saharan Africa revealed that there is little to no buy-in from countries and nationals of the continent (Mendiratta. 2011). Despite a considerable

increase in terms of the number of evaluations since 2004 (77 percent from 2004 onward), there are still significant disparities. For instance, impact evaluations in Africa are conducted much more frequently in anglophone countries (mainly Kenya, followed by Uganda) compared to francophone countries (18 percent).[2] In addition, only 11 percent of the studies[3] are conducted with African experts involved in the drafting of the research paper, although African partners (e.g., local NGOs, ministries, etc.) are involved in various stages of the implementation of the programs.

Over the past two years, discussions with key regional stakeholders, including the African Development Bank, the West African Development Bank, and the governments of Benin, Guinea Bissau, Niger, and Senegal have identified a real need for capacity building in order to produce high-quality research results, and to promote their use in the development and implementation of policies. While most governments in the West African region have monitoring systems in place, three key conditions limit rigorous impact evaluations: a lack of the human resources and institutional capacity that could help to conduct and use impact evaluations; a limited monitoring system for helping in the implementation of evaluations; and a low level of competence to conduct the required studies. WACIE was created to tackle these issues.

WACIE is a regionally based national initiative to improve welfare and development outcomes through decision making that is informed by research findings in the West African region. The initial phase of the program will begin in year 2017 and will last three years, followed by a consolidation phase, which may be carried out by a regional organization within WAEMU. The program is promoted by the government of Benin, with partial technical and financial support from 3ie.

Key Elements of the Program

As specified in the program documents, WACIE's overall objective is to improve welfare outcomes through informed decision making, based on evidence from research findings in the region. To this end, the program is characterized by several key elements:

- Initiating and managing consultations at the national level to identify key areas where evidence is needed for policy making and programmatic decision making
- Establishing and managing a funding mechanism to sponsor public policy impact evaluations
- Developing a set of standards, guidelines, and databases to guide and communicate evaluation results
- Designing, implementing, and disseminating assessments

[2] Angola, Cape Verde, Eritrea, Mauritius, Mozambique, and Rwanda are classified as neither English- nor French-speaking countries in the report. They account for about 12 percent of impact assessments.

[3] The statistics are based on 257 evaluations from the years 1982–2010.

- Providing training and capacity building to support the planning and implementation of impact evaluations, and the use of evidence in the decision-making process in the region

The proposed initiative covers a number of areas.

- **Capacity strengthening to produce and use research findings.**
  - The capacity to produce scientific evidence will be enhanced through the participation of local researchers in the funded studies.[4] Training in impact evaluation will also be offered within the member countries, in collaboration with the Regional Centers for Learning on Evaluation and Results (CLEAR).
  - The creation of partnerships, codevelopment, and coeducation of training programs with regional academic institutions such as the African School of Economics.[5]
  - The capacity to use research results will also be developed through one or more workshops targeted at decision makers that will be organized within the country.
  - The provision of grants to provide staff for trainings and conferences.
  - Support through an assurance of assistance for quality impact evaluation of local governments and partners in the region.
  - Development of a network to connect personnel, researchers, and decision makers in the region. Furthermore, study teams will be made available to conduct workshops with their respective implementing agencies, thereby enhancing impact evaluation know-how and the use of research findings in institutions for the conception and implementation of projects and programs.
- **Creation of a database of research results.** This will involve support for at least four impact assessment studies. These studies will focus on areas and research issues identified by WACIE member governments in West Africa. They will be based on consultation sessions supported by WACIE secretariat staff, and supervised by 3ie staff. Staff and study teams will liaise with line ministries to discuss opportunities for impact evaluation of selected programs and their implications, and 3ie will conduct the studies independently. The conception phase of the impact evaluation will be subject to formative evaluation of the implementation process.
- **Policy interpretation of research findings** 3ie has a well-developed system for supporting and monitoring the use of study results. Through this system, the project will support the production of the

[4] This is a condition that must be fulfilled in order to receive funding from 3ie.

[5] The African School of Economics is a private pan-African research university based in Abomey-Calavi, Benin.

operational synthesis of results, and one-page research recommendations on all studies, along with other guidelines derived from research findings that are relevant to help develop major policies.

- **Learning from international experience.** The program will include a grant component to enable participation in international conferences and workshops on impact evaluation, including preconference training. A total of 16 grants will be offered each year.
- **Program management.** The program will be managed by 3ie international staff based in New Delhi, supported by a secretariat located in Cotonou.

Logic of the Intervention

By the end of its implementation, the program is expected to contribute to improving the living conditions of the people of the WAEMU region through more effective public policies. The expected effects are:

- Improving the effectiveness and relevance of national and subregional projects and programs;
- Making use of systematic and sustainable evidence in the formulation and implementation of public policies; and
- Ensuring effective program management.

To observe these effects, the following causal pathways should be achieved:

- Evidence is produced and used to influence and support the formulation and implementation of public policies in each country, as well as subregional programs in the WAEMU area
- The WAEMU countries' national evaluation systems are strengthened and able to produce evidence for informed decision making

The intervention of WACIE is therefore designed to produce evidence that can highlight public policy choices in francophone West Africa, with a view toward contributing to the ultimate improvement of the living conditions of the people of the WAEMU area. We are, however, assuming that evidence may not necessarily lead to policy changes for various reasons: for example, because other policies have priority; or because doubts are raised about the evidence; or because trade-offs have been made to achieve political compromises.

In this respect, some of the risks are related to the lack of will of the decision makers or the political context in certain countries.

To lessen these risks, it is necessary to ensure the commitment made by all WACIE member states when they join 3ie. This commitment stipulates, among other things, that the member countries of 3ie will commit to using the results of the impact evaluations financed by 3ie in their countries, in order to improve the quality of public policies. Similarly, the ministers responsible for public policy evaluation in each WAEMU country will be involved in

defining the program's orientations, and will avert risks to the judicious use of the results and recommendations of evaluations.

To support the planning and implementation of impact evaluations and the use of evidence generated by such evaluations, WACIE plans to strengthen the technical capacities of the various stakeholders. Scholarships for training in impact assessment, or participation in technical workshops on impact evaluation in the subregion, will be awarded for this purpose. This ambition could be undermined by the lack of a training center or high-level technical workshop on impact evaluation in the subregion. For this reason, the program will support and encourage institutions and research centers that have the capacities to offer quality trainings or technical workshops on impact evaluation, to increase their offerings.

Milestones and Stakeholders

Six of the eight WAEMU countries (Benin, Burkina Faso, Côte d'Ivoire, Guinea, Niger, and Senegal), and the West African Development Bank, are now members of 3ie. 3ie has also already supported nine impact assessments (in Benin, Burkina Faso, Mali, Niger, and Senegal). These impact assessments were preceded by capacity-building workshops and consultations. 3ie has also supported four capacity-building workshops in the region (in Benin, Burkina Faso, and Togo).

WACIE will bring together as stakeholders the member states of WAEMU; WAEMU regional institutions (i.e., the WAEMU donor agencies); 3ie; and other nonregional donors (i.e., donor agencies outside of WAEMU), as well as other stakeholders.

- **WAEMU member states.** Each country participating in WACIE will be identifying and choosing a focal point for working with the WACIE secretariat. Focal points are public institutions with a legal mandate to work on impact evaluation issues with the support of their respective government authorities. The role of the focal points is to facilitate the implementation of WACIE at the country level: they will play a key role in the consultation process of identifying relevant projects in relation to the socioeconomic context of each country. In addition, they will ensure the effective involvement of local and national institutions in the initiative.
- **WAEMU regional institutions (WAEMU donor agencies).** These are the West African Development Bank and the WAEMU Commission. They represent WAEMU internal donors who will be supporting WACIE and helping to define strategic policy by participating in WACIE Advisory Committee.
- **3ie.** 3ie will act on the one hand as a catalyst throughout the initiative, by motivating regional involvement and participation in the program; and on the other hand, coordinating and supporting some WACIE activities. WACIE will be led by 3ie's main office in New Delhi. The design of WACIE assumes that there are resources for impact evaluation in the West African region, and that these resources need to be channeled through training and technical assistance.

- **Other nonregional donors (non-WAEMU donor agencies).** These are other technical and financial partners from the WAEMU member states that will cofinance the program: for example, the Agence Française de Développement (AFD), the African Development Bank, and others. They will help mobilize resources and define WACIE's strategic policy by participating in the its Advisory Committee. WACIE support modalities are presented later in this chapter.
- **Other stakeholders.** This group includes the project-executing agencies that will be selected under WACIE. It also includes research teams, and all institutions not included in the previous groups that will benefit from or contribute to the initiative.

STRENGTHENING PERFORMANCE AND EVALUATION MONITORING PROGRAMS: TWENDE MBELE

In March 2012, The South African Department of Performance Planning and Monitoring and CLEAR in anglophone Africa organized a regional workshop with seven African countries in Johannesburg in March 2012. At this workshop, Benin, South Africa, and Uganda recognized the similarity of their approaches to developing their national evaluation systems. Since then, they have collaborated by sharing guidelines and methodological tools, and attending training and events together. This has allowed for some cross-learning and the building of a slight, but effective, collaboration between the three nations. And in both anglophone and francophone Africa, CLEAR has worked with these three countries during the initiative.

Since 2013, the United Kingdom's Department for International Development (DFID) has been supporting the South African Department of Performance in their Strengthening Performance M&E project, which is focusing on the use of M&E in front-line service delivery and on citizen monitoring to inform the government on results delivered to the people of South Africa. The result of the project is that effective delivery of front-line services, a citizen-based surveillance system, and a results-based evaluation system support government accountability to the South African populace. A total of £2 million has been made available over 2013–15. The project has been successful: the systems have developed rapidly and have had impacts on the formulation and implementation of policies. This South African experience has been shared as a training in the framework of cooperation between Benin, South Africa, and Uganda.

These three countries have been active promoters of regional and international networks, including CLEAR and 3ie; both are supported by DFID. South Africa and the United Kingdom are also members of the Partnership for Open Government, which promotes accountability and transparency in government action. Other countries have been interested in what these three countries are doing, with study visits from countries such as Botswana, Kenya, Lesotho, Mauritius, Niger, Seychelles, and Togo. They are sharing their experiences and tools widely in the region, both in evaluation and monitoring, especially when they rely on the regional work CLEAR undertakes.

The Twende Mbele program aims to support emerging countries across Africa, enabling the three founding partners to intensify learning and to involve other partners that are determined to move forward with M&E. The program will also be built on global experience, using a range of M&E mechanisms that have been proven to improve government performance and accountability.

DFID is currently funding the African Networks Strengthening for Governance, Accountability, and Transparency (SANGAT) program, which supports three other African networks. SANGAT aims to take advantage of the existing projects, demand-driven processes, and peer learning in thematic areas. There are currently three SANGAT projects, focusing on areas from budget transparency and public finance to a high-level network dealing with major issues of prosecution and organized crime. Twende Mbele is being funded as a fourth component of the SANGAT program.

Key Elements of the Program

As specified in the project documents, the immediate result proposed for Twende Mbele is "the improvement of M&E systems (e.g., practices, policies, tools and procedures) demonstrated in partner countries based on shared experiences." The theory of change assumes that the M&E systems demonstrated in the project (immediate outcomes) will be expanded and further developed, and will be applied in additional countries over a period of six years (wider result) in order to improve the output and accountability of governments toward their citizens. This would result in improved services, a better use of M&E products by parliaments, and an improved performance culture.

The immediate results—demonstration of improved M&E systems—will be achieved by:

- The creation of an M&E application at the level of senior management, Parliament, and the public;
- The conduct of learning and sharing activities to build on, based on the experience of current partners;
- The development of specific M&E tools in collaboration with partners; and
- The implementation of programs to ensure ownership and cost-effectiveness.

Many countries undertake surveillance activities, but this program seeks to work specifically with those seeking to make a systematic assessment within government as well as monitoring, so that there is a desire for learning and deeper change, and more systemic effects on the output and impact of governments on the living conditions of their people.

Milestone and Stakeholders

The central objective of Twende Mbele is not to constitute a network where countries share their experiences a priori, but to help collaborate in the development and implementation of M&E systems that improve the performance and impact of government actions on citizens.

Established in 2015, the plan predicts that by 2018, the number of partner countries will increase from three to six. There is already high demand for countries to be included in the project, but participating countries must maintain a commitment level corresponding to their capacities. It is therefore important to define an engagement strategy to include other countries. For those countries that do not have a sufficient level of development of their national M&E systems, Twende Mbele has put in place an inclusion mechanism based on differentiations between countries. This option works primarily with countries that are truly committed to using M&E to improve the lives of their citizens.

An initial classification has placed countries in three broad categories:

- **Category A.** Those "knocking on the door" are committed to using M&E in change processes (the indirect indicator being that they are engaged in the evaluation)
- **Category B.** Those who are already initiating actions in the field of M&E, have individual champions, and are willing to become more involved
- **Category C.** Those with some interest but who are not actively pursuing M&E as a key element in improving government performance

The program wishes to actively target all of these countries, and include them to the extent possible in its activities. Their involvement and ability to advance the work will be evaluated using explicit criteria, and enabling them to become collaborative partners.

Potential partners in **Category A** are already very active in M&E, and are keen to improve what they are already doing for better government performance. In these countries, there is already the political will to expose failures and draw lessons from them. These partners could improve their activities by appropriating and contextualizing some of the more advanced tools that have been operationalized in countries outside of Africa. The most effective strategy to employ with these countries is to collaborate on the development of their national M&E systems in order to make them more effective and inclusive. This category includes Côte d'Ivoire, Ghana, and Niger. They can be considered main partners in the program.

The countries in **Category B** have some involvement in M&E, but have not yet committed to extensive systems, or faced rigorous evaluations that may reveal weaknesses. For those with significant interest in M&E, participation in learning and capacity-building activities can be beneficial. However, these countries need to demonstrate their interest in using these opportunities. This category includes Botswana, Ethiopia, Kenya, Lesotho, Rwanda, Senegal, Tanzania, and Zambia.

Category C countries may not be interested in being involved in Twende Mbele, but it may be appropriate to involve them in activities such as newsletters, and informing them of events they might be able to attend. The program should aim to spread M&E focal points across all African countries, and to share news about Twende Mbele as well as general M&E activities

and information that might stimulate further interest. This category includes Angola and the Democratic Republic of Congo.

It is essential that this process be based on self-selection, and that countries recognize where they stand in relation to Twende Mbele. They must also commit to the appropriate roles for each level of participation if they wish to participate.

LEARNING FROM THE BENIN EXPERIENCE

Becoming a member of 3ie has allowed Benin to benefit from financial and technical support for undertaking the impact evaluation of the Free Support for the Scholarship of Girls in Secondary School Level One Program. This program is designed to promote the schooling of young girls across the country and contribute to reducing the education gap with young boys.

The government of Benin is waiting for the evaluation findings in order to improve its sectoral reforms in the field of education. 3ie has also given financial support to many Beninese in the field of capacity building, to attend courses and seminars around the world that are linked to impact evaluation.

Difficulties Encountered

The major difficulty faced up to now in the area of cooperation to promote M&E—both at the country and subregional levels—is related to WACIE, which, while it is a country-level initiative, has program objectives that remain regional.

In fact, the promotion of WACIE requires WAEMU country members to belong to 3ie. Each WAEMU country member has the responsibility to fulfill this essential condition in order for the program to become a reality. Program implementation needs the strong engagement of all of its stakeholders, especially the WAEMU states, and communitarian institutions such as the WAEMU Commission and the West African Development Bank.

Perspectives

Benin's approach to its evaluation system has enabled the country to build its international reputation for the innovative cooperation programs it has developed with its partners. The ultimate expected result of the Twende Mbele program is the effectiveness of national M&E systems at the continental level: WACIE, which is a regional community program, offers promising prospects for the development of Impact evaluation at the regional level.

The sharing of experience and the development of collaborative governance tools encouraged by these programs will undoubtedly improve the quality of transparency in public management at the continental level.

After its pilot phase, WACIE will be carried out by a community institution, and will be a reference organization for WAEMU through the research results and evidence obtained. Through these results, WACIE will be able to feed the process of defining the community directives for WAEMU.

The movement that Benin will instigate through its inclusion in global evaluation networks will lead the countries involved to improve their quality

of governance: this could result in the improvement of their standings according to the Ibrahim Index of African Governance.[6]

CONCLUSION

Development policies and programs at the global level demand better performance and more efficiency not only from states but also from partners and donors. Evaluation is a mechanism used to report on the commitments made in order to ensure good governance: it is a strategic steering tool that can help to inform decision makers.

African states must therefore take ownership of evaluation, according to their sociocultural contexts, and they should pool their experiences in order to improve governance. This pooling calls for the establishment of regional and subregional cooperation platforms for the strengthening of national M&E systems, with the aim of improving the living conditions of the people.

Benin's experience in the field of public policy evaluation has shown that it is important to open up to other experiences in order to capitalize on achievements. It is within this framework that Benin has initiated the setting up of a regional program for capacity building and impact assessment in the WAEMU countries, in partnership with 3ie. This program will not only develop impact assessment within WAEMU countries, but will also strengthen national evaluation systems and the capacities of stakeholders to conduct the evaluation.

To achieve this, WACIE will establish a mechanism for the selection of evaluation areas, and analysis of the results and conclusions of evaluation reports. It will also lead consultations aimed at setting standards, guidelines, and supporting funding mechanisms.

Benin is also relying on the Twende Mbele program, which is the result of its cooperation with South Africa and Uganda, to strengthen the collective learning process at the subregional level through M&E performance improvement and evaluation practice. These cooperative regional initiatives in evaluation will open new perspectives for the development of evaluation in Africa.

REFERENCE

Mendiratta, Vibhuti. 2011. "Report on Impact Evaluation in Sub-Saharan Africa." Working Paper, UMR DIAL, Paris.

[6] The Ibrahim Index of African Governance (IIAG) provides an annual assessment of the quality of governance in every African country. Originally established in association with the John F. Kennedy School of Government at Harvard University, the IIAG now consists of more than 90 indicators in 14 subcategories, 4 categories, and 1 overall measurement of governance performance. These indicators include official data, expert assessments and citizen surveys, provided by more than 30 independent global data institutions. This represents the most comprehensive collection of data on African governance. http://mo.ibrahim.foundation/iiag/.

Part IV

Safeguards and Resettlement

Chapter 14

Environmental and Social Safeguards in India – A Critical Assessment

Shekhar Singh

Abstract. *This chapter describes the state of environmental and social safeguards in India, as applicable to development projects and activities. A theoretical framework and a brief historical background is provided to contextualize the contemporary situation. Though traditional human activities were for the most part environmentally sustainable, the development of technology and the growth of human population is putting increasing pressure on the natural environment. Skewed economic and social development, perhaps as a result of selective access to natural resources and technology, have also begun to show trends of inequity. This has resulted in human interventions that sometimes benefit the few at the cost of the many. In response, most countries have set up environmental and social safeguard regimes designed to assess the possible environmental and social impacts of human activities, to disallow those that are not viable, and to establish and monitor measures for minimizing and mitigating the adverse impacts of those judged to be viable. Unfortunately, in many countries—as, for example, in India—these measures have not been very effective because of vested interests both within and outside of the government, whose own objectives are better served by undermining, or rendering ineffective, all such safeguards.*

Shekhar Singh, Research, Assessment and Analysis Group, shekharsingh@gmail.com.

ENVIRONMENTAL SAFEGUARDS

Conceptual Framework

The proposition that most contemporary human activities disrupt the natural environment and its processes is widely accepted today. However, there is much dispute about which impacts are acceptable, and to what extent. The stress here is on *contemporary* human activities, as many argue that traditional rural and tribal societies lived in harmony with nature, and in some cases still do.

In India, two traditional groups that come to mind are the isolated tribes of the Jarawas and the Sentinelese, in the Andaman and Nicobar Islands. There is no evidence to believe that the presence of these groups has in any significant way degraded the ecosystem they inhabit. Apart from the fact that their numbers have been stable over many years, they reportedly have many rituals that ensure that they do not adversely affect their natural environment. One such ritual is the reported practice of hunting parties half-breaking a prominent branch in a prominent tree in the area where they have recently hunted a wild pig. This hanging branch serves as a warning to other hunters who might venture there, that a pig has been recently killed in the area, and therefore they should hunt elsewhere. In a few weeks, the half-broken branch dries up and falls to the ground, once again opening the area to other hunters.[1]

Such practices of the Jarawas—and presumably of the Sentinelese, about whom much less is known—ensure that their footprint on nature is kept to a minimum and does not have a permanent adverse impact. However, most other rural communities in India cannot rightfully claim that their survival strategies are in harmony with nature. The conversion from hunting-gathering to shifting or settled agricultural practices alone has transformed natural ecosystems all over India.

Whether historical natural processes are the best, or the only, way forward is now a somewhat moot philosophical question. The time when the answer to this question would have been relevant has long since passed. However, the limits of change and manipulation of the natural environment, and the consequences of getting them wrong, are still very relevant.

Environmental Safeguards and the Government

Governments have the unenviable task of determining how much use and disturbance of nature is permissible, and how to meet the basic needs and growing aspirations of their people without overstepping these boundaries.

[1] This story was told to me by Samir Acharya, founding president of the Society for Andaman and Nicobar Ecology (SANE) while I was holding hearings in Port Blair, as the Supreme Court of India appointed the commissioner for forests and related matters of the Andaman and Nicobar Islands (2000). Many such stories describing the conservation practices of tribal and indigenous people can be found in Bharucha (2016).

Most, perhaps all, governments have adopted policies whereby a certain proportion of the nation's area, representing various types of ecosystems, is conserved in its natural state. In India, these are the national parks, set up under the Wildlife Protection Act of 1972.

Other areas are classified such that only certain types of activities can be permitted there. In India these are identified as wildlife sanctuaries, reserved forests, conservation reserves, community reserves, notified ecologically sensitive areas and wetlands, and coastal zones, among other classifications: and they are protected under a host of laws and regulations. The proportion of area that a country protects in this way is mostly dependent on three factors: the richness and diversity of ecosystems and species found in the country, the demand for land and other natural resources for human use, and the way these are balanced against the political will of the government to conserve nature and to sustainably use natural resources. Unfortunately, most countries in the world seem to be struggling to get this balance right.

For the remaining areas, most countries have restrictions on the types of land or water use permitted and regulations concerning the extraction of resources, the destruction of natural habitats, and the release of effluents. These standards vary from country to country and from ecosystem to ecosystem, and are a function of the cost and availability of "green" technology; the levels of environmental awareness and activism among the populace; the commitment and ability of the government to ensure long-term sustainability of growth and development; and the inclination and ability of the nation to transfer its environmental costs onto others.

Evaluating Programs

In 1950, the government of India set up a Planning Commission modeled after the planning infrastructure in the then-USSR. As a part of the Planning Commission, a Programme Evaluation Organisation (PEO) was created to evaluate the various programs being undertaken by the government and supported by the Planning Commission. Over the last 60 years or so, many of the important programs of the government of India and the various state governments were evaluated by the PEO.[2] For many years, most of the evaluations focused on economic and social outcomes, and on cost and time efficiency. Gradually the scope of the evaluations expanded and new aspects were introduced, including environmental aspects. However, these evaluations were mostly ex post facto, or at best carried out midterm, and dealt with only a few specifically selected programs. They therefore were not adequate for assessing the social and environmental impacts of programs, projects, and activities in advance of their being initiated, nor for assessing their social and environmental viability. They did perform the important role of influencing the design and implementation of new and ongoing programs. Unfortunately, the Planning Commission, and along with it the PEO, were terminated in 2014.

[2] For details, see http://planningcommission.gov.in/reports/peoreport/index.php?repts=peobody.htm.

Regulating Use and Disturbance

Experience has shown that in matters related to the environment—because once damage is done, it might not be easily undone—it is not prudent to simply declare standards enforceable by law and hope that the deterrent effect of stringent penalties would adequately protect the environment. Therefore, most governments have adopted an environmental safeguards regime that requires projects and activities to be subject to prior assessment and clearance.

In India, the environmental safeguards regime was initiated in 1974 through an administrative order. In 1994, the requirement of prior environmental clearance for most projects was made legally binding under the Environmental Protection Act of 1986.

To appraise projects and recommend environmental clearance, various environmental appraisal committees (EACs) were set up at the national level under the Ministry of Environment and Forests,[3] separately for different types of projects. These EACs were chaired, and had as members, independent experts from outside the government. Officials from various of the concerned departments were ex officio members. Though the EACs still function, in 2006 powers were delegated to the state governments to appraise and grant clearance for certain categories of projects, essentially the smaller and less problematic ones.

The basic process of carrying out appraisal, granting clearances, and monitoring compliance essentially involves an environmental impact statement being prepared by an expert body hired by the project proponents for that purpose. The regulating ministry has guidelines concerning the preparation of these impact statements. The statement is then appraised by the appropriate EAC of the ministry.

The EACs recommend to the ministry whether a proposed project or activity should be given environmental clearance, with or without certain conditions, or if it should be rejected. These recommendations are based on an examination of the impact assessment statement; other relevant documents and information; and discussions with experts and concerned stakeholders.

For most types of projects, there is also a statutory requirement to hold public hearings involving interested and affected members of the public. In these hearings, the public is given an opportunity to express its views on the possible impacts of the proposed project; the suitability of the proposed preventive and mitigative measures; and the consequent viability of the project. EACs also sometimes carry out field visits to monitor and verify the situation on the ground.

Based on the recommendation of the EAC, the ministry issues a clearance, a conditional clearance, or a rejection. Legally, since the EAC is only an advisory committee, the ministry is not bound by its recommendations.

[3] The Ministry of Environment and Forests was renamed, in 2014, the Ministry of Environment, Forests and Climate Change (MoEF&CC). In order to avoid confusion, it is here consistently referred to as the environment ministry.

Apart from this environmental clearance, projects that have any liquid or gaseous effluents must also get clearances from the relevant pollution control boards. Where forestland is involved, either in the location or in the impact zone of the project or activity, a separate procedure for forest clearances is mandated, involving the Forest Advisory Committee. Where the project or activity is located in, or likely to impact, a wildlife protected area, or a protected species of fauna or flora, clearance is required from the National Board for Wildlife.

Once accorded, an environmental clearance can be suspended or revoked if the conditions for clearance are not complied with. Each project proponent is required to submit a report to the regional office of the environment ministry, under which it is located, every six months. The regional office has the responsibility of ensuring that the various conditions prescribed in the clearance are complied with. They are expected to do this based on these reports and on their own monitoring.

Major Challenges

On paper, India has a stringent and elaborate system of checks and balances with multiple authorities, professional bodies, committees, scientists and other professionals, and institutions, all of them identifying, appraising, assessing, and monitoring environmental impacts. However, internal contradictions within the government, and the machinations of external vested interests, have made this elaborate system ineffective, and often corrupt.

Internal contradictions within the government. The environmental safeguards regime, though initiated in the 1970s, was fully institutionalized only in the 1980s. At least in part, this institutionalization seemed to be the result of both direct and indirect international pressure, to which India had become susceptible.[4] There was also growing domestic media and judicial pressure, and a vocal environmental movement. Countering these pressures were domestic economic imperatives, the push for short-term gains that is the bane of a five-year election cycle, and the consequent demand for a rapid expansion of industrial and commercial activity, and of infrastructure. Growing human populations and aspirations created pressure to convert natural habitats into agricultural lands and human habitations.

India's political strategy relating to environmental safeguards seems to have evolved out of these opposing pressures. The 1980s saw the emergence of strong environmental policies and laws, and an expansion of environmental institutional structures. But it also saw the emergence of a plethora of strategies that effectively negated the effects of these strengthened laws and institutions and allowed "business as usual" to continue. It allowed the Indian government and political leaders, even while they were showcasing to the country and to the world the progressive safeguard measures they had put

[4] For a more detailed discussion on this point, see Singh (2011).

into position, to simultaneously escape the adverse political consequences of a slowdown in economic growth, albeit a temporary one.

The process of undermining the environmental safeguards regime seems to have been spearheaded by four distinct yet interrelated strategies. Initially, there was a tendency to bypass or ignore the newly established regulatory regime. This, however, led to extensive litigation relating to various projects in which litigants challenged the legality of the government, ignoring the regulatory agencies that they themselves had statutorily created.[5] A second, related strategy was to make sure that these regulatory agencies did the bidding of the government, and to refrain from setting up independent and objective regulatory agencies, despite orders from the Supreme Court of India to do so.[6] The third strategy was to make even these "controlled" regulatory agencies functionally ineffective by starving them of resources and personnel; and the fourth was to roll back the safeguards themselves.

Vested interests. Apart from internal contradictions confronting the Indian establishment, almost from the start there were various vested interests opposed to the proper implementation of environmental safeguards. At least four such interest groups emerged.

Perhaps the most benign of these were those who saw many of the environmental safeguards, especially those seen as imposed by Western nations, as unnecessary and unfair, and an impediment to the urgent need for providing shelter, livelihood, and food to millions of impoverished Indians. To them, natural resources had to be made available, on a priority basis, in order to meet the immediate survival needs of the poor, and not be diverted or earmarked for long-term conservation imperatives, many of which seemed to them to be based on principles that were unproven, or inappropriately applied to Indian conditions.

While acknowledging the primacy of the needs of the poor, conservationists argued that there were enough resources in the country to meet everyone's basic needs, while ensuring environmental sustainability. But to do this, the existing resources needed to be more equitably used and distributed. There was, according to them, no justification for compromising the future of the people of India, especially the poor, just because the government was not able, or willing, to redistribute resources, especially land, water, and forest resources, so that they could support the survival needs of the poor rather than the luxurious lifestyles of the rich.

The second, far less benign, interest group militating against environmental safeguards held that the safeguards inhibited national economic

[5] Perhaps the two best cases from that period are those against the proposed Tehri Dam and the Narmada project. For details about the Tehri Dam controversy, see the Supreme Court of India 2003 judgment on *ND Jayal and Shekhar Singh vs Union of India and others*, https://indiankanoon.org/doc/1875824/; and Warrier (2016). For details on the controversy surrounding the Narmada dams, see, e.g., Peterson (2010).

[6] *T.N. Godavarman Thirumulpad vs Union of India and Others*, 2011, https://indiankanoon.org/doc/1725193/.

growth and thereby prevented, or at least delayed, India's transformation into a world economic power. The fact that India is now among the fastest growing economies in the world has further reinforced this belief among many. This group ignored all concerns about the impact of an economy that was growing rapidly, but inequitably, on the poor and marginalized segments of the society. They also ignored the inevitability of a façade of rapid economic growth and expansion soon collapsing, if it was achieved in a manner that was not sustainable.

A third interest group that opposed the environmental safeguards regime, sometimes very aggressively, was comprised of the powerful lobby of Indian, foreign, and multinational corporations, who saw environmental restrictions as impediments to their growth and profitability. The efforts of the Indian government to attract foreign investment, recently spurred by the launch of the "make in India" campaign, has exacerbated this conflict. This interest group argued that the availability, in India, of cheap and plentiful skilled labor was not enough to attract foreign investment, and the deal needed to be "sweetened" with weakened environmental regulations.

The fourth, and perhaps the most pernicious, of the vested interests opposing the proper implementation of environmental safeguards are the rent seekers. Much money stands to be made, and is being made, by allowing the violation of environmental norms in exchange for hefty political "donations" and personal bribes. Many political parties, functionaries, bureaucrats, scientists, and other professionals, benefit from this system. Ironically, these interests are best served if there are, on paper, strong regulations and safeguards, but a systemic inability to ensure that the regulators do their jobs effectively and honestly.

The rent seekers also include public servants who are involved in granting contracts and clearing payments to builders and suppliers for government projects. These public servants seek, and often receive, pay-offs from the contractors who are hired to build the project, and from other suppliers. For this to happen, the projects have to be initiated and constructed, and therefore environmental and social safeguards have to be bypassed.

Safeguarding the interests of the "weaker." Apart from the above four vested interests, many countries around the world successfully transfer their own environmental costs onto other countries, both by dumping pollutants and by unsustainably exploiting their minerals and other natural resources. This represents another powerful vested interest that works against the safeguard regimes of victim countries.

The tendency to exploit the "weaker" by forcing them to absorb the environmental costs of the "stronger" does not occur only among countries, but also happens within countries. In India, the location of environmentally destructive activities (such as mines and dams), and of hazardous and polluting activities (such as chemical industries and coal-based power plants) is often influenced by the amount of economic and political clout held by the adversely affected communities. Certainly, the efficacy of the application of safeguards is profoundly influenced by the amount of political and economic power those likely to be adversely affected possess.

As a counterbalance to these interests and pressures, India is also host to strong environmental movements, a sympathetic media, and a supportive judiciary. Nevertheless, the combined interests that have rallied against the effective implementation of a strong regulatory regime seem to be winning, as described below.

Subversive Strategies

Bypassing or ignoring the regulatory agency. From the beginning, for reasons discussed above, the regulating ministries often came under pressure from other departments and ministries of the national government, from state governments, and even from the prime minister's office, to accelerate the process of environmental appraisal, and in some cases to grant undue environmental clearance to favored projects. In some cases, as will be discussed later, the concerned ministry succumbed to pressure. In others, they did not. In some of these latter cases, the central and state governments decided to ignore the regulating ministry and start work on the project before it had been granted environmental clearance and, in some cases, even before the environmental studies had been carried out.

These half-completed projects were then presented to the regulatory ministry as a fait accompli. The fact that much of the anticipated environmental damage had already occurred, and as such could not be prevented or minimized, even if the project had now been abandoned, strengthened the arguments in favor of granting it *ex post facto* clearance. The fact that a huge amount of public money had already been invested in the project created further moral and political pressure on the regulating ministry, despite the utter illegality and immorality of a project being initiated and half completed before the mandatory clearances were received.

In a few high-profile cases, the refusal of the regulatory ministry to grant clearance was overruled by the prime minister's office, and the regulatory ministry was directed to accord clearance. Perhaps the most famous example of this was the granting of environmental clearance, in 1987, to the Narmada Sagar and Sardar Sarovar dams, which were two of the largest dams on the Narmada River. Despite the environment ministry categorically stating that the projects were not yet ready for appraisal, let alone clearance, the prime minister's office overruled the ministry and directed that the projects be cleared, with a curious *pari passu* clause that mandated that studies and assessments be carried out concurrently with the construction. Following this logic, the projects would be ready for assessment only when they were fully constructed.

Many of the efforts to bypass or ignore the regulatory ministry were challenged in the courts of law and caused serious embarrassment to the government, and much adverse publicity. Perhaps because of this, there was a gradual shift to other strategies, as described below.

In 2013, there was a qualitative change in the efforts of the government to bypass the environment ministry. In January 2013, the government of India set up a Cabinet Committee on Investments (CCI) as a part of its

proposed National Investment Board.[7] The CCI was designed and empowered to intervene in instances where different approval processes, particularly those related to the environment, were thought to be impeding the economic growth of the country. The CCI had the power to review decisions taken by ministries in which projects had been refused approval, or there had been "undue" delays. It was also empowered to direct statutory authorities to discharge functions and exercise powers under the relevant laws and regulations within the prescribed time frames, for "promoting investment and economic growth."

This was widely seen as a strategy to gain political advantage in the forthcoming general elections of 2014. The mandate of this committee was essentially to bypass the environment ministry and other regulators, and to provide speedy, even almost automatic, clearances to proposed projects and activities that were pending with the ministry for more than three months, regardless of the fact that in many cases the required studies and assessments had not been completed and submitted by the project proponents. The CCI then proceeded to ensure environmental clearance to these projects without conducting any scientific appraisal, or even having access to any professional expertise (Press Information Bureau 2013). Going into the general elections of 2014, the Congress party claimed that it had granted environmental clearance to a large number of projects in the previous year.[8]

This was perhaps the most blatant and direct effort to bypass the environmental regulatory mechanisms and safeguards, obviously necessitated because the environment ministry was not fully compliant with the wishes of the government, despite being headed by a minister from the ruling party. It was also an unprecedented obfuscation of the responsibilities of various ministries and levels within the government.

Though there has been no other comparably blatant effort at bypassing the regulatory mechanism (and in effect dismantling it), the new government, which took office in 2014, has not shown greater concern for the environment than the previous one.

Compromising the scientific objectivity and integrity of the assessment process. Despite demands for an independent statutory body to appraise projects and activities, and to grant and monitor environmental clearances, this process continues to remain within the government. This is also despite the fact that in a ruling given in the case of Lafarge Umiam Mining Pvt. Ltd on July 6, 2011,[9] the Supreme Court of India emphasized the need for such an independent regulator. In another judgment, in the case of T.N. Godavarman Thirumulpad, the Supreme Court further reiterated that the central

[7] For details on the CCI, see http://cabsec.nic.in/writereaddata/cci/english/1_Upload_989.pdf.

[8] See, e.g., Sharma (2013).

[9] *T.N. Godavarman Thirumulpad vs Union of India and Others*, 2011, https://indiankanoon.org/doc/1725193/.

government was required to set up a regulator at the national level, which would have offices in all of the states; which could carry out an independent, objective, and transparent appraisal and approval of the projects for environmental clearances; and which could also monitor the implementation of the conditions laid down in the environmental clearances.[10]

The refusal to set up an independent regulatory mechanism was adversely commented upon by the Comptroller and Auditor General of India (CAG) in its report of 2016, which was prepared for submission to the President of India under Article 151 of the Constitution of India, to be presented to the Parliament:

> A National Regulator to oversee the entire process of grant of Environmental Clearance and monitoring is yet to be appointed despite directions of the Hon'ble Supreme Court. Environmental Clearances were granted to the Project Proponents without checking the compliance of the conditions mentioned in the previous Environmental Clearances and recommendations of the Regional Office. (CAG 2016, viii–ix)

Unfortunately, a high-level committee set up by the new government in 2014 recommended against the setting up of an independent authority for granting environmental clearances, citing the very reasons that had made such an independent authority desirable, as arguments against its creation.

> While all technical aspects of an application/proposal for clearance would be examined on merits by the NEMA, it was felt that the final approval or rejection powers should be retained by the MoEF&CC. This is because there may be many other factors, relating to relationship with neighbouring countries, need to address regional disparity issues, dealing with areas and regions with special problems and issues, and need to take national security issues into account etc. etc, which may singly or in combination add a further politico-economic-strategic dimension in the decision making process. (HLC 2014, 59)

Delegating powers to the state government. To make things worse, in 2006 a decision was taken by the government of India to delegate the power to grant environmental clearance for certain types of projects to the state governments.[11] This was a controversial decision for at least two reasons. First, there is a well-founded belief that state governments by and large are much less committed to implementing safeguards, especially environmental safeguards, than the central government is. It was this conviction that led the government of India, in 1980, to promulgate the Forest Conservation Act, which stipulates that no designated forestland can be diverted for non-forest use by the state government without prior clearance of the central

[10] *T.N. Godavarman Thirumulpad vs Union of India and Others*, 2014, http://courtnic.nic.in/supremecourt/temp/20219953612014p.txt.

[11] For details, see http://www.envfor.nic.in/legis/delegation.htm.

government. Statistics suggest that subsequent to the enforcement of this law, the amount of forest land being diverted drastically shrank.

There are many reasons for the seeming indifference of state governments to environmental damage. Usually the performance of the political parties that are in power in a state is judged by its ability to enhance jobs and incomes, to provide basic services, and to distribute "freebies" and concessions. Environmental conservation, primarily because of its long-term returns, is usually not a significant factor affecting the re-electability of the ruling political party.

Also, state governments usually function in a more unified manner, in which the head of the state, the chief minister, invariably exercises total power and control over all departments. There is little scope for environmental departments within a state to oppose or even delay and modify projects and activities that are politically important and that have the full support of the chief minister.

Evaluating the Performance of Government-Controlled Regulatory Authorities

Ignoring violations of the law. The regulating agency is mandated, under the Environmental (Protection) Act of 1986, to: "direct (a) the closure, prohibition or regulation of any industry, operation or process; or (b) stoppage or regulation of the supply of electricity or water or any other service" for any violation of the conditions of environmental clearance.

However, despite this, and despite there being numerous such violations, the regulating ministry has rarely taken action against projects and project proponents that were in violation of the conditions of clearance. The CAG, as part of its sample assessment, identified numerous violations in the two years under review:

> MoEF&CC had stipulated certain specific conditions in the EC either relating to sectors or to the project which were to be followed by PPs. It was observed that the monitoring agencies were not able to ensure compliance to the EC conditions. (CAG 2016, 69)

Furthermore:

> ...there was shortfall of 43 to 78 per cent (with reference to compliance reports of June 2015) in submission of half yearly compliance reports. Further, it was observed in audit that most of the PPs did not submit half yearly compliance reports timely and regularly and there was delay ranging from one month to 48 months in submission of the compliance reports. We noticed that the ROs did not issue reminders regularly for submission of compliance report to PPs. Also, no action was taken by the MoEF&CC against the PPs under the provisions of the Environment Protection Act, 1986 for non-submission of compliance report by PPs. (CAG 2016, 84)

The CAG went on to observe that despite numerous violations, no action was taken by the regulating ministry.

> In reply to a Parliament question, the Ministry submitted (July 2016) that no penalty was imposed by the MoEF&CC for violating conditions of EC in the last two years. We observed that MoEF&CC did not have a compiled database of cases/projects received by it from the ROs where the violations were reported by ROs after their monitoring/inspection. Data register with year wise breakup of such cases was also not maintained. (CAG 2016, 88)

Ignoring the recommendations of the EACs. A popular strategy to undermine the environmental safeguards regime that evolved in the 1980s was for the environment ministry to overrule the recommendations of the EAC. The fact that the EAC was only an advisory body allowed the ministry to adopt this strategy.

Ordinarily, given that the EAC is appointed by the environment ministry, the final decision should have been in conformity with the recommendations of the EAC. Where the ministry had additional technical inputs or findings that were contrary to those of the EAC, these should have been sent back to the EAC for consideration and comment. However, this was not done, and usually the ministry gave no reasons for rejecting or modifying the recommendations of the EAC.

Perhaps the most well-known of such cases was that of the Tehri Dam in the Himalayas. At 260.5 meters, the Tehri Dam is the highest dam in India, and among the highest in the world. Located in the Himalayas in what is known to be one of the most seismically active zones in the world (Category V), the EAC had unanimously determined, in 1989, that the environmental impacts and the safety concerns related to the project were such that it was not ecologically viable. Despite this, the environment ministry proceeded to grant environmental clearance to the project and gave no reasons why it chose to overrule the EAC.[12]

Another high-profile case was the first of the coal-based superthermal power stations in India, at Kayamkullam, Kerala. This power station was located adjacent to the ecologically fragile creeks of the coastal region of the state of Kerala. In 1991, the EAC rejected the location because of its ecological fragility, and suggested alternate locations that were ecologically less sensitive, and economically and logistically preferable. However, allegedly because the initial site was within the political constituency of a powerful political leader, the environment ministry overruled the EAC and cleared the project, without giving any reasons.

[12] For further details, see the Supreme Court of India 2003 judgment on *ND Jayal and Shekhar Singh vs Union of India and others*, https://indiankanoon.org/doc/1875824/. Also see Narrain (2003), IRN (2002), and IUCN, https://portals.iucn.org/library/sites/library/files/resrecfiles/GA_18_REC_057_Tehri_Dam_Project_India.pdf.

In a similar case, a proposed coal-based thermal power station located adjacent to a crocodile sanctuary in Dholpur, in the Indian state of Rajasthan, was rejected by the EAC in 1992, but cleared by the ministry, again without giving any reasons. In this case also, the EAC recommended shifting the location to a less ecologically fragile area, but the suggestion was rejected, allegedly because the original location was within the political constituency of the then chief minister of the state.[13]

Fortunately, in all these cases the triumvirate of people's movements, a sympathetic media, and a supportive judiciary, helped. A case filed in the Supreme Court of India ensured that the environmental safeguards related to the Tehri project were strengthened.[14] Unfortunately, the Supreme Court declined to take a view on the safety concerns, indicating, perhaps correctly, that this was less a legal issue than a technical one, for which they did not have the requisite expertise.

In both Kayamkullam and Dholpur, public and media pressure, and the threat of legal action, resulted in the projects being converted from being coal-based to naphtha-based and gas-based respectively, thereby reducing the adverse environmental impact on their surroundings.

Undermining the independence of the EACs. The EACs are functionally dominated by the chairperson, who is responsible for making all final decisions after considering the views and advice of the members of the committee, and of invited experts. Decisions in the EAC are not taken in a democratic manner, in which each member has a vote. This is in keeping with how most official committees function, with decisions made mostly by the senior-most functionary, and with other members operating more as advisers than as co–decision makers. Therefore, it is critical to ensure that the chairperson of an EAC is competent, independent, and of impeccable integrity.

The experience with EACs during the 1980s and early 1990s taught the environment ministry that overruling the EACs would attract much public and media criticism, and would give opponents a good legal basis to move the courts. Therefore, it quickly revised its strategy and started replacing the independent experts who had initially chaired the EACs, with retired civil servants, or others who were either sympathetic to the concerns of the project lobbies, or were pliable and could be pressured.

Compromising the independence of environmental consultants. The EACs were primarily dependent on the environmental impact statements provided to them by the project proponents. As these statements were prepared

[13] The Kayamkulam and Dholpur projects were appraised in the early 1990s, before the web became functional in India. Therefore, documentation regarding these and other such projects is not available on the Internet. However, the author was the chairperson of the EAC that appraised both these projects and has a copy of all relevant documentation. A relatively recent publication that describes many other such cases is Chainani (2007).

[14] For details, see Narrain (2003).

by consultants who were hired by the project proponents, there was always an inherent danger of conflicts of interest.

This situation was aggravated by the fact that the EAC had neither the resources nor the mandate to carry out fresh assessments, or even to empirically test some of the claims made in the environmental impact statements. At best, it could visit the site of the proposed project, make observations, and require additional studies to be done, or studies to be done again. However, usually these studies would be carried out by the same consultants. Occasionally there was a possibility of getting independent studies done, but only in high-profile projects.

The necessity of introducing a system in which the initial environment impact assessment could be carried out by a competent professional body that was independent of the project proponent, was stressed from time to time.[15] It was suggested that a panel of consultants and professional institutions could be maintained by the environment ministry, or by the Planning Commission, which could commission them for the task and pay them from funds recovered from the project proponent. Unfortunately, these recommendations have never been accepted, and no reasons have been given for the failure to accept them.

Compromising the functional efficacy of the regulatory agency. Most projects and activities were granted conditional environmental clearance, in which the clearance was based on adherence to certain conditions, especially preventive or mitigative strategies. There were also various statutory standards that such activities and projects had to comply with.

The responsibility of monitoring these projects to ensure that they were complying with the conditions of clearance was assigned to the 10 regional offices of the regulating ministry. Unfortunately, these offices were very inadequately staffed, and continue to be so. As a result, there is hardly any monitoring of compliance of the conditions of clearance. According to the CAG:

> There were only 15 scientists available for monitoring of Environmental Clearance conditions against sanctioned strength of 41. Regional Offices have not been delegated the powers to take action against the defaulting PPs and they had to report the violations of the Environmental Clearance conditions to the Ministry. (CAG 2016, 85)
>
> 24 State Pollution Control Boards/Union Territory Pollution Control Committees did not have in place sufficient infrastructure and manpower for monitoring despite having sufficient funds. (CAG 2016, 94)
>
> As per the information provided by MoEF&CC and its ROs, a total 9,878 Category A projects and 12,657 Category B projects were to be monitored by the ROs which had been given ECs since the inception of the EIA process, following the notification of 1994. (CAG 2016, 85).

[15] See, e.g., Singh and Banerji (2000) and Planning Commission (2007).

> As per MoEF&CC norms (July 2015) each scientist was to monitor at least five projects per month. Therefore, minimum 60 projects were to be monitored every year by each scientist…it may be seen that MoEF&CC/ ROs would not be able to monitor all projects under their jurisdiction even in a period of five years. (CAG 2016, 86-87)

Diluting standards. As described earlier, the initial strategy seemed to be to bypass or ignore the regulatory regime. This was followed by an effort to make the regulatory mechanism and the safeguards subservient to the whims of the government, and without any functional and scientific independence. The safeguards regime was also progressively made increasingly ineffective, so that it did not even have the ability to perform the required functions.

Essentially the dilution of the safeguards regime is being achieved by lowering the standards required; shortening the time available for conducting impact studies and assessments; and redefining the parameters that determine which projects qualify for prior assessment, and to what level, thereby excluding an increasing number of projects.

Commenting on the environmental impact assessment (EIA) notification and the amendments issued by the environment ministry, a joint committee of experts from the various Indian Institutes of Technology observed:

> In exercise of the powers conferred by the Environmental Protection Act, 1986 (GoI, 1986) Government of India (GoI) on 27th January 1994 made it mandatory for expansion and modernization of existing projects to have prior environmental clearance (EC) (MoEF, 1994). Thirteen amendments were made to it during 1994 to 2005…and then, in 2006 principle notification was replaced with a new one. The initial notification is no longer in effect, but it is our opinion that in comparison with the principle notification, the new one is weak in some of the areas, at least. (IIT 2011, 15)

With the installation of the new government in 2014, there now seems to be an added focus on the fourth strategy, that of dilution of the safeguards themselves, and of the processes involved in implementing them.

Soon after taking charge, the new government set up at least two committees to examine the ways and means by which environmental regulations could be "rationalized." The first of these was constituted in August 2014, and submitted its report in November 2014 (HLC 2014).

This committee recommended, among other things:

> …the identification of "no go" areas, which are in forest areas or inviolate zones—primarily with the criteria of over 70% canopy cover and "Protected Areas" which should not be disturbed except in exceptional circumstances, and that too only with the prior approval of the Union Cabinet. (HLC 2014, 11)

The disastrous implications of this recommendation can be judged by the fact that only about 3 percent of India's forests have canopy cover of over

70 percent (Forest Survey of India 2015). At present, all the legal forest area plus other areas that have tree cover (a total of nearly 30 percent) have legal restrictions on their diversion for nonforest uses. If the recommendations of the high-level committee are accepted, most of India's forested area, which in any case is well below the required 33 percent, would be opened up to industrial and other nonforestry uses.

Under the current regulatory regime, where forestland is allowed to be diverted for nonforest use, an equivalent area of nonforest land has to be brought under forest cover. In exceptional cases, compensatory afforestation can be permitted on degraded forestland where appropriate nonforest land is not available, and the overall forest cover of that state is at least 33 percent (the prescribed national minimum). This condition has ensured that the overall extent of forestland that either has tree cover, or has the legal protection that would allow regeneration of tree cover, does not decrease in the country.

Unfortunately, the high-level committee has recommended that this clause be dropped:

> The Committee recommends that this condition that there must be at least 33% forest cover in a State before approval is given for CA on degraded forest land should be done away with. (HLC 2014, 36)

The high-level committee has also recommended that:

> All the strategic border projects (border roads, fencing, Border Out Posts, floodlighting, surveillance infrastructure, power infrastructure) falling within 20 km. from the International Border, Line of Actual Control, Line of Control; and the projects in power sector and coal mining which are the growth engines for national economy may be given a fast-track treatment through special procedures. (HLC 2014, 57)

SOCIAL SAFEGUARDS

Unlike with environmental safeguards, until recently there were no social safeguards that were statutorily required for development projects and activities. For most large projects where human populations were being physically displaced, there was invariably a scheme or policy to manage the displacement and to minimize adverse consequences on the affected population. By and large, the focus of social safeguards was limited to the physical displacement of families and individuals.

When, in the mid-1970s, environmental appraisal processes were set into motion at the national level, along with various environmental parameters, human displacement was also mentioned. Therefore, while seeking environment clearance, projects also had to describe any human displacement that would take place, and lay out plans for rehabilitation. This became a precondition for getting environmental clearance, even though technically the regulation of human displacement did not come under the purview of the environment ministry.

It was only in 2007 that the government of India finally came out with the National Rehabilitation and Resettlement Policy.[16] It took another six years for this policy to get a corresponding statute. In 2013 the Parliament finally enacted The Right to Fair Compensation and Transparency in Land Acquisition, Rehabilitation and Resettlement Act (known as the R&R law).[17] This law, though not very strong, does provide a statutory basis for regulating the adverse social impacts of the acquisition of land for development purposes. The overall responsibility for enforcing this act lies with the national Ministry of Social Welfare.

Perhaps any R&R law can be assessed on the basis of at least four tests:

- Does it discourage forced displacement?
- Does it comprehensively define affected families/displaced persons?
- Does it provide for a just and humane compensation package and process?
- Does it provide for effective implementation?

Discouraging Forced Displacement

India's R&R law stipulates that forced displacement can only be done when it is in the public interest. It defines the public interest as including security concerns, infrastructure projects, the resettlement of project-affected persons, housing for specified disadvantaged groups, and the resettling of disaster-affected populations. It further stipulates that the social costs should be justified based on a prior social impact assessment (SIA). However, it does not establish any norms to guide or regulate the conduct of an SIA, and it exempts irrigation projects where an EIA is being conducted from also conducting an SIA.

The law bans the acquisition of multicropped irrigated lands, except as a last resort, though it exempts linear projects from this prohibition. It also stipulates that acquisition must be for the least displacing alternative, and of the minimum required area. Private companies can acquire land only if at least 80 percent of the affected families consent to it. The law does not make it mandatory to do either an accumulative impact assessment in an area, or on a community or an SIA of the overall development model and its components.

[16] Copy available at http://www.dolr.nic.in/NRRP2007.pdf. For a critique of the draft policy, see Singh (2006).

[17] Copy accessible at http://lawmin.nic.in/ld/P-ACT/2013/The%20Right%20to%20Fair%20Compensation%20and%20Transparency%20in%20Land%20Acquisition,%20Rehabilitation%20and%20Resettlement%20Act,%202013.pdf.

Defining Affected Families and Displaced Persons

The law defines an affected family as one whose land or other immovable property has been acquired. Members of scheduled tribes and other forest dwellers who are losing forest rights are also classified as affected families. It includes as displaced persons those residing in the area being acquired even though they might not own any land or property, and those whose primary source of livelihood will be affected. It includes the landless, tenants, share-croppers, artisans, agricultural laborers, usufruct rights holders, gatherers of forest products, fishers, hunters, and boatmen and women–provided they have been involved in these activities for at least three years prior to the acquisition.

Adult unmarried daughters and sons, widows, divorcees, and women deserted by their families who are residing in the affected area, are considered separate families. The law includes dependent minor sisters and brothers in its definition of family.

Providing a Just and Humane Compensation Package

Though various types of compensation are provided under the law, the major problem is that the law does not mandate that land must be given in exchange for land. This means that when poor farmers are displaced, they are not provided with other land where they can again take up farming. Though they are financially compensated, the expectation, if indeed there is any, that they could then use this money to buy equivalent or an even greater amount of land of equal or better quality, is not well founded.

Bitter and long experience has shown that land prices shoot up in areas where there is a sudden demand for land from displaced farmers, making it impossible for them to replace the land that they have lost, let alone improve on it. Also, most poor farmers have no experience of handling large sums of money, and are either cheated out of it, or spend it on immediate needs and wants rather than saving it to replace their productive assets. This leaves the farmers with no option but to go into some other profession, for which they are not trained, and are often not suited.

Effective Implementation

As with environmental safeguards, there are powerful interests opposed to the establishment and effective implementation of a progressive R&R regime. These include, in the main, the corporate lobby that sees its profits being eaten away when huge expenses have to be made to provide relief and rehabilitation for displaced populations. It also includes ministries and departments of the government, especially those charged with infrastructure development, who find it difficult to justify the overall economic benefits of the project (the cost/benefit ratio), if the costs of relief and rehabilitation are high.

There is also often unresolved tension within host communities, who are forced to share their resources with resettled populations. This is aggravated when populations are relocated in distant, or culturally antagonistic, locations. All of these factors have combined to inhibit the proper design and implementation of an effective social safeguard regime in India.

The R&R law does not, unfortunately, envisage an independent and statutory appellate body and monitoring authority to ensure that the process of rehabilitation is fairly and properly executed. This responsibility lies with the government which in most cases is neither willing nor able to carry out this function.

Though the law was enacted in 2013, the almost identical policy statement has been in force since 2007. Unfortunately, initial assessments reveal poor implementation. This can be seen from the observations of the CAG, which has surmised that "in over 80% of the projects sampled, the R&R conditions required to be followed were not specified in the environment clearance, despite there being a statutory requirement to do so" (CAG 2016, 60).

Unfortunately, the R&R law does not make the provisions of the law binding, as a fundamental right under the Constitution, nor does it make individual entitlements of project-affected people legally binding through contracts. Nor are officials charged with the responsibility of implementing provisions of the law made personally liable for any violations. The R&R law is somewhat unique among laws Indian laws, in the sense that it mandates no punishment or penalty for any functionary involved in infractions of the law: in fact, the only penalties envisaged are for members of the public who might knowingly supply false information to the authorities,

FUTURE DIRECTIONS

One lesson that emerges from the experiences described in this chapter is that for environmental and social safeguards to be effectively implemented, there is a critical need for regulators who are functionally, administratively, and financially independent of the government.

The experience of the past 40 years or so has also demonstrated that unless there is constant pressure from people's groups and movements, supported by a sympathetic media and a sensitive judiciary, the executive on its own is unlikely to pay much attention to either of these two sets of safeguards.[18]

It is also critical, in order for both the independent regulators and for people's movements to have increased credibility and impact, that there be periodic independent assessments by constitutional and statutory authorities in the assessment of the CAG. Assessments by independent scientific institutions, and by people's organizations would also be invaluable, so that the findings of all of these can be linked back to the initial appraisals of both ongoing and completed programs and activities, and can also be used to ensure that future ones are better designed and implemented.

[18] For a detailed set of recommendations relating to implementation of environmental safeguards, for which there is now experience of over 40 years, see Planning Commission (2007), 7–12. Though somewhat dated, most of the recommendations therein are still relevant today.

ACKNOWLEDGMENTS

I am grateful for inputs from Raman Mehta, Vishaish Uppal, and Misha Bordoloi Singh, while writing this paper.

REFERENCES

Bharucha, Erach. 2016. *Living Bridges: Folk Cultures of India, Then and Now.* HarperCollins India.

CAG (Comptroller and Auditor General of India). 2016. "Report of the Comptroller and Auditor General of India on Environmental Clearance and Post Clearance Monitoring." Report No. 39 of 2016 (Performance Audit), Ministry of Environment, Forest and Climate Change.

Chainani, Shyam. 2007. *Heritage and Environment: An Indian Diary.* Mumbai: Urban Design Research Institute.

Forest Survey of India. 2015. "Forest Cover. India State of Forest Report 2015." Ministry of Environment and Forests.

HLC (High-Level Committee). 2014. "Report of the High Level Committee to Review Various Acts Administered by Ministry of Environment, Forest and Climate Change." Government of India.

IIT (Indian Institutes of Technology). 2011. "Mapping of Policy Instruments and Governance Agencies for Environmental Clearance of Hydropower Projects in Upper Ganga Segment." Report Code: 008_GBP_IIT_PLG_ANL_02_Ver1_Dec2011. Indian Institutes of Technology, Kanpur.

IRN (International Rivers Network). 2002. "Tehri Dam." Fact Sheet.

Narrain, Siddharth. 2003. "The Judicial Go-Ahead." *Frontline*, October 10.

Peterson, M.J. 2010. "Narmada Dams Controversy—Case Summary." Case Study, International Dimensions of Ethics Education in Science and Engineering, University of Massachusetts Amherst.

Planning Commission. 2007. "Reports of the Task Forces on Governance, Transparency, Participation and Environmental Impact Assessment, and Urban Environmental Issues in the Environment and Forests Sector for the Eleventh Five Year Plan (2007–2012)." Government of India.

Press Information Bureau. 2013. "Summary of Decisions Taken by the Cabinet Committee on Investment since 30th January 2013." India Cabinet Committee on Investment, New Delhi.

Sharma, Nidhi. 2013. "Cabinet Committee on Investment Faces Clearance Hurdles." *Economic Times*, April 2.

Singh, Shekhar. 2006. "Displacement and Rehabilitation: A Comparison of Two Policy Drafts." *Economic and Political Weekly* 41 (52).

———. 2011. "Evolution and Implementation of Environmental Policy in India: Performance of the Congress Led National Governments in India: 1947–1984." In *A Centenary History of the Indian National Congress—Volume V: 1964-1984*, Pranab Mukherjee and Aditya Mukherjee, eds. New Delhi: Academic Foundation.

Singh, Shekhar, and Pranab Banerji, eds. 2000. "Large Dams in India." Prepared for World Commission on Dams. Indian Institute of Public Administration, New Delhi.

Warrier, S. Gopikrishna. 2016. "The Dam Debate." *Frontline*, February 5.

Chapter 15

The Imperfect Use of the Past in Resettlement

Inga-Lill Aronsson

Abstract. *This chapter discusses the use of the past in the implementation, knowledge production, and evaluation of resettlement projects. It argues that heritage and memories are neglected resources, and necessary analytical elements of the sociocultural dimensions of resettled societies. Sociocultural dimensions are the tangible and intangible resources that constitute everyday routine culture, supported and molded by the social relations, memories, heritage, and emotions that are attached to the landscape and environment. These dimensions are the least studied and the least understood in resettlement. It is further argued that the present resettlement models are insufficient to grasp the longitudinal consequences of resettlement. A consideration of heritage and memory would improve the model. The use of the past from a longitudinal perspective is explored through the ethnography of the Zimapán resettlement project.*

"Every society is a battlefield between its own past and its future"
—Eric Wolf (1959, 106)

Inga-Lill Aronsson, Uppsala University, inga-lill.aronsson@abm.uu.se.

The focus of the special session on resettlement at the International Development Evaluation Association (IDEAS) Bangkok conference in 2015 was to "look back to shape the future." This is both a theoretical and a practical challenge, as has been well documented, and has been expressed as "lessons learned" in resettlement literature and policy reports. Resettlement is not the only field that struggles with how best to accept the past. For peace work, it is of pivotal importance to recognize the past, as it is expressed in collective memories and traumas, in the process of reconciliation. Cultural heritage is another field whose theoretical focus and professional praxis is the past. The idea of a past to be used for present and future needs is also captured in the expression "learning from history." Its lack of success is well known, and these endeavors have failed to implement the old advice of "do no harm" reintroduced in this context by Anderson (1999). In development forced displacement and resettlement (DFDR), the difficulties are visible in the implementation, where dissonance prevails between the stakeholders.

As demonstrated from the 1950s on, DFDR projects are difficult to design, implement, assess, monitor, and evaluate, as the literature convincingly shows. Most recently, Smyth et al. (2015) have once again brought this to our attention, in their article outlining five "´big´ issues" for livelihood restoration that echoes requests of an urgent need to mitigate the negative consequences of resettlement. At the IDEAS conference, this ongoing problem was channeled into the resettlement evaluation thematic group led by Susan Tamondong.

This chapter addresses the special challenges attributed to the sociocultural dimensions of resettlement, which are "soft," "fuzzy," and difficult to discern, identify, and handle during implementation, using the present project models. They are related to economic recovery, but how and why they are is less known. Foremost, they are interconnected with the past routine cultures of society that the project is attempting to rebuild as a joint enterprise. International specialists have comparative and wider knowledge of DFDR, while local experts have in-depth knowledge about their life-world. These are the realms that knowledge production and the ensuing evaluation consist of.

Sociocultural dimensions are meant to include the tangible and intangible resources that constitute people's everyday, routine culture, which is supported and molded by the social relations, memories, heritage, and emotions that are attached to the landscape and the built environment. These resources are the peoples´ livelihood and life-worlds, and they are governed by "heritage as life-values" (Josefsson and Aronsson 2016) and "heritage as ambivalent" (Aronsson 2013). They are conveyed and framed by both spatial and temporal orders. They are both material and immaterial—they are concrete, visible, and durable as well as fuzzy and subtle, but they can be observed, recognized, and studied. There is nothing mystical, metaphysical, or esoteric about them. When a society is displaced and resettled, it "falls apart from within" (Aronsson 2002), and these are the orders that have to be reconstructed and reconstituted. These sociocultural dimensions are not static and cannot be frozen in time to easily fit a compensation matrix or an evaluation scheme.

THE PAST IN RESETTLEMENT

To explore the use of the past in resettlement, I will share my deep ethnographic knowledge of the Zimapán resettlement project in Mexico, which was executed 20 years ago. I have lived through this resettlement together with the people in the valley, whose landscape and society was inundated by the Zimapán hydroelectric dam in 1994. I returned there in 2013, and have had regular contact with the people. I have never been part of any executive power structure that has shaped my data. My perspective is longitudinal, with an emphasis on the reconstruction of society, and I explore the topic from an interdisciplinary perspective. The ethnographic context is pivotal, and I have tried to understand how and why certain choices and decisions were made, using the lens of heritage, collective memories, and collective traumas. This has triggered questions about what kind of knowledge was produced during the negotiations between the main stakeholders, and what bearing this knowledge had on the longitudinal results of the resettlement project.

Complex sociocultural data are the building blocks of society, and they frame knowledge production as well as the rebuilding of the resettled society. These building blocks are fluid, intangible, and embedded in a material world. How can we ethnographically describe these sociocultural building blocks, which are embedded in the cultural heritage of a resettled society undergoing rapid, induced change? What are the challenges for evaluation? These are the questions that are discussed here. The combination of theoretical concepts I have drawn on emanates from my background in anthropology and resettlement, humanitarian action, and cultural heritage.

THE ZIMAPÁN SCENARIO

The villages of La Vega, Vista Hermosa, and Rancho Nuevo in the Ejido Vista Hermosa, in the state of Querétaro in Mexico, were involuntarily resettled in the 1990s because of the building of the Zimapán hydroelectric dam. The reservoir is located on the border of Querétaro and Hidalgo: the dam wall is 203 meters high and 80 meters wide, and was built in a 400-meter deep canyon where the San Juan and Tula rivers join the Moctezuma. The reservoir covers 22 square kilometers and has two arms, each 12 kilometers long. One stretches up the Tula River and the other follows the San Juan River. The water level is calculated to be 180 meters at the highest point, and the dam was estimated to run at full capacity in 1998 in order to pay off the investments. However, this was not the case, according to local informants in July 2013, because of a lack of rain. They reported that they could see rain on the other side of the mountain range as usual, but when the clouds approached the reservoir side, they dissolved. During my fieldwork, I noticed a gap of an estimated 20 meters between the water surface and the indicated full-water level of the reservoir.

Traditionally, the people in the valley lived between two spheres of cultural heritage. Hidalgo is known for its Otomi heritage, but on the Querétaro side, the picture is less clear. The identity of the valley people is a mixture of Otomi, Spanish, and *mestizo* cultural heritage, but also with a claim to

Chichimecas heritage. The kinship ties to Hidalgo were extensive. Spanish was spoken, but with Otomi and Nahuatl words used for everyday, routine culture. The elderly generation spoke Otomi at home, but seldom in public. In 1994, I asked the villagers about their identity, and they said that they could not be Otomi because "we have forgotten how to speak this language" (Aronsson 2002). Instead, they identified themselves as "mountaineers" and "*ejidatarios*."[1]

Their livelihood was based on agriculture, combined with seasonal migration work in the United States. Religiously, they were divided: but Catholics and Seventh Day Adventists lived side by side, and were roughly distributed between Rancho Nuevo (Adventists), Vista Hermosa (mixed), and La Vega (Catholic).

The number of resettled people varies from 2,152 (in 1991) to 2,452 (in 1996).[2] Counting people in a resettlement project is difficult, because no community is spatially closed: family constitutions and homesteads are only stable at certain points in time.

In Zimapán, the gates of the dam were closed on November 27, 1993, at five o'clock in the morning, without prior notice. This was an emergency measure no one had wanted. The people had refused to move, and after countless negotiations, there was no other recourse than to close the gates and let the water fill the valley. Some families still refused to go: the water rose, and there were no roads, no drinking water, and no electricity. The animals were dying or fleeing. The remaining families were forced to move when the water level rose. A lifestyle had come to an end, and something new had to begin.

BELLA VISTA DEL RÍO

The valley people chose to move to the nearby semidesert plateau, Mesa de León, within the boundaries of the *ejido*, where they still had some rain-fed land. The new village of Bella Vista del Río consists of the three former villages, with each village demarcated, and clearly divided by wide concrete avenues. When the new village was built, it became a hybrid urban enclave within a rural environment. There was a strong economic contrast between the receiving homesteads and the new village.

Twenty years later, in 2013, the border between the new village and the plateau had become blurred. The resettlement had transformed the settlements and the previously simple and nonpermanent homesteads on the plateau. The old houses had been enlarged, and new houses had been added along the main road, and new houses encroached on the plateau´s remaining *ejido* land. The spatial order of the plateau had thus changed, and was consolidated by the material manifestation of permanent houses. In addition,

[1] An *ejidatario* is a member of a collective agricultural community established after the Mexican revolutions.

[2] Data are from unpublished reports by the Comisión Federal de Electricidad (CFE) in 1991 and 1996.

new enterprises had been established. There was a hotel, restaurants, car mechanics, hardware stores, information technology shops, butchers, mobile food trucks, and typical small *tiendas*, and the old church had been improved. All of this indicates an economic upswing.

In the new village, which is still encircled by a high fence, I was told by the resettled people that the people on the plateau had gained, at their expense. To understand this statement one must know the history of the place. Before the resettlement, the valley had functioned as a node in the region: and the people who lived on the plateau were landless squatters who had arrived, one family after another. The *ejidatarios* had allowed them to build nonpermanent houses, each one surrounded by a small garden, and as long as they themselves did not need the land, they allowed the landless occupants to have small herds of goats. There was no water on the plateau: therefore, the squatters were allowed to come down to the valley to fetch water once a week, to wash their cars and clothes, and swim in the river (which "belonged" to the valley people). They also received or exchanged products such as prickly pears, for the fruit and vegetables that grew in the valley, with its three to four harvests a year of corn, beans, tomatoes, and thousands of fruit trees. The valley people had a high socioeconomic status, which was directly linked to the richness of the valley.

The resettlement changed all of this: the former valley people lost status and reputation, while the former squatters on the plateau gained status, and became more self-assured. Tough negotiations followed between the valley people and the former squatters, who did not want to leave their houses and lots. They came to an agreement, and the people stayed. By 2013, some families from the new village had also moved to the plateau.

The decision to place the new village on the semidesert plateau was a joint one, decided with a majority vote. The villagers had the option of moving down to the town of Ezequiel Montes, which was closer to the planned restitution of their inundated farmland, but that option was rejected for several reasons. The main argument was that they wanted to keep their villages, with its social matrix of landless people and landowners, intact: this is in line with anthropological theories on social coherence and solidarity. Later on, the landowners rejected the restitution farmland, and instead accepted large cash compensation (Aronsson 2002). The Comisión Federal de Electricidad resettlement team had argued against cash compensation, but in vain. A few years later, many of the families had lost their money due to unwise investment and consumption.

The village soon became known as the "women's village" because the men had left for the United States to find work. The migration cycle had thereby changed from that of the 1990s, when the men had migrated in harmony with the agricultural cycles. After the resettlement, they stayed in the United States much longer, and finally they did not return at all. By 2013, more than 1,000 people had left, including entire families and single women.

The name Bella Vista del Río means "beautiful view of the river," but there are no signs of the former river. Names, places, and landscapes in Mexico (and elsewhere) are usually coherent: they support narratives and function as social memories. In this case, however, there is a cognitive

dissonance between the name and its environment. My interpretation of this is that the only thing left for the people now was a cognitive category of a beloved landscape, which was etched into their minds and bodies. They felt they needed to bring this with them, but if so, it was an unconscious process on the deepest level of collective memories. I have formulated this earlier as:

> Consequently, naming was a strategy aimed at reconstructing and upholding a socio-cultural continuity. People brought with them the mindscape of the valley, loaded with emotion, and transformed it into names. This may also have been the beginning of a process of enculturation into the new life in the new village… (Aronsson 2002)

Further research is needed to see how this process of enculturation is formulated and expressed. In 2016, Bella Vista del Río can be followed on YouTube, Facebook, Twitter, Instagram, and other social media, and is well integrated into the global communication network. It is mainly the younger generation that is connected. The Google street-view camera car team has also visited the new village. As my task here is to analyze a cultural transformation, this could be highlighted as something extraordinary, but I am hesitant to do so. It would be a form of "exoticizing" the village based on an outsider´s stereotypical view of rural peasant society. The Internet and social media would inevitably have found their way down to the valley with time. It is the speed and the profoundness of the changes that creates a dissonance between the generations and between families that stands out. Any resettled society will be exposed to this kind of dissonance, and there is less need to problematize modernity as such than to figure out how to get the pieces to fit together: because in 2013, there was a general feeling that the new village was in a state of disharmony.

THE PLACE OF HERITAGE IN RESETTLEMENT

Heritage has never been applied to resettlement, except in its narrowest sense of archaeological sites. A working definition of heritage is that heritage is about using the past as a resource for present needs. Consequently, a selection of the past is singled out, elevated, and labeled "heritage" in an institutionalized setting. This selection also includes "difficult" or "dark" heritage that has been translated into popular tourist sites.[3] Within this field of "difficult" heritage, there are various types, such as dissonant, unwanted, and "uninherited" heritage. The terms dissonant and unwanted heritage refer to contested heritage. Uninherited heritage is heritage that exists but that does not seem to have any value. A dissonant heritage carries the burden of history, the mistakes and atrocities that at any time can burst open again, and cause open conflict: it is always present beneath a calm surface. Both

[3] Numerous sites of dark and difficult heritage have become tourist hot spots, e.g., Dachau concentration camp in Germany, Terror Haza in Budapest, and the site of the destroyed Buddha statues in Afghanistan.

low-intensive and long-lasting conflicts are examples of this. I have discussed the heritage concept elsewhere and have accepted its ambivalence, but also its omnipresence (Aronsson 2013; Josefsson and Aronsson 2016). Hence, heritage is grounded in both space and time: and this is fundamental to our understanding of the resettlement process.

The power of heritage is that it connects us to the past and makes us believe that there are existential, "God-given" values that can help us return to and restore lost values and collective memories from the past. These urges to reconcile with the past and make it comprehensible are so strong that people who have been forced to migrate due to armed conflicts try to recreate their past, sometimes in a way that denies reality. An example of this is Palestinian families living in refugee camps in Lebanon, who have kept the house keys to their long-gone homes in occupied Palestinian territory. The key is a materialized memory made sacred. The past has been frozen in time and made sacred, and is therefore beyond renegotiation and reconstruction. There are qualitative differences between a refugee setting and a DFDR displaced community, but I assume, despite the lack of ethnographic evidence from longitudinal resettlement research, that making selected elements sacred also takes place in the latter.

The link between prevalent resettlement theories (Cernea 1997; Downing and Garcia-Downing 2009; Scudder and Colson 1982), and heritage would thus lie in a processual view of a past that would offer a framework for the understanding of social disarticulation; and a reconstruction of routine culture with reference to collective memory and collective trauma as signifying practices. Evaluation practices would gain by learning to use these highly qualitative dimensions in resettlement projects.

The Intricate Use of the Past

As in the case of the Palestinian refugees who hold onto their house keys—frozen in time, sacred, and beyond reconstruction—it can be assumed that people in DFDR projects suffer from similar dissonances that hamper them in the reinvention of new routine cultures, and thereby influence the process of transformation. In Zimapán, the different types of heritage (dissonant, unwanted, and uninherited) were all at play in the reconstruction and reinvention processes before, during, and after the physical displacement of the people.

Different types of heritage always coexist in a society, but the difference lies in the intensity of the selection process evoked by the resettlement. Under normal circumstances the selection process is slow and well marked: this house, tree, site, bridge, temple, and ritual. In DFDR projects, the resettled people have to make these decisions and selections, not only for singular objects, places, and traditions, but also for all their heritages and memories, in an all-embracing enterprise, during a very short time period (in Zimapán, four years). This is of course riddled with conflict, which adds an additional dimension to the displacement and resettlement process.

A particular heritage always belongs to someone, which implies that someone else will be disinherited. The selection of the past always signifies a

power dimension. The disinheritance of a particular group may be short-term, and a mistake in the process of selection, but it may also be long term, widespread, intentional, important, and obvious (Ashword and Tunbridge 1996). Whoever has the power to legitimize the selection is crucial when a particular place, monument, tradition, and/or memory is elevated and made into heritage. The process usually goes hand in hand with the institutionalizing of a particular heritage. The status, legitimization, elevation, and use of the past as a resource may lead to *dissonant* heritage, which is a vital part of being assigned the status of heritage (Smith 2006). Who is doing the interpretation, and how it is received by the people, is also decisive (Ashword and Tunbridge 1996).

There is a built-in tension associated with the creation and definition of the values, meanings, and symbols of a particular heritage. In Zimapán, the resettlement triggered several processes associated with this array of different heritages: some of them were contested, others were accepted, but none were harmless. The resettlement also trigged a political turnover, with the political power being transferred from the smallest, oldest, Catholic village La Vega, to the biggest, newest, Adventist village Rancho Nuevo. This power struggle was deeply anchored in history.

Politics and Legacy

La Vega claimed to be the oldest settlement in the valley and the bearer of a heritage going back centuries. The villagers tied this legacy to the Otomi identity, the Catholic religion, and their economy. The people insisted that they originated from the Otomi town Tecozautla in Hidalgo, the traditional marketplace of the valley. They had settled in the valley because of their kinship with the Otomi families across the river in Hidalgo. They had built a cable car that carried goods, people, and animals across the river, and had sustained a walking path to the market in Tecozautla, which was used by all. In the valley, they controlled the water through a small irrigation dam behind the village, which was occasionally used as a weapon in conflicts with the other villages. A spatial analysis reveals that in the past, La Vega was at the front of the valley; they were thus the gatekeepers and guardians of the valley.

The political *ejidal* structure confirms this. La Vega had upheld the office as *ejidal* president in the majority of the mandate periods (9 out of 12), since the official foundation of the *ejido* in 1937 (Aronsson 2002). This legacy was broken with the resettlement, when the *ejidal* president was shot and killed during project implementation. This human tragedy is forever documented in "The Ballad of Zimapán," which was written and performed by the La Vega brothers (Aronsson 2002, 285). From here on, La Vega began to withdraw from the negotiations, and the power was transferred to Rancho Nuevo.

In the negotiations, La Vega was depicted as the most "traditional" of the three villages, while Rancho Nuevo became the "progressive" one. Somehow this generated the idea that Rancho Nuevo was better able to cope with the resettlement and future needs. This impression influenced the negotiations and informed the design and idea of the new village.

From a heritage point of view, however, La Vega *was* the traditional village. For one thing, in contrast to the other two villages, it had never changed its name. Place names are not random, and they are ethnographic evidence of an unbroken continuity with the past. Furthermore, this sense of continuity was combined with a strong identification with the landscape, in accordance with Wolf's "bundle of relationships" and Ingold's "unfolding fields of relationship" (Ingold 2000; Wolf 1959, 106). Although all three villages paid deep attention to the landscape, there was something that made La Vega claim that their land was more productive than that of the others, and that therefore they were entitled to more compensation. This claim can be seen as an economic argument, but it might also be part of a feeling of alienation in the resettlement process that stressed "progress" without consideration of a more comprehensive understanding of the past.

In the new village, the La Vega sector seems to have reinvented (or never lost) the routine cultures of the valley to a greater extent than the other sectors. Only time will tell, but after 20 years there are signs that La Vega has found a way forward by balancing new and old structures, a way that promotes a reestablishment of routine culture that is "calmer" and more adapted to the past routine cultures that existed down in the valley. The spatial reinvention of the section confirms this—there are *milpas* with all kinds of vegetables, prickly pear, fruit trees, and animal corrals in conjunction with the houses. The spatial analysis also reveals an adaptation of scale: the school with a well-kept garden (La Vega had the first school in the valley), and a small church. The built environment is in harmony with the size of the village, and lacks the grandeur of the rest of the new village. Furthermore, there are no high walls surrounding the houses, and they are not as massive as in the other sections; the people can see and talk to their neighbors from their porches.

This familiar *ambiente* has been consolidated by the fact that the new village extends beyond the La Vega sector, where the villagers build their own houses, streets, paths, and gardens. Everyday activities were familiar, and included such mundane tasks as the burning of garbage, the hanging of laundry, chatting with the neighbors over the fence, parking the car, and attending to the animals. These routine cultural activities have created a familiar sound and olfactory landscape. The silence and desertedness that dominated the other sections of the new village are absent.

One tentative interpretation is that the other two villages, Rancho Nuevo and Vista Hermosa, have become too different from their past, and thereby have lost viable elements for repairing and consolidating themselves to the same extent that La Vega has done.

The focus of the negotiations was, thus, not the past, but the future, which was accentuated by the fact that the power to legitimize the past was put in the hands of Rancho Nuevo, whose past had been driven by radical change, revolution, and the liberation from the hacienda in the valley, and the adoption of a new religion, Seventh Day Adventism. A low-intensive power struggle had always existed between La Vega and Rancho Nuevo, with Vista Hermosa functioning as the mediator.

A resettlement triggers and reinforces the existing sociocultural elements that rest deeply in the spatial and temporal orders manifested in

heritage and memories. Along these lines, the political turnover reconnected to intricate past sociopolitical structures, and directed Rancho Nuevo into a position—and maybe a perceived right as well—to define and interpret the past for the present and future needs, which culminated in the strains of the resettlement. The struggle between La Vega and Rancho Nuevo was therefore more than a struggle over political and economic resources. It was a struggle about the right to define the future by the use of the past, which was articulated in the sociocultural dimensions of the society. The resettlement "unfolded" the past. To reduce it to a struggle over resources is to diminish its potential. This could have been used as a creative force in the rebuilding of the society: the ethnography was there, but it was neither seen nor used. An opportunity was lost.

The Devil as Heritage

Heritage is always present, even if the society denies its past, because the process of remembering is also a process of forgetting. In other words, part of remembering the past is selecting which memories to forget. This creates not only a *dissonant heritage*, but also an *unwanted heritage* and an *uninherited heritage* (Grydehøj 2010). In the valley, there were all kinds of heritages. Here I will briefly discuss "the heritage of the devil."

In Mexico, the devil is frequently mentioned in the ethnography of resettlement (Barabas and Bartolomé 1973). The devil is also a recurrent theme in South America: it is associated with the fetishization of evil, and is seen as a mediator in conflicting views on the objectification of the human condition (Taussig 1980).

According to local tradition, the Devil lived in the canyon that bears his name, and his body constituted the symbolic landscape of the valley. The dam wall was built in his canyon, and that disturbed him deeply. When the construction of the tunnels in the mountain began—the "opening up" of the mountain—he appeared before the dam workers in the shape of a huge woman dressed in black in 1989 and said that the mountains were his children (Aronsson 2002, 158).

The Devil's body symbolically constituted the natural and cultural landscape in the valley by lying down outstretched in the valley in a northeast to southwest direction, with his head as the village La Vega to the southwest, his stomach as the village Vista Hermosa (in the middle), and his legs as the village Rancho Nuevo to the northeast. His feet faced his canyon.

In the new village, the Devil reappeared in the village´s symbolic spatial outline, but with one crucial difference. Instead of lying outstretched, he is now in a fetal position. The spatial order corresponds to the positions in the valley: the head is the La Vega sector, the stomach is the Vista Hermosa sector, and the legs are the Rancho Nuevo sector, but now the Devil is cringing. This was brought to my attention in a spontaneous discussion that took place in the new village. The story was told accompanied by big smiles, as if the people were distancing themselves from this information. Surprised, I asked what they meant by this, and I was told that even the Devil had to give in to the World Bank. The symbolic representation was hence not questioned

per se: the important thing was seen as, rather, the change in the Devil´s body posture. Is this ethnographic evidence of a transformed and reinvented expression of a diminished routine culture? Or is it just another esoteric ethnographic anecdote? In any case, what makes it conspicuous is that it is a dissonant and unwanted heritage made visible, and maybe in the future it will be an uninherited heritage as well, present in the collective memory of the resettled society.

The Villages' Spatial Orders as Heritage

In this project, the World Bank recommended that the resettled villages should be lifted up and placed in the same spatial order as they were in the original village outline. For Bella Vista del Río, the placing of the three villages was preceded by tough negotiations between the villagers without the involvement of other actors. The negotiations resulted in a spatial order that placed Rancho Nuevo and Vista Hermosa closest to the main road and the main entrance. La Vega was placed at the back of the village, farthest away from the main road. The village sections were thus placed in accordance with the valley´s spatial outline as it appeared at project start.

The past tells us, however, that La Vega had once been the gatekeeper of the valley, located at the front, facing Hidalgo, before it was spatially turned around in favor of Rancho Nuevo in the 1960s. This happened when the new road was built to Cadereyta in Querétaro. Rancho Nuevo had promoted the road, and La Vega had argued against it, and refused to collaborate. The road was built with state money and labor from the valley, and it entered the valley in Rancho Nuevo, which therefore moved into the front position, while La Vega was demoted to the back. This was the spatial order that was recreated in the new village.

Twenty years after the resettlement, the pieces of the new village do not seem to fit together harmoniously, echoing an early ethnographic observation I had made. In 1994 I saw a truck parked in the new village in the La Vega sector. A motto painted on the truck stated, "My village is in agony" (Aronsson 2002, 195). A message like this makes it clear that any "lesson learned" policy has failed utterly in this situation.

The Ballad of Zimapán as Intangible Heritage

Another example of heritage that appeared during the resettlement was "The Ballad of Zimapán." The heritage of a place is maintained and reproduced across the generations through stories, songs, and poetry. In "The Ballad of Zimapán," which was composed and performed by the La Vega brothers in 1994, the collective trauma of the society has been documented forever. The ballad consists of nine verses that describe the agony of having to leave the valley of *Ejido* Vista Hermosa, the feeling of being betrayed, and the anger of being targeted by the state, forced to move, and be "developed" in Ezequiel Montes, a town located closer to the main national culture and the majority society. The ballad was attainable for analysis during the implementation, but it was not used. It was not even considered to be of importance for the

understanding of the cultural response to the displacement. In 2013, a follow-up video, *Dueto C.V. Bella Vista del Río*, was available on YouTube.[4]

Both the ballad and the statement painted on the truck are cultural expressions of a collective trauma that these villages experienced and lived through. This goes beyond individual pain, depression, and memory, and has made its way to a higher level of abstraction. This pain belongs to the collective memories of the community, and its heritage. The collective trauma and the feeling of victimization go hand in hand.

LOST AGRICULTURAL LAND AND NOTIONS OF FREEDOM

As mentioned earlier, the new village was built on the plateau with the argument that the villages wanted to preserve their social cohesion, and not separate the landowners and the landless. The *ejidatarios* rejected the replacement land located closest to the town Ezequiel Montes. This was a majority decision taken at a general assembly. In 2013 the reasons for the rejection had been reformulated, and the matter of social cohesion was no longer mentioned. The new statements were: "they wanted their freedom," "their independence," "they were used to dealing with their own stuff," "it is too crowded," and "it is too close to the municipal authorities." But these arguments mirror the arguments of the conflict about the road in the 1960s. The citizens of La Vega insisted then that "they did not want the law of the municipality to enter the valley" (Aronsson 2002, 77).

These arguments could be attributed to the place-attachment model, but I find that they rather express underlying existential themes, such as a hesitation to come too close to the national society with its values, lifestyle, and demands from authorities. There is a perceived ontological distance between the valley and the national culture that is connected to a long-term strategy of retaining a heritage, collective memories, experiences, and notions of freedom that go far back in history. In 2013, other voices were heard, claiming that it might have been a mistake not to accept the replacement land.

CONCERNS ABOUT MEMORY AND SACRALIZATION

In DFDR projects different types of heritages and memories are at play in the reconstruction and reinvention processes. An understanding of the past is needed for a reinvention of routine culture, but at the same time a sacralization of the past could block its transformation through generations. The structures at play, heritages and memories, are not passive, but active, agents that either impede or speed up the reinvention and reconstruction processes. I suspect that people in DFDR projects live for years with a dissonant, unwanted, and uninherited heritage that they are forced to make livable and adapt to a new life-world. They have to learn to forget to remember, and

[4] https://www.youtube.com/watch?v=yIOA-9uwsC0.

remember to forget, while they fight for their livelihood. They have to learn how to practice selective forgetfulness for the benefit of their community.

The Zimapán resettlement shows that the creative forces of society and the components of reinvention are connected to these intersubjective memories, but more research is needed to exploit its full potential for policy and evaluation. For example, Misztal, with reference to Halbwachs, points out that memory is a community issue that is embedded in societal values rather than merely a psychological function of cognitive capacities (Halbwachs 1941/1992; Misztal 2005). This line of thought entails that memories have a stabilizing effect on societies because of their normative and calming functions; memories give people a sense of meaning and place in the world (Schwartz 2000).

In DFDR projects, a focus on memories could, however, be both sensitive and dangerous. I am concerned that if the "wrong" memories are triggered, we might end up with a very complicated resettlement, with a sacralization and freezing of the past, similar to what displaced people in conflict zones experience. Furthermore, efforts to mitigate this might even create a state of mind that would be part of this "freezing" and dwelling in the past and a sense of victimization. The victimization syndrome has been a problem for a long time in resettlement.

Finally, I encountered methodological problems in the field when I tried to ethnographically document memories in Bella Vista del Rio in 2013. I did not seem to be able to formulate the right questions: and people did not relate to my question, "What do you remember from the past in the valley?" However, a breakthrough moment came early one morning in a villager's kitchen, when I was told that when they talked about the valley and their past life, they "cried and remembered the food, fruit, vegetables, smells and sound, all down there."[5] These are sensory memories, bodily experiences, and emotions felt and explained in a holistic view of the past. I realized that the cognitive approach I had been using was insufficient. I concluded that the use of a phenomenological methodology is more suitable in approaching the memory complex. This would challenge any evaluation scheme for resettlement, because the data that is produced with such a methodology is qualitative.

KNOWLEDGE PRODUCTION AND EVALUATION

During implementation, a complicated kind of knowledge production is generated between the local people and the implementing company. The quality of this knowledge governs the outcome, and badly implemented projects have severe consequences, as can be seen by studying the World Bank Inspection Panel (which had 106 cases in 2016).[6] I have argued elsewhere that DFDR projects are so complex that they have similarities to art installations, and

[5] L.C. Bella Vista del Rio, personal communication, July 2013.

[6] The World Bank Inspection Panel, http://ewebapps.worldbank.org/apps/ip/Pages/AllPanelCases.aspx.

performances that result in ad hoc solutions (Aronsson 1992). de Wet has also observed this complexity, but relates it to the capacity of the stakeholders to create a "moral space" for meaningful communication (de Wet 2009, 86). In a similar vein, Hermans, El-Masry, and Sadek (2002) discuss participation by stressing the pedagogical aspects of communication. My objection to these approaches is that knowledge production in resettlement must never be reduced to a pedagogical training exercise between more or less rational agents. There are aspects of the pedagogical methods that need improvement, but that is not all that is needed. We need to recognize how the past is embodied in material objects and expressed in the intangible narratives that determine the knowledge product that is to be executed and evaluated. The past has to be critically evaluated and reflected upon during implementation, as well as in its evaluation.

In Zimapán, during the implementation the main stakeholders became engaged in a ritual dance that had less to do with solving the project´s everyday problems than with the upholding of self-defined positions. The stakeholders were locked into past structures and positions that regulated their behavior and attitudes. The behavior became routine, and the understanding and competencies became self-generating categories, similar to performative rituals that influence praxis, resulting in the standards (policy) not informing and shaping praxis. Standard policy guidelines and praxis had failed to build a problem-solving platform for the purpose of generating operationally useful knowledge for all stakeholders. The knowledge was there, but policy and praxis did not interlock.

Knowledge is not only linked to power: first and foremost it has an ability to make itself "true" (Foucault 1977, 27). For resettlement implementation and evaluation, this is relevant, because all classification schemes (understood as spatial-temporal segmentations of the world) have a tendency to become "true" and taken for granted with time. They become standards.

The knowledge produced during resettlement is always categorized and put into boxes to become operational entities. The problem arises when a knowledge category becomes "true" without any consideration of its potential value for the project. The principle must be that in a participatory-informed project all knowledge (both local and expert) must be scrutinized and assessed from a systematic perspective. The romanticizing of local knowledge, based on a relativistic view of culture, can be devastating. Equally devastating is blind faith in blueprint knowledge, based on a classification scheme that has become "true in itself." Instead of either of these, all knowledge produced must prove its solution value regardless of whose knowledge it might be. The challenge is in dealing with contradictory knowledge, and finding ways to identify, analyze, and address it.

In the goal-free evaluation method, power and self-generating categories are in focus, because it is assumed that if an external evaluator "intentionally avoids knowledge of and reference to the program´s stated or official goals and objectives" neutrality could be upheld (Youker, Ingraham, and Bayer 2014). The evaluator moves backward in the project to discern the effects of the implementation without any informed knowledge about the project goals. Goal-free evaluation is mainly associated with qualitative

data collection methods, a multilayered approach, and evaluation indicators. However, in DFDR projects the stakes are high, and when the project moves forward, everything is intensified: there is therefore a risk that these soft life-skills indicators may be set aside in favor of the material and compensatory aspects crucial for livelihood.

Participatory methodologies aim to incorporate local knowledge and empower local people. This requires that the locals be trained in participatory methods, and that they build their capacity. Participatory evaluations have shown, however, that there is a clear division in the tasks and responsibilities: program staff design evaluation and data analysis, while local participants collect the ethnographic data.

The intricacy of DFDR projects inevitably leads to complexity theory and its application for evaluation. Briefly, complexity theory is not one single, coherent body of thought, but rather consists of bundles of interacting stakeholders, objects, and processes bound together by interest or functions. These interactions are nonlinear, open to feedback, and difficult to predict. Because of their uncertainty and nonlinearity, complex social systems are difficult to evaluate, and there is no consensus in the research literature about what can be useful for their evaluation (Walton 2014).

A final observation from the praxis in Zimapán involves the monitoring team. With time, their reports came to evince strong similarities with the implementer's reports. In fact, it became almost impossible to see the difference between a monitoring report and a management one. The World Bank conclusion was that the team was young and inexperienced, and therefore could not uphold their position in the face of the management of the Comisión Federal de Electricidad (World Bank 1997). From my everyday experiences with the team, I think this had to do with their lack of trust in their own knowledge. Much of this knowledge belongs to the "soft" socio-cultural dimensions, and thus did not fit the matrix—the expected (or "true") knowledge categories. Most of this tacit knowledge was related to the past and how to live a good life, but it did not find its way into any of the reports, whether they were management, monitoring, or evaluation reports. There was no appropriate language, no classification schemes that could be used, and these observations were therefore left aside. If we add to this complexity the knowledge that "open-ended, non-fixed, non-politicized collective memory is good for cooperative relationship" (Misztal 2005), we are faced with an even more complex system, which nevertheless might be a step closer to introducing the use of the past in resettlement—although it will undoubtedly still be an imperfect use of the past.

ACKNOWLEDGMENTS

The Magnus Bergvalls Foundation, Sweden, funded my fieldwork in Mexico 2013. Thank you.

REFERENCES

Anderson, Mary B. 1999. *Do No Harm: How Aid Can Support Peace—or War*. Boulder, CO: Lynne Rienner Publisher.

Aronsson, Inga-Lill. 1992. "Impresiones de un Proyecto de reasentamiento." *Alteridades* 2 (4): 51–59.

———. 2002. "Negotiating Involuntary Resettlement. A Study of Local Bargaining during the Construction of the Zimapán Dam." Occasional Papers 17, Department of Cultural Anthropology and Ethnology, Uppsala University, Uppsala, Sweden.

———. 2013. "Det Ambivalenta Kulturarvet." *Dragomanen* 15: 89–102.

Ashword. Gregory J., and John Tunbridge. 1996. *Dissonant Heritage: The Management of the Past as a Resource in Conflict*. Chichester, UK: Wiley.

Barabas, Alicia, and Miguel Bartolomé. 1973. "Hydraulic Development and Ethnocide. The Mazatec and Chinantec People of Oaxaca, Mexico." IWGIA, 15. Copenhagen.

Cernea, Michael. 1997. "The Risks and Reconstruction Model for Resettling Displaced Populations." *World Development* 25 (19).

CFE (Comisión Federal de Electricidad). 1991. "Regimenes de Propiedad. Afectados Por el Proyecto P.H. Zimapán." Unpublished report.

———. 1996. "Informe Final Sobre el Programa de Reasentamiento del Proyecto Hidroeléctrico Zimapán." Unpublished final report to the World Bank.

de Wet, Chris. 2009. "Does Development Displace Ethics? The Challenge of Forced Resettlement." In *The Crisis of Forced Displacement and Resettlement*. Anthony Oliver-Smith, ed. Santa Fe: School for Advanced Research, Advanced Seminar Series.

Downing, Theodore E., and Carmen Garcia-Downing. 2009. "Routine and Dissonant Cultures: A Theory about the Psycho-socio-cultural Disruptions of Involuntary Displacement and Ways to Mitigate Them without Inflicting Even More Damage." In *Development and Dispossession: The Anthropology of Displacement and Resettlement*, Anthony Oliver-Smith, ed., 225–53. Santa Fe: School for Advanced Research Press.

Foucault, Michel. 1977. *Discipline and Punish: The Birth of the Prison*. New York: Pantheon Books.

Grydehøj, Adam. 2010. "Uninherited Heritage: Tradition and Heritage Production in Shetland, Åland and Svalbard." *International Journal of Heritage Studies* 16 (1-2): 77–89.

Halbwachs, Maurice. 1941/1992. *On Collective Memory*. L.A. Coser, editor and translator. Chicago: University of Chicago Press.

Hermans, Leon M., Nader El-Masry, and Tarek M. Sadek. 2002. "Participative Water Management: Social and Ecological Aspects: Linking Actors and Models for Water Policy Development in Egypt: Analyzing Actors and their Options." *Knowledge, Technology and Policy* 14 (4): 57–74.

Ingold, Tim. 2000. *The Perception of the Environment. Essays on Livelihood, Dwellings and Skill*. London: Routledge.

Josefsson, Johan, and Inga-Lill Aronsson. 2016. "Heritage as life-Values: A Study of the Cultural Heritage Concept." *Current Science* 110 (11): 2091–98.

Misztal, Barbara A. 2005. "Memory and Democracy." *American Behavioral Scientist* 48 (10): 1320–38.

Schwartz, Barry. 2000. *Abraham Lincoln and the Forge of National Memory*. Chicago: University of Chicago Press.

Scudder, Thayer, and Elisabeth Colson. 1982. "From Welfare to Development: A Conceptual Framework for the Analysis of Dislocated People." In Involuntary Migration and Resettlement: The Problems and Responses of Dislocated Peoples, Art Hansen and Anthony Oliver-Smith, eds. Boulder, CO: Westview Press.

Smith, LauraJane. 2006. *Uses of Heritage*. London: Routledge.
Smyth, Eddie, Michael Steyn, Ana Maria Esteves, Daniel M. Franks, and Kemal Vaz. 2015. "Five 'Big' Issues for Land Access, Resettlement and Livelihood Restoration Practice: Findings of an International Symposium." *Journal of Impact Assessment and Project Appraisal* 33 (3): 220–25.
Taussig, Michael. 1980. *The Devil and Commodity Fetishism in South America*. Chapel Hill, NC: University of North Carolina.
Walton, Mat. 2014. "Applying Complexity Theory: A Review to Inform Evaluation Design." *Evaluation and Program Planning* 45: 119–26.
Wolf, Eric. 1959. *Sons of the Shaking Earth*. Chicago: Chicago Press.
World Bank. 1997. "Implementation Report, Mexico. Hydroelectric Development Project Loan 3083-ME, Infrastructure and Sector Leadership Group, Mexico Department, Latin America and the Caribbean Regional Office." Evaluation Report No. 16835, World Bank, Washington, DC.
Youker, Brandon W., Allyssa Ingraham, and Nicholas Bayer. 2014. "An Assessment of Goal-Free Evaluation: Case Studies of Four Goal-Free Evaluations." *Evaluation and Program Planning* 46 (10).

Chapter 16

Evaluating the Benefits and Costs of Resettlement Projects – A Case Study in the Philippines

Marife M. Ballesteros

Abstract. *The efficiency and impact of two types of resettlement modes undertaken by the Philippine National Housing Authority are compared. Production efficiency was measured from the cost/benefit ratio (CBR) of the present value of the total project cost and estimated housing rental values of specific resettlement projects developed from 2004 to 2011. The socioeconomic impact analysis involved a small sample of households, matched based on household characteristics using propensity matching. Results show that the use of government resources in in-city developments is more efficient: for in-city projects the government spends from ₱0.62– ₱0.76 for every peso of housing benefit; for off-city projects, the benefit is an estimated CBR of ₱1.72. The study noted that the initial gains of off-city resettlements—lower investment and administrative costs, and the provision of a house-and-lot package to affected families—are erased by compromises on the quality and sustainability of resettlement sites. The household income of resettled families is lower, and participation in the schooling of children is also lower in off-city sites. The government must prioritize in-city resettlement through longer-term planning, and consider alternative in-city housing options*

Marife M. Ballesteros, Philippine Institute for Development Studies, ballesteros.marife@gmail.com.

for affected families. It should also review policies on off-city resettlements, specifically the selection of sites and the extent of involvement of the community in resettlement planning.

The rapid pace of urbanization in the Philippines has led to housing challenges that are visibly manifested in poor housing conditions. In 2015, it was estimated that about 18 percent of the total population was living in blighted conditions. Poor housing is most evident in cities, specifically the capital city of metropolitan Manila, which is home to about 2 million slum dwellers. This condition is as much the result of unplanned urban growth as it is of low income levels. In general, the country lacks integrated urban planning, and there is poor coordination between spatial and structural transformation in cities (World Bank 2016). A result of this unplanned growth is haphazard land development, during which the illegal occupation and conversion into residential settlements of land for public use is commonly observed. Over time, there has been a proliferation of these illegal residential settlements (also known as "informal settlements"), and an increase in the number of families occupying them.

This condition has constrained the government to build the critical infrastructure needed for effective and efficient urban services. Metro Manila was among the world's top 20 most populous megacities in 2015, but despite rapid urbanization, the city has not benefited much in terms of economic growth and poverty reduction, compared to other Asian countries (World Bank 2016). The city has poor connectivity with peripheral towns and cities, and even within Metro Manila there are areas outside of the main transit routes, or areas not linked to citywide social services and infrastructure.

In the last two decades, the Philippine government has put into action an infrastructure plan to address these inefficiencies. In particular, the plan identified expressways, railroads, and a flood control system as the major infrastructure projects for the expanding Metro Manila region. These projects are expected to involve massive relocation and urban renewal in some parts of the city: involuntary resettlements thus are unavoidable.

The Philippine policy and institutional framework for dealing with involuntary resettlement provides for humane procedures for relocation and resettlement. Affected communities are engaged in a consultative process that covers housing options, government resources, livelihood support, and the protection of vulnerable persons. However, resettlement action plans have been primarily focused on providing housing, and often have failed to account for the loss of incomes and the social networks of households. Oftentimes, the government tends to look at resettlement in terms of short-term results: that is, how to move out affected families and provide them with permanent homes in the shortest time and at the least expense. This action tends to favor off-city resettlement, given the tedious process of searching for adequate land to fit the large number of affected families, and the limited availability of low-priced land in the city. Moreover, providing affected families with their own houses and lots seems to present a better picture than housing them in urban, multistory buildings.

However, the lower initial cost and house-and-lot package in off-city resettlements do not necessarily result in a cost-effective government investment compared to the cost of in-city projects. Off-city projects can be counterproductive, since the displacement of families has an impact on their livelihoods and other economic opportunities, as well as on their social networks and psychological well-being, their access to basic services, and their opportunities for skills development. In other words, the initial gains from off-city resettlement projects can have adverse effects on families' overall welfare.

In cases where off-city resettlement cannot be avoided, the provision of adequate basic infrastructure in the resettlement area must be ensured prior to relocation. Resettlement sites should also at least be in municipalities identified as subregional or provincial urban centers, and not in rural municipalities where economic opportunities are scarce. The inadequate basic infrastructure and lack of opportunities for jobs and livelihood is the main problem for off-city resettlements in the country: many of the involuntarily displaced people end up returning to the city after their resettlement.

This study provides a quantitative methodology for assessing the benefits and costs of resettlement projects that are implemented by the national government. Specifically, it compares two resettlement modes—in-city and off-city—and determines which mode provides the greater efficiency and the best social and economic outcomes.

The analysis considers resettlement projects of the National Housing Authority (NHA), the central government agency in the Philippines mandated to undertake housing production for families in the lowest 30 percent of income. The NHA is the lead agency in the resettlement of families affected by infrastructure projects of the national government: between 2003 and 2012, it has carried out massive resettlement of families for the proposed construction of the North and South Rail infrastructure project, which will link Metro Manila to peripheral towns and cities. This period has also coincided with the resettlement of families living along riverbanks, to address the flooding problem in Metro Manila.

The first section of this chapter provides an overview of NHA resettlement modalities and processes during this period. The second section presents the methodology used to evaluate the efficiency and impact of NHA off-city and in-city resettlement projects. The following section discusses the results of the efficiency and welfare analysis, comparing in-city and off-city projects. The final section presents conclusions and provides recommendations.

RESETTLEMENT MODALITY: PROCESS AND PROCEDURES[1]

The NHA classifies resettlement projects into either in-city or off-city projects. In-city projects refer resettlement sites that are developed in the same city or municipality where the affected families reside. Off-city projects refer to resettlement sites outside of the original settlement. Off-city resettlement

[1] This section draws on Ballesteros and Egana (2013).

areas are usually in distant locations, about 40–50 kilometers from the original settlement.

In off-site settlements, individual houses and lots can be provided, due to the availability of large areas of contiguous low-priced land. Resettlement sites within Metro Manila are smaller plots that entail the construction of multistory housing, with a higher density population. Multistory development also requires higher investment and maintenance costs, which means a higher per-unit cost of housing, and is usually unaffordable for most families residing in informal settlements. Under the North and South Rail Project, affected families in Metro Manila were mostly resettled outside of the city.

The success of a resettlement project is to a large extent dependent on how well it is implemented. The NHA adheres to a humane approach to resettlement. It carries out the resettlement process in several phases, during which social preparation is the central activity. Social preparation involves identifying beneficiaries and resettlement sites, and mobilizing resources. Affected families are organized, and the community goes through a capacity-building process in order to establish the social, organizational, and institutional norms and mechanisms that will enable resettled families to cope with their relocation, and encourage them to work together in partnership with concerned institutions and stakeholders. This activity covers two phases of the NHA's resettlement work program. It is the most critical stage in the resettlement process, since it involves the buy-in of both the community and the proposed resettlement sites, the involvement of several stakeholders, and the creation of committees and subcommittees at the level of the local government and the community. It also requires the longest time, because the NHA has to formalize agreements with both the sending and receiving local government units (LGUs), the community, and the developer. While this entire phase is programmed to be accomplished within three to six months for 1,000 affected families, delays often occur, due to the number of stakeholders involved in the preparatory work. There can also be prolonged resistance, or disagreements among the affected families. Often a longer consultation period is needed in order to resolve collective action and/or coordination problems with the government and other entities.

The NHA primarily applies the developer-constructed approach to resettlement projects.[2] Under this approach, the NHA partners with private developers to undertake the development of resettlement sites and the construction of housing, based on standards of socialized housing.[3] The NHA

[2] An alternative approach is the housing material loan, which is an incremental housing approach whereby NHA provides the developed site with core housing (i.e., a box house), and beneficiaries take charge of housing improvements based on their affordability level. Although this approach has been observed to have better outcomes, it is not popular with the NHA: the agency finds it tedious both administratively and physically, since it has to be concerned with the process of material acquisition and housing construction as well (Ballesteros and Egana 2013).

[3] Socialized housing refers to housing projects for the underprivileged and homeless, following the national law Batas Pambansa 220 on subdivision development

accredits the developer, who then provides a list of proposed resettlement sites, with approved development permits and locational clearance, and the pricing of the housing units.[4] This approach is administratively less costly to the NHA, since the agency does not have to engage in land banking, and simply contracts the private developer to supply the developed site and housing units. On the other hand, there is incentive for the private developer to engage in the project because of the captive market. The developer does not have to look for buyers for each unit, since the entire development is being purchased by the NHA for the beneficiaries of its resettlement program. The private developer can also use the resettlement project for compliance with the Balanced Housing Development Act, which requires developers of proposed subdivision projects to develop an area of socialized housing equivalent to 20 percent of the total area, or the total cost, of proposed projects.[5]

Upon accreditation by the NHA, the developer offers the community the site, and schedules site visits for community officers. The community officers have to formally endorse the project to the Local Inter Agency Committee (LIAC), which in turn endorses the project to the NHA.[6] The endorsement from the community, and the recommendation of the LIAC, enables the NHA to finance and purchase the housing units for each community member from the selected developer.

It is important to note that the community can select only projects offered by NHA-accredited developers, which in most cases might just be a choice of two sites from one developer, or two sites from two different developers. Moreover, although the NHA provides the criteria on site suitability for resettlement projects, the endorsement of the LGU, based on the approved subdivision plan and locational clearance, is sufficient for the NHA. However, the approved subdivision plan considers only the land use and environmental

standards and price ceilings, as determined by the Housing and Urban Development Coordinating Council and the National Economic Development Authority. During the period in review, the price ceiling for socialized housing was set at ₱400,000 per housing unit, which is usually applied in highly urbanized areas. In towns and municipalities, the NHA sets a lower price as determined by its board.

[4] The development permit and locational clearance are certifications issued by the LGU that certify a specific site for residential use and suitable for residential subdivision development.

[5] Urban Development and Housing Act (RA 279 of 1990). The developer may partner with other developers to invest in these projects.

[6] The LIAC is formed at the start of the resettlement process. The members consist of representatives from the sending LGUs, local nongovernmental organizations, and representatives of national government offices (e.g., the Department of Interior and Local Government, the Housing and Urban Development and Coordinating Council, the Department of Social Welfare and Development, the Department of Environment and Natural Resources, the Department of Public Works and Highways, the NHA, the National Poverty Commission, the Metro Manila Commission, and the Presidential Commission for the Urban Poor). The LIAC is chaired by the local housing board representative and cochaired by the NHA.

suitability of the site: it does not include the socioeconomic feasibility of the area, such as conditions of employment (e.g., distance to employment centers); access to schools, markets, and tertiary hospitals; or distance of the site to energized sources of water and power. The NHA argues that social services (e.g., schools, health centers, etc.) can be provided over time, and that the concerned national agencies should include the construction of social facilities for resettlement sites in their respective budgets. As to basic utilities such as water and power, the NHA simply requires the selected developer to provide shallow wells and power generators in areas that are far from energized sources.

After social preparation is completed and the approval of the site has been obtained from community officers and LIAC, the relocation of affected families follows. The relocation phase starts when the NHA, the community officers, and the developer have signed contract agreements. This phase involves preparatory work such as a period of dismantling structures at the evacuated sights, and preparation for staging areas if needed. In most cases, the site has been prepared prior to relocation, except for the individual power and water connections, which are usually provided at a later period. Upon completion of the preparatory activities, the actual relocation usually takes about a month for 1,000 beneficiaries (an average relocation rate of 50 families per day). Weather conditions can slow down the process. There are relocation guidelines that must be followed. The NHA and representatives from the Philippine Commission on Human Rights, and the Presidential Commission for the Urban Poor ensure that relocations are undertaken within the legal guidelines. The NHA takes the lead in relocation activities, with support from the sending LGU. The sending LGU also provides financial assistance of not less than ₱1,000 per family. Some sending LGUs, especially in more prosperous cities, provide additional compensation, such as a week's supply of groceries and/or the extension of health privileges to their former constituents for a period of one year. At the resettlement site, the relocated families are received by the NHA local office, and the assigned community representatives.

The post-relocation phase starts with the termination of the relocation operation, and turnover of the evacuated sites to the concerned government agency—usually the Department of Public Works and Highways, or the Philippine National Construction Corporation, or the Philippine National railways. The developer has to also turn the resettlement project over to the NHA and the community. The resettlement process is deemed completed at this stage.

Once a site has been approved for resettlement, the developer is only responsible for the site development, and for the construction of core housing. The developer is not responsible for the construction of community facilities, or for the installation of water and power services to individual households.[7] In off-site areas that are far from energized sources of water and power, developers are only required to provide bulk water facilities

[7] Utility companies usually require 90 percent occupancy of subdivisions prior to connections.

sourced from shallow wells, and from generators for power supply. These facilities have to be maintained by the community and the NHA upon turnover of the resettlement project. While shallow wells and generators are considered stopgap or temporary measures, the community may have to wait several months or even more than a year before they can be connected to the local water or power systems. Meanwhile, these resettlement areas have to bear the higher cost, and less effective water and power systems, compared to households that have access to the local water districts and power lines.

Table 16.1 shows the status of in-city and off-city resettlement sites for the affected families of the North and South Rail Project. Upon relocation, in-city resettlement sites are provided with basic infrastructure facilities and improved houses made of strong materials. In off-city resettlement sites, although all affected families are provided with a house and a lot, and with better subdivision roads and houses made of stronger materials, basic services—that is, water and power—are not fully available upon relocation. In some completed sites, access to water and power are rationed, and are only available at specific times. There are also cases when some of the shallow wells are not operational, due to water potability. The NHA has not been able to readily address this situation, due to the high cost of energizing the sites, the limited budget of the agency, and the low income levels of families in the resettlement site. The installation of additional generators and water pumps would require a higher subsidy per family. On top of this subsidy, the NHA has to subsidize the maintenance of these machines. The sites may eventually be further improved, but these improvements depend on the availability of funds from the NHA, or grants from local politicians or external funders.

The NHA has established an estate management office in each of the resettlement areas, but their duties are mainly focused on loan collection and monitoring.[8] The NHA considers the resettlement program as a cost-recoverable program in which beneficiaries share in the cost of the development by paying for the cost of housing unit and lot over time. The site development cost is part of the government subsidy, while the cost of housing, including the lot, is a loan to each family that is paid for on a monthly basis over a period of 30 years. NHA collection performance, however, has historically been low, with an average collection rate of only 30 percent. This is one reason why the NHA has not been able to disengage from resettlement sites, as these sites remain assets of the agency unless they are fully paid for by the community. Moreover, the NHA has not been able to turn over the common areas of resettlement sites to the host LGUs. LGUs usually treat these sites as NHA properties, and are not keen to take on the responsibility for maintenance of common areas and utilities. There are several reasons why this is so: one, LGUs usually do not generate real property taxes from these areas; two, there could be development problems, such as landslides, maintenance of

[8] The NHA's monitoring and evaluation system is limited to occupancy and collection performance of resettled families. There are no systematic records on whether the resettled families have left the area or whether the housing has been transferred or sold to current occupants.

TABLE 16.1 **Summary status of resettlement projects for affected families in Metro Manila, North and South Rail Infrastructure Program (as of 2016)**

Resettlement site	Land area (hectares)	No. of developed sites	Lots generated	No. of occupant families	Sites completed		Status of power and water connection in completed sites 6 months after relocation
					No.	% of developed sites	
All	798.33	45	113,357	106,869	30	71.0	
Metro Manila	85.09	6	13,313	12,941	6	100.0	
Bulacan (off-site)	181.61	14	27,497	27,236	10	71.0	Two sites with no local power and water connection; two sites with no local water connection; water supply from shallow wells, but not all are functional; in some completed sites, there are portions with no water connection, affecting some 1,853 households
Pampanga (off-site)	203.70	6	13,198	12,903	6	100.0	In some completed sites, 600 households with no local power and water connection
Laguna (off-site)	181.92	9	35,402	32,091	4	44.0	In some completed sites, some shallow wells are not functional; others rely on commercial water distributor
Cavite (off-site)	24.25	2	4,117	3,861	1	50.0	Some shallow wells not functional
Rizal (off-site)	121.76	8	19,830	17,837	5	63.0	Some completed sites have bulk metering

SOURCE: NHA North South Rail Project Report 2014; focus group discussions.

NOTE: Developed sites are sites turned over by developers to the NHA, and considered ready for relocation. Completed sites are developed sites that have been turned over to beneficiary households: land development and housing construction is completed, and households have local power and water connection.

water and power systems, and weak community ownership of the area; three, some off-site resettlement sites are located in fourth or fifth-class municipalities that do not have enough funds to support the social services needed by the communities and the new settlers in the short to medium term, and therefore the LGUs require continued support from the NHA.

METHODOLOGY

The study used both efficiency and welfare measures to compare the two resettlement modes—in-city and off-city—that were undertaken by the NHA in cases of involuntary resettlement in Metro Manila.

To measure the efficiency of government investments for in-city and off-city resettlement projects, cost/benefit ratios (CBRs) were derived, based on estimated present values of the total costs of housing provision and the expected returns on the investment. This method captures the production efficiency in the use of government resources (Olsen 2000). The cost data were obtained from NHA records that include data on production, financing, and maintenance costs of specific in-city and off-city projects. The benefits were derived from the estimated value of market rents of housing in the location. The analysis assumes that the housing investment has a useful life of 30 years.[9]

In addition to comparing efficiency, the human welfare effects of in-city and off-city resettlement projects were also measured, using small-sample analysis of affected families. The data were obtained from the socioeconomic survey funded through the Social Impact Monitoring Project of the World Bank in 2010. The surveyed families were resettled families who had previously been living along the waterways of the Tullahan and Pasig Rivers, and were victims of Typhoon Ondoy (Typhoon Ketsana) in 2009. These families were initially moved to evacuation centers after their houses were washed out by the typhoon. The NHA, with assistance from the local housing board, selected from the list of evacuated families those who would be resettled in NHA resettlement sites in-city or off-city.

A total of 180 sample households was surveyed in the two sites; 100 households in the off-site resettlement in the province of Laguna (about 65 kilometers from Metro Manila), and 80 households that were resettled in a site in Pasig City, in the eastern part of Metro Manila. The 180 sample households were matched using household characteristics that were not affected by the resettlement project (e.g., age and educational level of the head of household, average household size before resettlement, etc.). The propensity-score matching performed on the sample households resulted in 163 matched households. Regression analysis on the matched households was used to determine differences in pre-identified outcome variables such as

[9] Socialized housing, given the type of building materials used, has a lower useful life than regular housing, which is estimated to have a useful life of between 50 and 70 years.

monthly household income and expenditures, school attendance of children, health status, and the employment of women.

COMPARISON OF IN-CITY AND OFF-CITY RESETTLEMENT: EFFICIENCY AND WELFARE IMPACT

Production Efficiency of Resettlement Projects

Table 16.2 compares production efficiency in the use of government resources for in-city and off-city resettlement projects. The cost components include both the investment costs and the operating and maintenance costs (including interest subsidies) that the NHA or the government incurs in the management of the resettlement site. Note that under the resettlement program, the NHA purchases the developed lots and housing from developers, and the beneficiaries amortize to the NHA the purchase price of the house and lot unit at a subsidized interest rate for a maximum period of 30 years. The benefit or return on the investment is the estimated imputed rents of the housing over a period of 30 years.

Based on recent NHA resettlement projects for Metro Manila, the cost of housing in off-city sites is only about ₱348,000 per unit compared to ₱917,640 per unit for in-city projects. The cost difference is due to the higher land prices and higher construction costs for multistory buildings for in-city projects. On the other hand, while off-city projects have lower land prices, these projects incur additional costs for the construction of community facilities. Controlling for size of resettlement and period of construction, resettlement sites in Metro Manila actually incur lower expenditures for community facilities, since schools, health centers, and livelihood infrastructures (e.g., markets) are already existing in the area, and are accessible to the community. Moreover, the community can readily connect to the local water and power districts, since the site is within the energized area. The households also remain constituents of the same city or municipality: thus they are already known by the LGU and are included as recipients of local services.

For off-site settlements, the total project development cost is lower, but this advantage is erased by additional investments in both physical and social infrastructures. Aside from government subsidies on the house-and-lot packages, resettled families are also given a housing subsidy that includes the utilities' expenses for installing power and water in the area, either by the provision of shallow wells or power generators, or as advance payment to utility companies to facilitate individual household connections.[10] In addition, resettled families are provided with a livelihood subsidy in the form of physical infrastructure and skills training programs. The infrastructure to support livelihoods includes capital outlays for the construction of livelihood facilities such as livelihood centers, tricycles, jeepneys, transport sheds, and/or market

[10] Deep wells are installed in areas that are not yet served by existing local water systems, and generators are provided for temporary power utilities.

TABLE 16.2 **Cost/benefit ratio: selected NHA resettlement projects**

Cost item	Manila: in-city (multistory)	Caloocan: in-city (H&L)	Trece, Cavite: off-city (H&L)
Total project cost/unit	917,640	203,346	146,942
H+L cost/unit	841,667	172,107	113,183
School building (for 1,000 units; 15 classrooms, 3-story)	—	23,000	23,000
Total investment cost	917,640	226,346	169,942
Other costs			
Interest subsidy on interest-free housing loan component, PV 30 years	151,974	34,715	34,715
O&M, 5% of investment cost for multistory; 1% for H&L, PV 30 years, 8%	9,176	2,263	1,699
PV O&M	516,530	25,482	19,132
MOOE and PS of school = ₱1409/year, PV 30 years, 8%		15,862	15,862
Livelihood program, noninfrastructure (one time)		3,000	3,000
Total interest and operating subsidy	668,504	79,059	72,709
Total cost, PV	1,586,144	305,405	242,651
Market rent, PV 30 years; 8%	2,081,667	493,667	141,392
CBR: total cost/market rent	0.76	0.62	1.72
CBR: total investment cost/market rent	0.44	0.46	1.20
CBR: total cost/market rent (assumes gov't owns land at end of 30-year period)	0.47	0.32	0.81

NOTE: Costs are in ₱ per unit or household; H&L = house and lot; MOE = maintenance and operational expenses; O&M = operating and maintenance; PS = personnel services; PV = present value. Investment costs for different construction years adjusted for comparability based on inflation. Market rent is based on average rental rates from the 2009 Family Income and Expenditure Survey and rental rate index of 0.08 in Metro Manila and 0.04 in provinces. Land values increase using regional consumer price index.

talipapa centers.[11] On the other hand, for skills training, the NHA allocates about ₱3,000 per beneficiary household to link the resettled communities to skills training, job placement, scholarship programs, and livelihood-based projects, including credit or loan assistance from other national government agencies. However, the NHA does not monitor whether these trainings and facilitation activities result in actual employment or livelihood.

In terms of benefits, while off-city projects are designed as house-and-lot packages, the value of the housing units measured in terms of housing rental value is much less than the in-city housing, especially since the sites are often located in third or fourth income-class municipalities.

The results of the CBR analysis show that in-city developments, specifically resettlements in Metro Manila, are more cost-effective. It costs the government less than one peso (between ₱0.76 and ₱0.62) to produce one peso of housing benefits in Metro Manila. In off-city sites, the cost exceeds the benefit, based on a CBR of ₱1.72. This implies that in the long term, the return on investment is negative. The benefit is higher for in-city housing because of the higher economic value of the property after development. Off-city locations have a lower rental value, because these sites are usually outside the city or town centers, and in lower-income municipalities. The cost effectiveness of the projects is thus affected by the economic potential of the area as well as the value that beneficiaries attach to the resettlement housing.

Considering that after the 30-year lifespan of the housing unit, the NHA usually retains the land, given the low loan repayment performance of beneficiaries, the value of the retained asset becomes part of the benefit from the investment. Land values are assumed to increase over time, thus the CBR is lower when the value of the land is considered.

The Socioeconomic Impact of In-City versus Off-City Resettlement

The results of the welfare analysis show that off-city relocation distances people from livelihood, and pushes them into poverty. The income of the off-city relocated households is lower by about ₱3,000 after adjusting for cost-of-living differences (table 16.3). The reduced expenditure on basic needs (food, water, electricity) of households in off-city resettlement implies deepening food insecurity, and could be a coping mechanism to deal with a reduction in income and a change in the nature of employment.

On the other hand, health expenditures appear to have increased, although the result is not statistically significant. The lower transportation costs may reflect changes in employment. While there is no significant difference in the proportion of employed households, there are significantly more women employed in the off-city relocation sites. This change indicates that women may have taken on domestic jobs, or livelihood projects (e.g.,

[11] These are informal wet markets housed on temporary structures made of mixed materials (wood and salvage materials).

TABLE 16.3 **Results of socioeconomic analysis (based on small-sample analysis)**

Variable	Off-city	In-city	Difference (off-city–in-city)	p value	Significance
Average monthly household income	7,456.4586	10,602.9494	−3,146.4908	0.000	***
Average household expenditure	5,882.0586	9,972.5556	−4,090.4971	0.000	***
Average household savings	1,574.4000	630.3937	944.0063	0.224	n.s
Average household expenditure: food	3,336.4924	5,064.2564	−1,727.7641	0.000	***
Average household expenditure: transportation	709.2988	793.3841	−84.0852	0.530	n.s
Average household expenditure: water	206.4023	332.3209	−125.9186	0.000	***
Average household expenditure: electricity	387.6078	705.6084	−318.0006	0.000	***
Average household expenditure: medicine	202.2216	125.1561	77.0655	0.324	n.s
Proportion of household members (age 6–22) attending school	0.2758	0.3482	−0.0724	0.038	**
Proportion of employed household members	0.3716	0.3312	0.0403	0.185	n.s
Proportion of employed household members: men	0.4728	0.5537	−0.0808	0.158	n.s
Proportion of employed household members: women	0.3415	0.2248	0.1167	0.028	**

SOURCE: Socioeconomic Survey of Resettled Families Social Impact Monitoring Project, World Bank.

NOTE: n.s. = not significant; *** significant at alpha = 1%; ** significant at alpha = 5%; * significant at alpha = 10%. Adjusted for cost-of-living differences using the Consumer Price Index. All expenditure values are in ₱. Used propensity matching to come up with comparable households; sample size = 163 matched households.

sari-sari stores), while the men are still looking for jobs in nearby areas. There are cases where the employed family member rents space in other informal settlements in the city and goes home to the family only during weekends or holidays.

A disruption in education was also noted in the off-city sites. The proportion of school-age children (6–22 years old) has dropped, despite the construction of new schools in the area. One possible explanation is that the new schools may not adequately serve the students, since the resettlement schools are considered satellite classes of regular Department of Education employees, and there are several cases of a reported absence of teachers in the area. It may take time for the Department of Education to hire additional teachers for these new schools, and given the already high student/teacher ratio in the public schools, the additional load for teachers further compromises the quality of education. The results also indicate dissipation in community social capital, from high community involvement to lower interest in participating.

This has also been observed in the case of health centers. A community volunteer usually stays in these clinics, mainly to dispense over-the-counter medicines.

CONCLUSIONS AND RECOMMENDATIONS

Both off-city and in-city resettlements are meant to improve the housing conditions of resettled families, and both should be welfare-enhancing. However, this study has shown that badly planned off-city resettlements are costly, and actually reduce the well-being of resettled families. Compared to in-city resettlement, government resources are not used efficiently in off-city resettlement. In the long term, the government may even have to spend more for off-city resettlement, since the resource requirements in terms of social and physical infrastructures tend to rise exponentially the greater the distance to the relocation site from the original site, or from the city proper. Moreover, the adverse impact on welfare refers not only to reduced income, but also to lower participation of school-age children in schooling, and increases the responsibility of women to earn needed income for the family.

Off-city resettlements are often hastily undertaken, and are located in marginalized areas, far from livelihood and employment facilities. While affected families are provided with houses and lots, there are compromises in the quality of the developments. These areas are also far from sources of water and power; thus resettlement sites are often deficient in basic services. In other words, the objective of expediency rather than efficiency and effectiveness has dominated the choice in the construction of these resettlement sites, and has adversely affected the welfare of resettled families in off-city sites.

A policy that advocates for in-city resettlement is far superior to off-city resettlement. The initial investment is high, but the socioeconomic outcomes are better. Problems of poor maintenance of multistory housing can be addressed through better estate management, while land costs can be minimized through lease arrangements or rental housing options. Income-based

subsidies for the costs of the housing units should also be explored by the government.

In cases where off-city resettlements are inevitable, the government must improve the choice of resettlement sites, and explore an incremental housing strategy in which the households and communities are more involved in the construction of the housing and in community development.

The results of this case study have revealed policy shortfalls, and also how the policy—and compliance with the policy—needs to be strengthened. This study also highlights the need for an inclusive urbanization process, so that negative externalities can be minimized and the marginalized people are not made to suffer the impacts of development.

There is also much to learn from a social impact assessment of resettlement projects, specifically how households, and/or the community and the government, can undertake the reconstruction process in such a way as to restore economic and social well-being. The NHA has resettled more than 100,000 families in infrastructure development projects in the last 10 years alone. Involuntary resettlements are expected to continue under the current administration's accelerated infrastructure spending in the medium term.

This case study supports other, qualitative studies that were undertaken to assess the socioeconomic conditions of households in various NHA resettlement sites. Impact evaluation of housing programs in the country has been constrained by data limitations: thus, many of the assessments of resettlement housing have dealt mostly with implementation issues, activities, and inputs. The current approach provides more rigorous analysis of resettlement impact. While it was based on small-sample estimates, the methodology can be applied to larger samples for more robust results.

REFERENCES

Ballesteros, M., and J. Egana. 2013. "Efficiency and Effectiveness Review of the National Housing Authority Resettlement Program." PIDS Discussion Paper 2013-28. Philippine Institute for Development Studies, Manila.

Olsen, Edgar. 2000. "The Cost Effectiveness of Alternative Methods of Delivering Housing Subsidies." Working Paper, Department of Economics, University of Virginia, Charlottesville.

World Bank. 2016. "The Philippines Urbanization: Harnessing the Benefits of Urbanization for Growth and Poverty Reduction." Paper presented at World Bank Urbanization Workshop, January 21, Manila.

Chapter 17

Livelihoods in Development Displacement – A Reality Check from the Evaluation Record in Asia

Susanna Price

Abstract. *Development, widely considered a solution to long-term population displacement, can paradoxically create more displacement. This chapter explores this paradox through the lens of evaluation studies. Early evaluation studies identified a gap between country laws, which positioned development displacement and resettlement as a subset of property and expropriation laws, and international policy, which centralized livelihood measures, living standards, and outcomes for people affected. The chapter explores the international policy conceptualization of livelihoods as embedded in a sociocultural context, requiring strategies to recreate livelihoods, monitoring and evaluation (M&E) and their results in terms of livelihood outcomes. It compares international policy perspective and evaluation outcomes with selected evolving Asian country safeguard systems, to examine the extent to which livelihoods are addressed and evaluated. The gap between international and national standards is narrowing, but livelihood measures form the weakest point in many laws concerning land takings. Differences in time frames, focus, mandates, and resources in project preparation and implementation reflect these divergent objectives. Methods for assessing livelihood*

Susanna Price, Australian National University, susanna.price@anu.edu.au.

risk, planning livelihood support, and for M&E of livelihood outcomes, are rare in country frameworks. Some approaches that may provide a way forward in building the knowledge base on livelihood success and sustainability through evaluation at the country level are presented.

Globally, the number of people forcibly displaced due to conflicts and disasters has escalated to record-breaking levels and protracted time frames, renewing the pressing call for longer-term solutions that foster sustainable livelihood creation. Some experts expect that development may offer such a solution to forced displacement (UNDP 2013). The question is, will development itself swell the number of displaced people lacking livelihoods, and thus only add to the problem?

This chapter approaches this question by first reviewing the case being made for livelihood creation to mitigate the costs of displacement generally. It then explores livelihoods in development displacement more specifically, through the lens of the evaluation record, to understand the key elements shaping livelihood loss and potential reconstruction. Evaluations find that people affected by development may lose their income and livelihoods along with their housing, or independently of housing. The case for considering livelihood as an essential and critical requirement in longer-term solutions to displacement is examined. The chapter also examines the impact that country laws, procedures, practices, and capabilities for land taking and transfer—that is, the "country framework" or "country system"—have on development displacement outcomes, and on livelihoods in particular.

This analysis serves to highlight a gap, identified at the earliest stages by international resettlement specialists, between international policy and borrower country frameworks. Lost income and livelihoods in particular fell into the gap. The World Bank's policy was based on an understanding of the often complex sociological processes through which displacement could damage incomes and livelihoods, and living standards generally, through which livelihoods could, eventually, be recreated. In contrast, in 1991 the World Bank's general counsel found that, among borrowers, legal issues in resettlement were treated as a subset of property and expropriation law, that basically aimed to clear land for development purposes in return for cash payments to recognized owners (World Bank 1994). In these circumstances the actions and recordkeeping focused on transfer and status of the land rather than the Resettlement Plan, its livelihood measures, and its socioeconomic outcomes for land losers. Noting that cash would not suffice to prevent impoverishment where land and labor markets and safety nets, were undeveloped, and where compensation funds risked diversion, siphoning off, or delay, the World Bank recommended "policy reform" and other actions to address the gap and, in effect to bring the people affected into sharper focus (World Bank 1994). In other words, project proponents would be encouraged to move beyond cash compensation to take responsibility for regenerated income flows, livelihoods, and living standards among people dispossessed by the developer's own projects (World Bank 2004).

This gap is still very evident. International lender policies assign, in various circumstances, responsibility to the governments that expropriate or restrict access to land involuntarily; to "clients" implementing and operating the project (IFC 2012); or to the borrower, defined as the "recipient of Bank financing for an investment project, and any other entity responsible for the implementation of the project" (World Bank 2017, 3). Human rights standards require "competent authorities" to ensure that anyone forcibly displaced by development has access to livelihood (UNHRC 2007).

By centralizing a livelihood objective, international lenders such as the World Bank and the International Finance Corporation (IFC), raise the question of socioeconomic rehabilitation, which may require a deeper understanding of sociocultural patterns of community interactions in relationship to land and resources. This entails assessing risks and impacts for various categories of affected people who may use land and other assets differently, and, in consultation with all stakeholders, developing income and livelihood options that meet social and economic parameters. It may mean detailed assessment of loss of income; and strategies to replace or to cost the reestablishment of working agricultural and commercial enterprises. It may mean matching skill sets, diagnosing training needs, and mobilizing social security or welfare provisions. It may mean formalizing and costing time-bound measures and plans; and monitoring outcomes for people's livelihoods and lives.

Conversely, payment of compensation in cash can seem a simpler and easier option that absolves project sponsors of responsibility for any further remediation, and readily hands the decision on the use of compensation to project-affected people, who may opt to replace the lost asset, if they can; look for alternative income sources; consume the proceeds; mix several of these options; or do something else. That strategy assigns the livelihood risks that arise in displacement squarely to the affected people. It also raises the possibility of misdirection; siphoning off or delay of cash compensation funds intended for affected people; and hardship among people affected where countries lack social welfare. Many officials among borrowers in Asia nonetheless express a preference for this approach.

This chapter will explore the dynamics at work in some of the differing expectations around this central theme, and offer a perspective based on evaluation studies, discussions and interviews conducted over many years. International standards, whether originating from a focus on human rights or on international lending policies on involuntary resettlement, recognize the importance of the country role in policy implementation. Across Asia, country laws, procedures, and practices for land takings are changing. Is there now greater recognition of the importance of livelihoods in crafting sustainable solutions for long-term displacement? The chapter explores this question and concludes with some relevant recommendations.

THE CASE FOR ADDRESSING LIVELIHOODS IN DEVELOPMENT DISPLACEMENT

On a conceptual level, the case for addressing livelihoods in the context of forced displacement is multifaceted. The 193 countries of the United Nations

(UN) General Assembly adopted the UN's 2030 Sustainable Development Goals (SDGs) in 2015. This has revived the attention given to sustainable livelihoods, which "resonate" with all 17 of the SDGs, and underpin the realization of the SDG targets, particularly those that aim to end poverty and hunger; achieve sustainable growth; reduce inequalities; promote decent work for all; and use the earth's resources in a sustainable manner (Biggs et al. 2015). The SDGs encompass all countries, both developed and developing, and include developing a plan for monitoring and evaluation (M&E) with indicators that work at the regional, national, global, and thematic levels: and evaluation plans that must include social, environmental, and economic indicators. These factors raise the stakes for a globally adopted and evaluated development plan, and they have brought renewed attention to the approaches for developing sustainable livelihoods (Biggs et al. 2015).

The SDGs foresee population displacement as a major risk to achieving sustainable development—and this has implications for livelihoods. As population displacement arising from conflicts and disasters breaks records and becomes more protracted, the challenge of finding longer-term solutions intensifies. Livelihood creation may be viewed as an essential component of longer-term solutions, especially for those displaced who have the least negotiating power and the fewest skills to access "new" economic opportunities. The United Nations Development Programme (UNDP) estimates that, globally, more than 200 million people are unemployed, with 74 million young people aged 15–24 looking for work. Some 600 million new jobs will be needed in the coming decade, without which UNDP expects the risk of further destabilization and intensification of population displacement. UNDP's resilience-based development approach, for example, builds livelihoods for both displaced people and their hosts (UNDP 2013). The SDGs present development both as a means of preventing further displacement by diminishing its drivers, and as a solution to protracted displacement, by turning short-term refugee costs into longer-term gains; lowering the costs of migration; and increasing the contribution of migrants to their host countries or communities through building livelihoods (UNDP 2013).

Development, which is the intended solution to displacement, paradoxically creates more displacement—at least 15 million displaced persons each year (IDMC 2016). Building infrastructure, for example, has social and environmental impacts and externalities that, if not properly managed, can result in unmitigated displacement that further disrupts livelihoods. This chapter explores this paradox through the lens of evaluation studies and of emerging new directions in laws, regulations, and procedures on compulsory acquisition.

LIVELIHOODS IN DEVELOPMENT DISPLACEMENT: SOME THOUGHTS FROM EVALUATIONS

The involuntary resettlement policies of international financial institutions place livelihoods at the center of resettlement objectives, defining livelihoods broadly, for example, as "the full range of means that individuals, families, and communities utilize to make a living, such as wage-based income, agriculture, fishing, foraging, other natural resource-based livelihoods, petty trade, and

bartering" (IFC 2012, 1). The World Bank definition, as set out in the new Environmental and Social Framework, is similar (World Bank 2017).

Livelihood replacement or recreation can be complex. For example, based on significant project experience, the IFC prefers providing replacement land where livelihoods are land-based, or where land is collectively owned (IFC 2012); and access to alternative resources where livelihoods are resource-based, together with resources for their preparation and development. Those people losing income from lost or damaged commercial activities may be eligible for compensation for reestablishing the commercial activities elsewhere; compensation for lost net income during the period of transition, replacement land, and for the costs for relocating and re-establishing plant, equipment, and other items (IFC 2012). These efforts may entail, in addition to replacement or replacement-rate compensation for income-generating assets and income loss, measures to reestablish investment and development assistance such as the land preparation, credit facilities, training, or job opportunities needed to enable affected people to improve their living standards, income-earning capacity, and production levels; or at least to maintain them at preproject levels.

Lenders may require "gap analyses," which record any differences between their own policy positions and the country frameworks, together with the supplementary gap-filling measures that lenders may agree on with borrowers in each project case. Livelihood gaps reflect the difference between international policies and country laws and standards on land takings (World Bank 2014). Engaging with these income and livelihood issues raises questions on the availability of resources and time for planning; for identifying those specifically at risk of losing income and livelihoods; for assessing the compensation amounts due in different circumstances; for canvassing feasible livelihood-supporting options and opportunities or social welfare possibilities, in close consultation with affected people; and for monitoring and evaluating impoverishment risks and livelihood outcomes.

Development displacement is conducted in a very specific context and time frame, with legal, valuation, financing, consultation, disclosure, and appeals dimensions. An early, internationally financed resettlement evaluation found that favorable country policy and legal frameworks, together with sufficient financing, capable institutions, and local involvement, were the foundation of successful livelihood restoration, which from the beginning was held to underpin successful resettlement (World Bank 1994). An Asian Development Bank evaluation in 2000 found similar results, and also recommended more attention to livelihood risk assessment and restoration, together with stronger M&E of outcomes. A subsequent evaluation by the World Bank confirmed the importance of livelihood reconstruction: a careful assessment of impacts found that more people lost income and livelihood than were physically displaced from their housing for the sampled World Bank Group projects.[1]

[1] An evaluation of operations over the period fiscal 1999–2008 found that 41 percent of people affected were physically displaced; the rest faced impacts on livelihoods (IEG 2011).

While evaluations generally have confirmed the importance of conducive country frameworks, it has been difficult to foster such frameworks. Successive evaluations found that international involuntary resettlement policies, when applied in loan financing, offer better risk assessment and more comprehensive corresponding mitigation plans to address income and livelihood loss than do country frameworks generally. International policies offer fairer compensation and other assistance for nontitled landowners; and broaden monetary compensation to include loss of income and measures that aim to restore livelihoods taking account of sociocultural context (ADB 2000; IEG 2011; World Bank 1994, 2014). However, the extent to which these additions in planning are carried through into implementation, and reflected in monitoring and management depends at least partly upon the level of congruence in borrower frameworks, and more generally, borrower commitment to these ideas.

In densely populated Bangladesh, for example, land is a particularly critical asset in social and cultural as well as economic terms. Even though land in Bangladesh is not only a means of livelihood but also "a sign of social power, pride, status, security and happiness" (Al Atahar 2013, 306) the government has not approved a national resettlement policy that would recognize and address the wider implications of the substantial losses that are experienced when land is acquired for development purposes, including loss of livelihood, and its interrelationship with these wider social variables. In 2016 the country's Ministry of Lands approved a new Land Acquisition Act, but it does little to address the wider concerns of landowners and land users, beyond speeding up the acquisition steps and raising compensation levels, typically paid in cash with no additional assistance (Zaman and Khatun 2017). India, in contrast, has made a significant effort to address resettlement, rehabilitation, and livelihood issues in its new law, as discussed below.

It is worth pausing to recap the underlying legal powers in forced development displacement. Legal instruments for land expropriation or transfer have power to trigger the displacement. If upfront negotiations fail between willing buyer and willing seller, or are deemed inappropriate, it is the state's exercise of eminent domain or compulsory acquisition that provides a legal foundation for it to expropriate, in the public interest, the property of individuals for development purposes. This overrides, in most cases, their constitutional rights to property, whether it is their "property" by legally verified ownership, or by use rights. Most country constitutions allow expropriation, or compulsory acquisition, upon payment of "just terms," "equitable compensation," or a similar phrase, to citizens for the loss of their property. If, rather than acquiring the land the project simply restricts access to it, or activity upon it, as with a power transmission line with tower footings, or a fragile environmental area that is being protected, other laws, regulations, and guidelines may apply. The law and any associated regulations generally determine who is eligible for compensation and other assistance, and for which kinds of losses. As a subset of property and expropriation law, without a livelihood or rehabilitation objective, in most cases these laws fail to recognize the full extent of losses and what might be required to address them.

Several countries have requested that financiers allow them to use their own country safeguards rather than financier policies. Early signs are that these country safeguard analyses focus closely on the wording of legal instruments for compulsory acquisition, and their application in practice. Other parameters also deserve careful attention in such assessments. These include time frame and planning cycle constraints, valuation methods, and grievance and appeals mechanisms that provide a fair outcome, as explored in the following sections. These parameters are discussed sequentially in what follows.

Time Frame and Planning Cycle Constraints

Development displacement accompanies the project-planning cycles of feasibility, design, appraisal, approval, and implementation. If resettlement planning is required, it is situated in a specific, and often a very tight time frame. While scoping and socioeconomic surveys can and must begin earlier, compensation and resettlement plans often cannot be finalized until completion of the detailed technical design that will allow the assessment of impacts, through a census and asset inventory. The risks to livelihood must be assessed quickly, and those at risk of losing income and livelihoods must have choices put before them. Compensation must then be delivered before construction begins, forestalling arbitrary eviction without compensation. This tight time frame favors quick cash payments. It presents a challenge for livelihood measures, which may involve longer-term activities such as training, project employment, production, business development, and various forms of social assistance. This time dimension has been little noted or explored in international financier evaluation studies. Nor have the variations in planning cycles between countries and sectors received much attention.

Valuation Methodology—Replacement Costs, Including Social Costs

The valuation methodology determines the ultimate value of compensation offered. Country laws, which had used lower asset tax values, and various administrative formulas, with lesser compensation rates for less certain categories of land holdings, are now moving in some cases to independently set market rates. International practice uses replacement rate, which adds to the market rate the additional real costs to people affected by involuntary acquisition, such as administrative and transaction costs and relevant moving and transfer costs (Pearce 1999). If the payment is delayed, inflation may erode the potential for the compensation to replace lost assets.

Replacement of losses is the overall principle that international financiers advocate for—and also that replacement land must be offered to land-dependent rural producers. There are some very important questions to consider here. Is fair-market appraisal used as a basis for asset valuation? Is the highest and best price regularly used, based on accurate data? In addition, are out-of-pocket expenses covered? And how are nontangibles valued?

The loss of land may represent not only lost assets and income, but also the loss of security for old age and disability, and loss of the locus of social networks. Land is often an essential element in the formation of households, social and cultural systems, and psychological well-being. Replacing land with cash or short-term work may destroy these systems, and with it prospects for sustainable livelihoods (Cernea 2008; Downing and Garcia Downing 2009).

The full social costs of displacement to affected people may only become apparent well after displacement as households, communities, and their production systems begin to unravel. These costs may include the loss of hard-to-quantify social networks that reflect and sustain communities, and offer both economic and social benefits, including health care, informal support, marketing networks, reciprocal labor exchanges, and backup in hard times. They also include the loss of any nonpriced social and cultural assets, such as commonly owned forests, water bodies, and grasslands, which provide communities with sacred sites and ritual objects, food, grazing land, fuel, medicines, and salable items. Land may also represent the only source of security in old age or infirmity. It may encapsulate and represent an unquantifiable but fundamental sense of belonging and identity that underpins psychological well-being. The loss of production systems and other assets may start unraveling the cohesion of households, neighborhoods, and communities—the social characteristics that underpin the inter- and intra-household agreements that in turn underpin livelihoods (Downing and Garcia-Downing 2009; Lam 2015). In short, in losing their tangible assets, in addition to income, households may lose essential subsistence, insurance, social support, and their place-based identity.

Consultation, Disclosure, Grievance Redress, and Appeals

Disclosure of critical information to the people affected, and the establishment of grievance redress mechanisms help to set a fair process, in which dispossessed people have access to information and can, without prejudice, lodge an appeal and expect to have it heard fairly and in a reasonable time frame. In these respects, development displacement takes place rather differently than most other forms of displacement—and with care, the mitigation strategies can be built in ahead of the act of displacement.

HOW ARE LIVELIHOODS ADDRESSED IN NATIONAL FRAMEWORKS OR SYSTEMS FOR EXPROPRIATION?

The rapidly developing Asian region features several significant new laws concerning land acquisition. This legislation increasingly reflects independent asset valuation, social analysis or social impact assessment (SIA), consultation opportunities with affected people, and negotiation as a basis for compensation. Most Asian countries now allow project-affected people to challenge land acquisition procedures in court—if they can afford to challenge. In addition to physical assets, many countries now recognize and compensate for loss of economic activity and improvements on land. At least three countries

(Cambodia, India, and Indonesia) offer land-for-land replacement options, but none of them require that the replacement land be ready for cultivation. India, Indonesia, and Vietnam offer relocation allowances where relocation is necessary (Tagliarino 2017).

India's 2013 law, The Right to Fair Compensation and Transparency in Land Acquisition, Resettlement and Rehabilitation Act (LARR),[2] replaced the Land Acquisition Act of 1894, which was based on the state's power of eminent domain. As signaled in its title, where it applies, the LARR accepts the proposition that land acquisition may threaten livelihoods; and that in such cases, livelihood rehabilitation strategies are necessary. It recognizes that "affected families" include those without title to land, who nonetheless have depended on the land for their primary livelihood for the preceding three years. Nontitled people losing livelihoods may benefit from livelihood reconstruction through increased compensation rates, SIA, and consent from the people affected. LARR includes consultative planning, negotiation, and grievance redress. The LARR is part of a broader legal framework of rights and guarantees that increase its legitimacy in protecting affected people (Mariotti 2015). India is the only Asian country with a legal requirement to minimize displacement by exploring alternatives. It has constituted a group of experts with social science expertise, to review SIAs and alternative project designs that would minimize displacement.

Indonesia recently introduced a landmark Law on Land Acquisition for Development Purposes in the Public Interest (Law 2/2012), which became effective in 2014 after the issuance of implementing regulations. The law replaces a series of presidential decrees and other regulations that generally required only the payment of lesser or no compensation to those land users without formal title; and calculated compensation starting at lower rates, based on tax value. Law 2/2012 introduces the concept of independent market appraisal for lost assets, requiring "reasonable and fair compensation" that covers land, assets on land, structures, plants, other objects relating to land, and other nonphysical appraisable losses, including loss of jobs, businesses, the costs of changing location or profession, and the loss of value in remaining assets. While not an explicit requirement, this could equal replacement cost. The law allows compensation to be paid as cash, replacement land, resettlement, shareholding, or other forms agreed upon between the parties (Article 36).

There is no mention of livelihoods in the law. In cases where the affected household is selecting the resettlement option, this may include livelihood assistance under a related law of 2011. Generally it is unclear how Law 2/2012 will deal with audit rules that require depreciation to be deducted from asset compensation; or with budgeting regulations that do not allow "double counting" in the form of additional livelihood measures on top of compensation. Law 2/2012 does include a greater opportunity for affected people to seek consultation, negotiation of compensation, and redress of

[2] The text of the LARR can be found here: http://www.prsindia.org/uploads/media/Land%20and%20R%20and%20R/LARR%20(2nd%20A)%20Bill,%202015.pdf.

grievances; and these mechanisms may offer opportunities to counter earlier regulations that in effect limited livelihood assistance.

Vietnam's new Constitution (2014) and Land Law No 43 (2013) strengthen legal protections for people affected by development displacement, with provisions to identify them, inform them, and consult with them prior to any acquisition; to recognize certain customary land-tenure rights; to recognize and compensate for loss of economic activity on land; to provide replacement land as a compensation option; to compensate before possession; and to pay a relocation allowance when people must relocate.

In 2102, Cambodia introduced a Law on Land Expropriation that strengthens information and consultation requirements, allows compensation for loss of economic activity, and encourages land-for-land compensation, but does not require that replacement land to be ready for cultivation. Sri Lanka introduced an authoritative but nonbinding National Involuntary Resettlement Policy in 2001, a Compensation Policy in 2008, and several gazette notifications under the Land Acquisition Act of 1950, the latest of which (2013) which applies very selectively, mainly to certain transport projects; brings compensation payments to replacement rates; and significantly boosts consultation and negotiation possibilities for people losing land.

The Kyrgyz Republic, like some other Central Asian republics, has transformed its legal and regulatory framework to allow privately owned land for its citizens. However, the legal basis for compulsory acquisition lacks clarity, and there is little commitment to consultation with the people affected in practice, or to the development of livelihood programs.

China has introduced measures in the reservoir sector that address livelihoods (Cernea 2016), and has several other laws and regulations that provide some assistance for livelihood that covers expropriation in certain circumstances. It still lacks an overall, transparent and consistent law on land acquisition that covers all sectors; unambiguously sets a livelihood objective; and requires a resettlement plan and SIA as a basis for developing livelihood options.

While a complete analysis is beyond the scope of this chapter, there are positive examples of the movement toward fairer, more consultative, and more transparent land acquisition laws and regulations across Asia. However, the legal requirements generally stop short of statements in law that define livelihood standards; propose risk assessment tools and methods to determine when livelihoods are at risk; formulate income and livelihood measures in meaningful consultation with those affected; mobilize the necessary expertise, management capacity, and financing; and establish requirements to evaluate whether livelihood objectives have been achieved. The laws do not take the additional step of recognizing the rationale for livelihood measures: the unquantifiable social costs, opportunities foregone during downtime, and the transition and reconstruction costs that so often accrue for people along the way.

Nationally ratified human rights conventions, declarations, and treaties would, if applied in tangible ways within the process of planning and managing resettlement, strengthen consultations and protection for a range of vulnerable groups.

NEGOTIATING FOR LIVELIHOODS

Many new laws and regulations offer increased scope for negotiation, even within the context of compulsory acquisition. Does the negotiation process hold something that would foster income and livelihood choices for people affected? This section briefly explores this possibility.

Negotiation may take place around choices within a framework for involuntary land acquisition, or entirely outside the existing framework, as an agreement negotiated between a willing buyer and a willing seller, in the form of a market transaction. The former type of negotiation is subject to the applicable national legal framework, which has the legal power to involuntarily displace people, even while offering them an opportunity to negotiate on some elements.

Most Asian countries require such negotiation: the exceptions are Bangladesh, Taiwan, and Thailand (Tagliarino 2017); and Sri Lanka, except for selected projects named in the 2013 Gazette. Most reports indicate that such negotiations revolve around the level of cash compensation in terms of the assets recognized, and the level of compensation offered, with little scope for negotiation on the reconstruction of livelihood opportunities.

Even within this involuntary framework, however, the laws increasingly provide opportunities for negotiation. Among Asian countries, Indonesia, for example, in its new Land Acquisition Law of 2/2012 (Articles 34, 37, and 38), introduced a specific requirement for negotiation of compensation between the land administrator and the "entitled parties," with the express intention that the acquisition is not carried out entirely under duress. Before the possibility of expropriation for development in the public interest, for example, Law 2/2012 requires the government to consider other options, including buying land under a "willing buyer-willing seller" transaction. If the land parcel is less than five hectares, Presidential Regulation 40 of 2014 permits the acquiring agency to negotiate directly with the land user in the form of sale-purchase land exchange, or other means agreed on by the parties. Beyond this, however, the law envisages a negotiation with affected people with regard to compensation options, including money, replacement land, resettlement, shareholding, or other forms of compensation, as agreed between the parties (Article 36). All of these options hold possibilities for restoring livelihoods, but nothing is explicitly spelled out.

The second type of negotiation takes land transactions into the market arena, which may be shrouded by commercial in-confidence concerns. There are some indications that private sector models can be more flexible, expansive, and responsive to the articulated concerns of affected people when compared to models applied by government agencies. Private developers can work outside the government planning, project cycle, and budgeting systems that may limit the options available for government projects, particularly where livelihood measures are not mandated by law. For example, a private oil and gas project developer in Indonesia provided a resettlement plan for the affected communities. The plan was considerably more generous than a comparable government-funded project would have been (Price 2015). The plan articulated a resettlement-with-development objective that was higher

than either the international or national standards in place at the time. It also required a high level of public scrutiny and disclosure, and strong corporate commitment.

Negotiating directly with affected people as an integral part of a framework for land acquisition, compensation, and involuntary resettlement increases the transparency of the process, possibly leading to fewer complaints (Tagliarino 2017). However, a "willing buyer, willing seller" arrangement might not necessarily mean a level playing field between buyer and seller that will result in a fair outcome. Rather, it can signal the influence of asymmetries in power and information. It may reflect a loss of entitlements to fair treatment designed to forestall the impoverishment of affected people, in favor of a nebulous system of negotiated rules and remedies that in effect do away with entitlements altogether (Bugalski 2016).

Private developers and proponents may simply rely on government to clear the land with no questions asked; or to waive any requirements for environmental and social safeguards, including even basic compensation. This has been the case with certain notorious mining ventures, such as the Freeport mine in Indonesian Papua; or, more recently, in agriculture, as in some of the Cambodian Economic Land Concessions. Asymmetries in status, resources, power, and information are key features of interactions between the negotiating parties. If developers have financing from international banks with involuntary resettlement policies, or have signed onto voluntary agreements such as the Equator Principles, this may offer additional protection to the affected people. However, land deals negotiated between unequal parties may still result in divided and disempowered land-owning groups, and their resulting marginalization and impoverishment, even where both international and human rights standards are respected (Narula 2013). New guides are being written with the aim of informing and supporting small landholders faced with the daunting prospect of negotiations with powerful, well-resourced developers who often have government backing (e.g., a community guide to negotiation issued by Inclusive Development 2016).

RECOMMENDATIONS FOR ENHANCING LIVELIHOODS IN DEVELOPMENT DISPLACEMENT

> ...in order to have continued relevance and application, livelihoods perspectives must address more searchingly and concretely questions across ...four themes:...knowledge, politics, scale and dynamics. These are challenging agendas, both intellectually and practically. For those convinced that livelihoods perspectives must remain central to development, this is a wake-up call. The vibrant and energetic "community of practice" of the late 1990s has taken its eye off the ball. A certain complacency, fuelled by generous funding flows, a comfortable localism and organisational inertia has meant that some of the big, emerging issues of rapid globalisation, disruptive environmental change and fundamental shifts in rural economies have not been addressed. Innovative thinking and practical experimentation has not yet reshaped livelihood perspectives to meet these challenges in radically new ways. (Scoones 2009)

Changing laws and practices are creating new ways of addressing losses in incomes and livelihoods across Asia. Some recommendations are presented below.

- Include explicit livelihood objectives in the country's property and expropriation legal instruments, or introduce a new law that recognizes the need for livelihood rehabilitation, as India has done; as well as supporting guidelines and procedures.
- Mobilize resources and capacities to address livelihood objectives through the development of appropriate management mandates, staffing skills, and feasibility assessments and arrangements.
- Develop definitions and diagnostic tools and strategies to assess quickly and effectively whether livelihoods are at risk among the people affected.
- Test principles, valuation methods, and forms of compensation and other assistance that may help rebuild livelihoods. For example:
    - The *value to the owner compensation principle*, which is made up of market value together with other losses suffered by the claimant; and
    - Payment of other *consequential financial losses*, such as the cost of finding alternative accommodations; extra costs for living in a new district; fees for discharging mortgages; temporary business losses pending removal; loss of business goodwill; and the costs of notifying customers and clients about the removal, and other related losses.
- Recognize the principle that it *is an interest in land* that is actually acquired: this comes close to recognizing nontitled people.
- Proactively test baseline socioeconomic surveys as a basis for subsequent M&E, by project sponsors, developing effective feedback links to enhance livelihood program outcomes.
- Test more sensitive needs assessments as a basis for developing for livelihood programs, taking into account the needs and priorities of different groups.
- Research the question of which types of compensation contribute more effectively to restore lost income and livelihoods.
- Explore and test a wider range of feasible livelihood options. For example:
    - Benefit sharing, such as in-kind assistance, project employment, and related on-the-job training, direct revenue sharing, development funds, links with employers' and government programs and equity sharing.
    - Other options include assistance for business development, access to credit and other services, and other forms of training and skills development. These forms of assistance may be reasonably and readily deployed by the average project sponsor or proponent, whether in the public or private sector, although

without proper needs assessments, targeting, and monitoring of results, or without the support of experienced livelihood practitioners, the measures offered to affected people may be poorly utilized and quickly abandoned.

- Certain other options require the waiving of local taxes and preferential rates for financing livelihood reconstruction. This requires concurrence with revenue-raising bodies that might be outside the land acquisition framework. Finally, putting into place social safety nets through pensions, project insurance, contingency funds, vulnerability support schemes, and/or project special funds may be a logical way to proceed, but may also raise practical problems of specifically targeting the affected people. Safety nets usually operate through national-level coordination, again necessitating concurrence from national-level agencies.
- Consider accessing land on a lease basis so that the land is not lost in perpetuity to the original owners.
- Build the knowledge base of what works effectively and how, through the development of M&E methods of measuring livelihood outcomes.
- Share results from the development displacement experience with livelihood programs more generally.

CONCLUSION

This chapter argues that finding livelihood solutions is an urgent matter as development, conflict, disasters, and, increasingly, environmental change displace ever more people around the globe. Of all these forms of displacement, development displacement has a long track record in addressing livelihood issues, and in prioritizing international policies and standards of livelihood improvement, or at least restoration. The earliest evaluation of involuntary resettlement highlighted the sociocultural context of livelihoods and the links between livelihood restoration and the overall objective of poverty reduction (World Bank 1994). Despite its importance, however, livelihood outcomes have routinely suffered not just from the absence of systematic data and analysis, but also from lack of visibility in the form of articulated objectives in legal instrument, and methods for assessing losses. This has meant a corresponding lack of resources for planning, monitoring, and evaluation of livelihood options; a lack of sufficient time allocated to planning cycles for significant and meaningful consultation on a wide range of choices as a basis for preparing livelihood programs; and deficiencies in asset valuation methods for addressing loss of income and the reconstruction of livelihoods.

Compensation at replacement cost forms a critical basis of the overall strategy to rebuild livelihoods in international policy formulations: but resettlement specialists have found that more is required if livelihoods are severely affected. While there is increasing congruence between international and national standards in legal and regulatory instruments, livelihood measures form the weakest point in many national laws concerning land acquisition.

Whereas the livelihood objective is central to international resettlement standards, it barely appears, if at all, in most of the recent national legal and regulatory initiatives.

Yet even where internationally financed projects include livelihood measures in resettlement plans, or as stand-alone livelihood restoration plans, there is no guarantee of their outcomes. This aspect of policy application merits a major rethinking in terms of rationale, legal formulations, financing, management, and application. It could benefit, for example, from a reexamination of methods for identifying the affected people who are most at risk and will need livelihood support, and meaningfully consulting with them on a range of feasible livelihood options. It could also benefit from a comparative assessment of valuing and compensating for lost assets and incomes as a basis for livelihood reconstruction, and the management arrangements under which these decisions are made. Similarly, a comparison of livelihood outcomes from the application of country laws that recognize and address the livelihood imperative, such as in India, compared with outcomes when country laws do not engage directly with livelihoods, could be illuminating. Is it possible to address these issues through negotiations with the people affected? Do these negotiations offer better livelihood protections, or simply reinforce existing patterns of social exclusion?

International involuntary resettlement policies offer much in terms of methods and procedures for defining livelihood scope and standards; in identifying those at risk of losing livelihoods; in formulating livelihood measures in close consultation with those at risk; in setting forth time-bound, costed plans for identifying management arrangements; in addressing the socioeconomics of recovery for livelihoods at risk; and in methods for measuring, monitoring, and evaluating outcomes. They also offer possibilities for a safeguard on negotiated settlements. Innovative country practices are now adding to the body of knowledge on these matters, reflecting better understanding of the underlying sociological parameters supporting livelihoods in many contexts. International human rights norms and standards can serve as additional tests of public interest through project hearings, while the concept of consent offers a different approach to defending livelihoods at risk. For land-dependent communities with little negotiating power, dismantling productive rural livelihoods may be a step too far, especially under legal and regulatory frameworks that do not recognize the need for livelihood reconstruction. Development may represent a strategy for longer-term solutions to loss of livelihood, but only if livelihood objectives are explicitly named in laws and negotiation procedures; addressed through specific risk identification methods; supported by a range of feasible and consultatively developed livelihood strategies; underpinned by fair legal and grievance mechanisms that are accessible to all those affected; and independently monitored and evaluated to ensure good outcomes.

REFERENCES

ADB (Asian Development Bank). 2000. "Special Evaluation Study on the Policy Impact of Involuntary Resettlement." ADB, Manila.

Al Atahar, Syed. 2013. "Development Project, Land Acquisition and Resettlement in Bangladesh; A Quest for a Well Formulated National Resettlement and Rehabilitation Policy." *International Journal of Humanities and Social Science* 3 (7).

Biggs, E., et al. 2015. "Sustainable Development and the Water-Energy-Food Nexus: A Perspective on Livelihoods." *Environmental Science & Policy* 54: 389–97.

Bugalski, N. 2016. "The Demise of Accountability at the World Bank?" *American University International Law Review* 31: 1–56.

Cernea, M. 2008. "Compensation and Investment in Resettlement: Theory, Practice, Pitfalls, and Needed Policy Reform. In *Can Compensation Prevent Impoverishment? Reforming Resettlement through Investments and Benefit-Sharing*, M.M. Cernea and H.M. Mathur, eds., 15–98. New Delhi: Oxford University Press.

———. 2016. "Foreword." In *Development-Induced Displacement in India and China: A Comparative Look at the Burdens of Growth*, Florence Padovani, ed. Lexington Books.

Downing, Theodore E., and Carmen Garcia-Downing. 2009. "Routine and Dissonant Cultures: A Theory about the Psycho-socio-cultural Disruptions of Involuntary Displacement and Ways to Mitigate Them without Inflicting Even More Damage." In *Development and Dispossession: The Anthropology of Displacement and Resettlement*, Anthony Oliver-Smith, ed., 225–53. Santa Fe: School for Advanced Research Press.

Inclusive Development. 2016. *Avoiding Forced Evictions: A Community Guide to Negotiation and Advocacy*. Phnom Penh: Inclusive Development.

IDMC (Internal Displacement Monitoring Centre). 2016. *GRID 2016: Global Report on Internal Displacement*. Geneva: IDMC.

IEG (Independent Evaluation Group). 2011. *Safeguards and Sustainability Policies in a Changing World: An Independent Evaluation of World Bank Group Experience*. Washington, DC: World Bank.

IFC (International Finance Corporation). 2012. "Performance Standard 5, Land Acquisition and Involuntary Resettlement." IFC, Washington, DC.

Lam, L.M. 2015. "Conservation-Led Displacement, Poverty and Cultural Survival: The Experiences of the Indigenous Rana Tharus Community in Far-Western Nepal." In *Responses to Displacement across Asia–Pacific: Strategies for Development, Disasters and Climate Change*, S. Price and J. Singer, eds., 205–20. Abingdon: Routledge.

Mariotti, Chiara. 2015. "'Something Old, Something New, Something Borrowed, Something Blue': A Critical Evaluation of the Newest Indian Land Acquisition, Resettlement and Rehabilitation Act (2013)." In *Responses to Displacement across Asia–Pacific: Strategies for Development, Disasters and Climate Change*, S. Price and J. Singer, eds. Abingdon: Routledge.

Narula, S. 2013. "The Global Land Rush: Markets, Rights, and the Politics of Food." Public Law and Legal Theory Working Paper No. 412, New York University School of Law, New York.

Pearce, D. 1999. "Methodological Issues Economic Analysis." In *The Economics of Involuntary Resettlement: Questions and Challenges*, M.M. Cernea, ed. Washington, DC: World Bank.

Price, S. 2015. "Is There a Global Safeguard for Development Displacement? In *Development-Induced Displacement and Resettlement: Revisiting the Knowns and Revealing the Unknowns*, I. Satiroglu and N. Choi, eds, 127–40. Abingdon: Routledge.

Scoones, I. 2009. "Livelihoods Perspectives and Rural Development." *Journal of Peasant Studies* 36 (1): 171–96.

Tagliarino, Nicholas, 2017. "Country-Level Legal Protection of Landholders Displaced by Expropriation: An Investigation of National Laws in 50 Countries across Asia, Africa, and Latin America Comply with International Standards on Resettlement and Rehabilitation." Draft paper. International Network for Displacement and Resettlement, Society for Applied Anthropology Annual Conference 2017.

UNHRC (United Nations Human Rights Council). 2007. "Report of the Special Rapporteur on Adequate Housing as a Component of the Right to an Adequate Standard of Living, Miloon Kothari." A/HRC/4/18.

UNDP (United Nations Development Programme). 2013. *Livelihoods and Economic Recovery in Disaster Situations*. New York: Bureau for Crisis Prevention and Recovery.

World Bank 1994. *Resettlement and Development: Bankwide Review of Projects Involving Involuntary Resettlement*. Washington, DC: World Bank.

———. 2004. *Involuntary Resettlement Sourcebook*. Washington, DC: World Bank.

———. 2014. *Involuntary Resettlement Policy Review, Parts I and II*. Washington, DC: World Bank.

———. 2017. *Environmental and Social Framework*. Washington, DC: World Bank.

Zaman, M., and H. Khatun, eds. 2017. *Development Induced Displacement and Resettlement in Bangladesh: Case Studies and Practices*. Nova Science Publishers.

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; Rachel Eager, Centre for Development Impact and Itad, 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.[1] 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

[1] 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).[2] 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."

[2] 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 (Nakhooda 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.

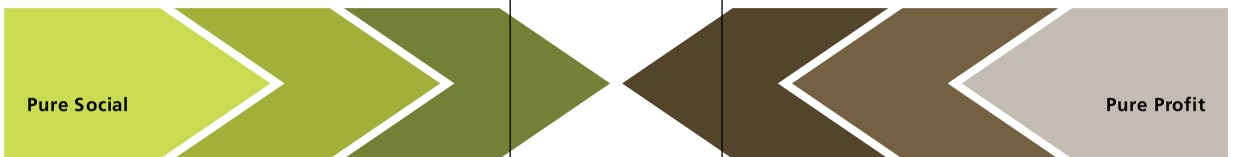
There is a broad range of approaches to socially or environmentally responsible capital, existing across a spectrum that ranges from philanthropic giving to traditional, profit-driven investment (Avantage Ventures 2011). As shown in figure 18.1, investment modalities can be broadly categorized into five 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.[3] 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

[3] 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**



Type	Philanthropic Donations	Venture Philanthropy	Program-Related Investing	**Impact Investing**	ESG Investments	Socially-Responsible Investments	Traditional Investments
Investment Approach	Pure charitable giving with no expectation of financial return	Donations are given as seed capital with the expectation of operational sustainability through mentoring by investors	Return of at least the principal is expected; possibility of market return, as well as of social performance	**Investments in companies whose primary goal is delivering social and environmental good, whilst also delivering competitive market returns**	Limiting investments to companies who track and evaluate their performance agaist key environment, social and governance metrics	"Do no harm" investing in listed stocks that avoid ethically and environmentally questionable companies (e.g. tobacco, coal, casinos)	100% driven by maximisation of short-term financial gain

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 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").[4] 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

[4] 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.[5]

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,

[5] 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-word 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.[6] 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,

[6] 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."[7] 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).

[7] 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.
Nakhooda, 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.*

Adinda Van Hemelrijck, Independent Consultant, adindavanh@gmail.com.

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;[1] 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

[1] 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.

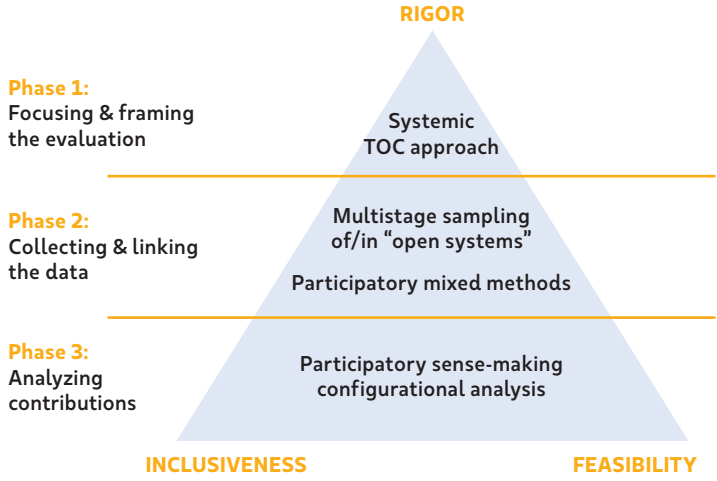
THE PARTICIPATORY IMPACT ASSESSMENT AND LEARNING APPROACH

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,[2] 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).

[2] 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).[3]

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.[4] 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

[3] 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).

[4] 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).[5] The final step involves integration of the findings for

[5] 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,[6] and the Roots and Tubers Improvement and Marketing Program (RTIMP), which was conducted countrywide in Ghana.[7] 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,[8] 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

[6] 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.

[7] 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.

[8] 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.[9] 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.[10] 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

[9] 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.

[10] 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.[11] 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

[11] 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.[12] 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

[12] 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.[13]

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

[13] *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.[14] 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.[15]

[14] 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.

[15] 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[a]	Strength of causality[b]
Impact Claim – Poverty Reduction					
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*		
Contribution Claim 1 – Enhanced Market Linking					
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[a]	Strength of causality[b]
Contribution Claim 2 – Enhanced Production of Roots and Tubers					
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)*			
M2a+M2b+(M2c)+M1c→C2a M2c→C2b			*Insert main findings from the KIIs with FFF facilitators, extension agents, DDAs, DADU officers, and CSIR, KNUST & UCC research leaders*		
		Insert main findings from the constituent feedback sessions with (non-)FFF participants (farmers and processors)			
Contribution Claim 3 – Enhanced Processing of Roots and Tubers					
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)*			
M3b+M3c+C1a→C3c			*Insert main findings from the KIIs with GPCs, BACs and PFI local branches*		
		2–3 constituent feedback sessions with (non-)GPC participants (farmers and processors)			

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.

Copestake, 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 P*articipation: 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 Glowen 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 Wörlen, 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.*

Takaaki Miyaguchi, Kyoto University of Foreign Studies, takaakinet@gmail.com.

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."[1] 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

[1] United Nations Framework Convention on Climate Change, "Glossary of climate change acronyms and terms," 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 (million $)	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 Wörlen 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 (Wörlen 2011). This meta-evaluation is one such attempt to apply the TONC concept to CCM projects that go beyond these fields.

According to Wörlen (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 Wörlen 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.[2] 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

[2] 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	Financier	Consumer	Supply chain	Policy maker	Financier	Consumer	Supply chain	Policy maker	Financier	Consumer	Supply chain	Policy maker	Financier	Consumer	Supply chain	Policy maker	Financier	Consumer	Supply chain	Policy maker	Financier	Consumer	Supply chain	Policy maker	Financier
I	●	●	●	●	●	●	●	●	●	●	n.a.	n.a.	●	●		n.a.			●	●			n.a.		●			●
M	●				●	●	●	●			n.a.	n.a.	●	●		n.a.	●						n.a.				●	
M	●	●		●	●	●	●	●	●	●	n.a.	n.a.				n.a.	●	●					n.a.					●
P	●	●	●	●		●	●	●	●	●	n.a.	n.a.	●	●	●	n.a.							n.a.		●	●		●
P	●			●	●	●	●	●	●	●	n.a.	n.a.	●	●	●	n.a.	●	●	●				n.a.			●	●	●
T	●	●		●		●	●	●		●	n.a.	n.a.	●			n.a.		●		●	●	●	n.a.	●		●	●	●
V	●	●	●			●	●		●	●	n.a.	n.a.	●		●	n.a.							n.a.			●	●	
V	●	●			●	●	●	●	●	●	n.a.	n.a.	●			n.a.	●			●			n.a.		●	●	●	●
	68% (22/32)				88% (28/32)				81% (13/16)				58% (14/24)				38% (12/32)				13% (3/24)				63% (20/32)			

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; ■ 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.

Wörlen, 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; Jyotsna Puri, Green Climate Fund, 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.[1] 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[2] 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

[1] 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.

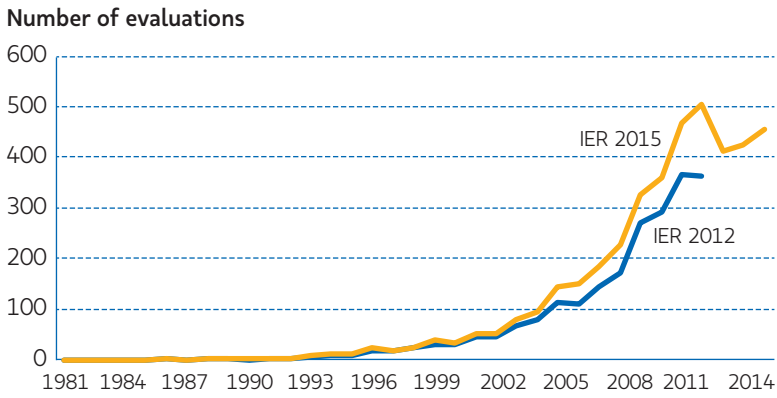
[2] Usually a mnemonic for smart, measurable, achievable, realistic, and time-bound.

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,[3] 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

[3] 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.

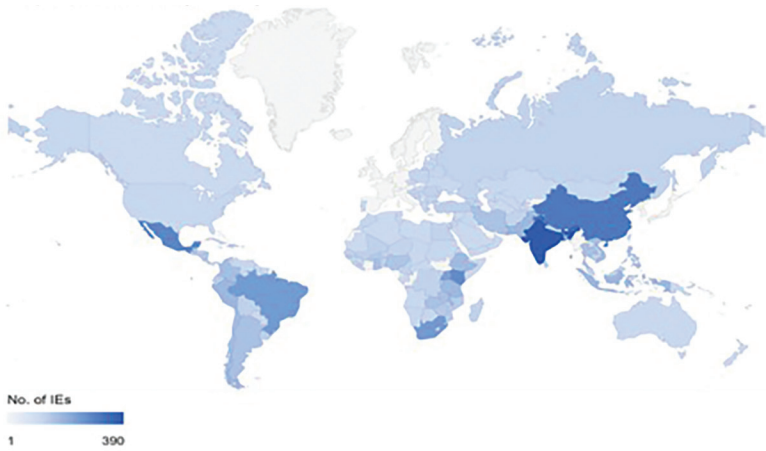
Geographic Gaps

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.[4] 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

[4] 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.[5] 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.

[5] 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 detraction 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.

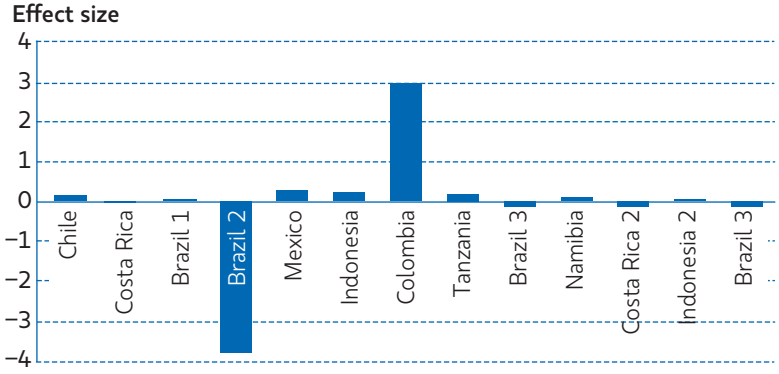
Credible and high-quality impact evaluations are critically dependent on defining the appropriate questions and "system boundaries," that is, defining the type and nature of the interventions that will be examined through impact evaluations. This in turn implies that impact evaluations are only able 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.[6]

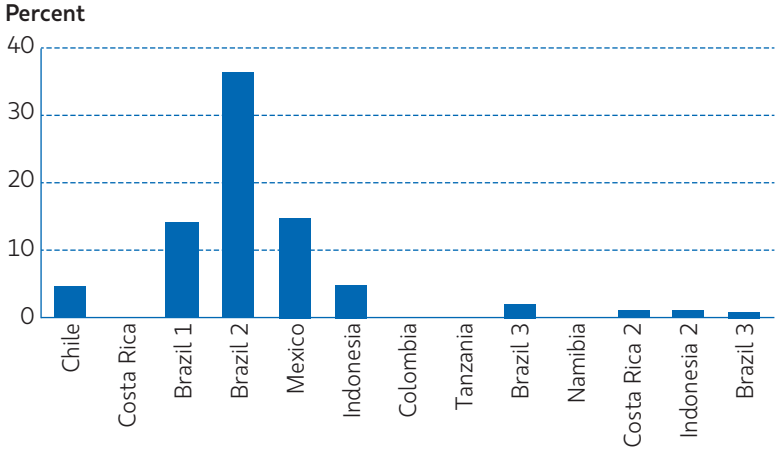
[6] See PLOS Collections (2016) for a summary of statistics from different forestry programs.

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 (Barooah, 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 Geleen, 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.*

Gwendolyn Wellmann, Independent Consultant, gwendolyn@gwendolynwellmann.com.

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,[1] 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

[1] 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.[2] 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.[3] 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,[4] 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

[2] All the mining companies referred to in this chapter belonged to the ICMM between 2002 and 2014.

[3] One international NGO was contracted by several gold mining companies to provide this service.

[4] 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).[5] 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 **Akoben 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.[6] 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 [PNDCL 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

[5] Each mine or mining project was assessed individually, irrespective of how many mining projects a company owned.

[6] See Agbesinyale (2003); Akabzaa, Seyire, and Afriyie (2007); Hilson (2002, 2004); and Kesse (1985).

times its contribution in 2000.[7] 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).[8]

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.[9] 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

[7] 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.

[8] 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.

[9] 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
	Very far	Far	Close	Very close		Very far	Far	Close	Very close	
Distance	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**

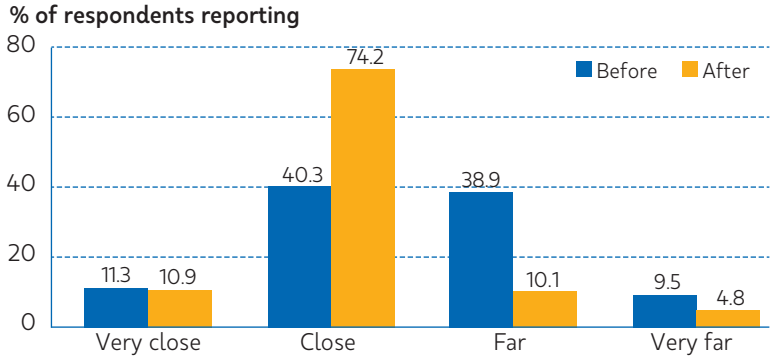


FIGURE 22. 3 **Percentage of respondents reporting various levels of water quality before and after company intervention**

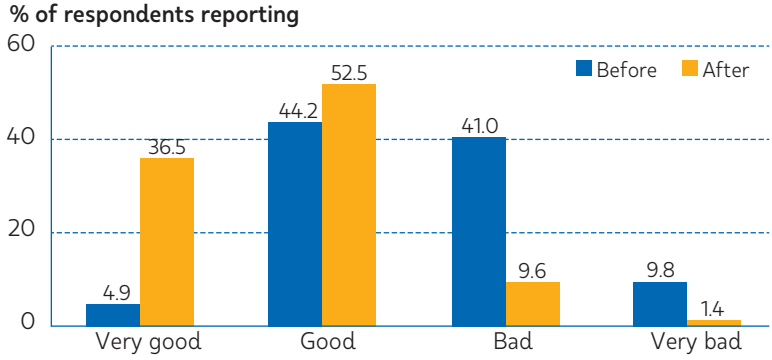


FIGURE 22. 4 **Percentage of respondents reporting various distances to toilet facility before and after company intervention**

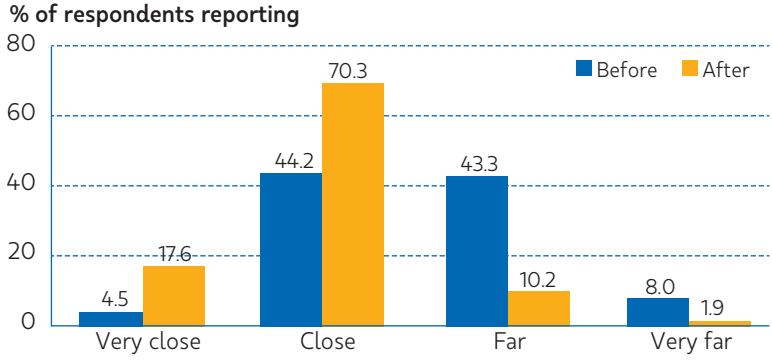


TABLE 22.5 **Household respondents' assessment of access, adequacy, and reliability of water before and after company intervention**

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.[10] 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.11 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[12]), and often initiate projects in the same communities, while government departments are

[10] "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.

[11] Global Reporting Initiative: https://www.globalreporting.org.

[12] 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.

Contributors

Taiwo Peter Adesoba is a young evaluation professional from Nigeria and Founder of the Centre for Knowledge, Learning & Evaluation in Nigeria. He has worked as a monitoring and evaluation (M&E) specialist on development programs funded by the President's Emergency Plan for AIDS Relief, the United States Agency for International Development (USAID), the U.K. Department for International Development (DFID), and the Global Fund. His most recent appointment was as State Monitoring, Evaluation, Research and Learning Coordinator at Christian Aid UK, where he was responsible for developing and strengthening M&E systems on an integrated community case management project in Nigeria. He is studying for a master's in environmental health at East Tennessee State University, and holds a BS in microbiology from the University of Ado Ekiti, Nigeria (now Ekiti State University).

Rashmi Agrawal has more than 30 years of experience working in various social sectors with the Indian government in association with international and community organizations. She has served as Director of the National Institute for Labour Economics Research and Development (NILERD), and works in the areas of monitoring and evaluation (M&E), skill needs assessment training in human resource development planning, gender issues, and development evaluation. She is a member of IDEAS and has served on its board, as well as chairing the IDEAS Thematic Group on Utilisation of Evaluations. She launched the Evaluation Community of India (ECOI), and is Cochair of the EvalGender+ Task Force on Capacity Building. She has published extensively on the empowerment of women and education of the disabled, as well as on evaluation use, policy, and capacity development.

Awny Amer is an independent research, monitoring, and evaluation consultant based in Egypt. He has 30 years of experience in development and evaluation, and has led and conducted a variety of research and evaluation processes at the global, national, and regional levels, primarily for United Nations agencies and international nongovernmental organizations. He is a board member of

the International Development Evaluation Association (IDEAS), the African Policy Centre, the Egyptian Research and Evaluation Network, and the Community of Evaluators; he is also a founder of the Middle East and North Africa Evaluators Network, and an active member of the African Evaluation Association. He has also translated some key research and evaluation publications, including *The Road to Results* textbook.

Inga-Lill Aronsson is an Assistant Professor in Cultural Heritage Studies at Uppsala University (UU). Her research interests are longitudinal studies of resettlement and heritage, memories, and reconciliation in postconflict areas. Among her publications are *Negotiating Involuntary Resettlement* (2002), *The Gendered Museum* (2005, published in Swedish), "The Ambivalent Concept of Heritage" (2013), "Heritage as Life Value: A Critical Investigation of the Heritage Concept" (2016), and "On Negotiations in Resettlement" (2017). She is a member of the board of the Latin American Studies Forum, UU, and of the Research Colloquium of the Swedish Research Institute in Istanbul. She is the former Director of the Network on Humanitarian Action. She holds a PhD in cultural anthropology from UU.

Marife Ballesteros is currently Vice President at the Philippine Institute for Development Studies. She has been involved in several evaluation studies for policy and program reforms. She has written several papers on topics pertaining to economic development and is a leading national expert in the fields of housing, land, and urban development policy. She is a member of the Society of Industrial and Organizational Economics (SIOE) and the International Development Evaluation Association (IDEAS). She earned her PhD in social sciences at the Radboud University in Nijmegen, the Netherlands.

Chris Barnett is the Technical Director of Itad (a consultancy for monitoring, evaluation, and learning); and formerly the Director of the Centre for Development Impact (CDI). His portfolio focuses on evaluating impact and broadening the use of evaluative methodologies, particularly in relation to civil society, governance, and livelihoods. An experienced team leader, he is currently the Project Director for the quasi-experimental impact evaluation of the Millennium Villages Project in Ghana; Project Director of a 10-year study of the CDC Group's investment mobilization; and methodological expert on the evaluation of Ideas to Impact, a global program of innovation prizes. He has recently published on evaluation ethics and assessing the social effect of impact investing.

Sonia Ben Jaafar is the Managing Director of EduEval Consultancy, a firm based in the United Arab Emirates that focuses on monitoring and evaluation (M&E) in multiple sectors. She has worked with a number of governments, nongovernmental organizations, and corporations in the Middle East and North Africa to integrate evidence-based practices for programs and policy. In her work that focuses on M&E in the education development sector, she

strives to help create sustainable improvement that effectively increases local capacity and achievement. A former elected board member of the International Development Evaluation Association (IDEAS), she continues her commitment to the professionalization of M&E in the Middle East and North Africa. She holds a PhD from the Ontario Institute for Studies in Education of the University of Toronto.

Lee Cando-Noordhuizen is a Consultant with the International Development Evaluation Association (IDEAS), and most recently with the Global Environment Facility Independent Evaluation Office (GEF IEO). She has more than 10 years' professional experience in the fields of climate change and sustainable development: policy, strategic communications, program coordination and project management. She has worked for various United Nations agencies, including the Executive Office of the UN Secretary-General, the United Nations Development Programme (UNDP), UNICEF, and the United Nations Framework Convention on Climate Change (UNFCCC). She also worked briefly with GIZ (German Agency for International Cooperation) and the Center for International Forestry Research (CIFOR).

Warren Crowther is a consultant for the United Nations, having worked for eight of its agencies in 39 countries. Concurrently, he was a professor for 38 years, and cofounded two postgraduate programs, including the Postgraduate Evaluation of Policies, Programs and Development Projects at the University of Costa Rica. He was also a visiting professor at many other universities. Crowther advises many nongovernmental organizations and community projects, especially in the subject areas mentioned in table 9.1 in chapter 9.

Rachel Eager is Head of Knowledge Management at Itad, and Deputy Director of the Centre for Development Impact (CDI). She was previously Deputy Head of Impact, Innovation and Evidence at Save the Children. She is an experienced monitoring and evaluation practitioner, with a focus on innovative approaches to impact evaluation. She provides methodological and design expertise across a range of evaluations, including the U.K. Department of Health's Fleming Fund, a £265 million one health program. She has recently published on the evaluation of a women's empowerment project in Tajikistan.

Abdoulaye Gounou is Chief of the Bureau of Public Policy Evaluation in Benin; previously, he served as a Technical Adviser to the president and several ministries in Benin, working on issues of policy evaluation and the promotion of private sector development over the last decade. In this capacity, he has been involved in process and impact evaluation of public policies, capacity development events, and collaborative projects. He has over 25 years of experience in financing development and promotion of the private sector. He holds a diploma of advanced professional studies from Senghor University in Alexandria, Egypt, in business administration and a project management and master of laws degree.

Tara Devi Gurung is Planning Monitoring Evaluation and Reporting (PMER) Coordinator for the Nepal Red Cross Society (NRCS). She is responsible for coordinating PMER initiatives and capacity-building activities in the NRCS. Her areas of expertise include project design (logical framework approach and theory of change), strategic planning, proposal writing, monitoring, evaluation, reporting, emergency plan of action development, reporting, case study writing, PMER trainings, and evaluation management. Her academic background is in microbiology. She holds a master's degree in microbiology.

Kabir Hashim is Minister of Public Enterprise Development in Sri Lanka and Chair of the Global Parliamentarians Forum for Evaluation. He has been serving as a member of Parliament in Sri Lanka since 1994. In 2001, he became a minister; in 2015, he became the Secretary General of the United National Party in Sri Lanka, the largest single political party in the country. He was instrumental in initiating parliamentarians' forums for evaluation in South Asia and globally. In the early 2000s, he initiated a campaign to emphasize political participation in evaluation; the campaign ultimately succeeded in engaging more parliamentarians in promoting an evaluation culture, particularly in developing countries. Hashim has been a keynote speaker on engaging policy makers in evaluation in many international forums, including at the United Nations headquarters in New York.

Ana Luisa Guzmán is General Coordinator of the Latin American and Caribbean Network of Monitoring, Evaluation and Systematization and a professor at the University of Costa Rica, Faculty of Education and Postgraduate Evaluation of Policies, Programs and Development Projects. She is a member of the Coordinating Group of the Capacity Development in Evaluation Project (FOCEVAL) in the Ministry of Planning and Economic Policy of Costa Rica, with support from the German Institute for Development Evaluation (DeVAL), the Regionalization Platform for Strengthening Capacities in Evaluation, and the Network of Latin American Women to Strengthen Women's Personal and Social Development. She is a seasoned independent consultant in the fields of education, childhood/adolescence, gender, and the evaluation and monitoring of programs and development projects.

Emmanuel (Manny) Jimenez is Executive Director of the International Initiative on Impact Evaluation (3ie), a nonprofit organization which provides grants for the rigorous assessment of the effectiveness of development projects and programs and supports the use of such evidence in decision making. He came to 3ie in 2015 after many years at the World Bank Group, where he provided technical expertise and strategic leadership in a number of research and operational positions. These included serving as Director of the Bank's operational program in human development in its Asia regions and as Director of public sector evaluations. Previously, he was a member of the economics faculty at the University of Western Ontario in London, Canada. He received his PhD from Brown University.

Asela Kalugampitiya is a Sri Lankan lawyer currently working as the EvalPartners Executive Coordinator. He manages initiatives related to the Global Evaluation Agenda 2016–2020 and EvalPartners global activities. Active in the monitoring and evaluation field for over 15 years, particularly in gender equality and policy development, he has worked for international organizations including the International Development Law Organization, United Nations Women, the United Nations Population Fund, the International Labour Organization, and CARE International. Kalugampitiya was instrumental in initiating the Engaging Parliamentarians in Evaluation campaign, which led to the launch of the Global Parliamentarians Forum for Evaluation. He was also instrumental in the successful campaign for the International Year of Evaluation 2015 and organizing the Global Evaluation Week held at the Nepalese Parliament.

Jeremy Kohlitz has eight years of experience working on development issues in the Pacific island region, primarily in the water, sanitation, and hygiene sector. His current research seeks to bridge different disciplinary perspectives on climate change vulnerability and the resilience of community-managed water services in a Pacific island setting to produce more sustainable water service outcomes. Kohlitz holds an MSc (with distinction) in water and waste engineering. He is a PhD candidate at the Institute for Sustainable Futures, University of Technology Sydney.

Takaaki Miyaguchi is an Associate Professor at Kyoto University of Foreign Studies in Japan. His research topics include the role of the private sector in climate change, and evaluation of donor-funded climate change mitigation and adaptation projects. Before starting an academic career in 2012, he worked at the United Nations Development Programme and other United Nations' agencies for eight years, developing and managing both climate change mitigation and adaptation projects in more than 20 countries. He holds a PhD in global environmental studies from Kyoto University, Japan.

Linda Morra Imas is Cofounder and Codirector of the International Program for Development Evaluation Training. She was a Chief Evaluation Officer and Evaluation Capacity Building Adviser for the World Bank Group, and a Director at the U.S. Government Accountability Office. Now an independent consultant building on more than 30 years of experience, she advises and provides monitoring and evaluation training worldwide. She coauthored the comprehensive textbook on development evaluation, *The Road to Results*, and is known for her work in developing evaluation competencies and professionalization. She holds a doctorate in evaluation from the University of Virginia and is an honorary member of the International Development Evaluation Association (IDEAS).

Indran Naidoo has overseen a significant reform of the evaluation function at the United Nations Development Programme (UNDP), moving it toward full independence, enhanced credibility, and an expansion in the scope and

coverage of UNDP work. Through the professional model under his tenure, the organization has seen the Independent Evaluation Office take the lead and full responsibility for its work at the global and country levels, with changes evident in the methodologies and approaches toward country-level and corporate assessments, and quality assessment of decentralized evaluations and national evaluation capacity. He has substantial experience in evaluation thought leadership, and has served as an instructor for the International Program for Development Evaluation Training and globally as a panelist on the subject.

Rituu B. Nanda works part time with the Institute of Social Studies Trust (ISST). She also works with an international nongovernmental organization, Constellation, on community development and knowledge management. She has 10 years of work experience, primarily in public health in implementation, community engagement, monitoring and evaluation, knowledge management, and program management. At ISST, Nanda has been facilitating an online community of practice on gender and evaluation. She also conducts gender and equity–focused evaluation. She is a founder and core group member of the Evaluation Community of India.

Rajib Nandi is a Research Fellow and Officer-in-Charge at the Institute of Social Studies Trust, New Delhi. He has 20 years of experience in research and evaluation. His areas of work cover gender and development, solidarity economy, information and communications technology, environment, HIV/AIDS, social movements, program evaluations, and evaluative studies. He is a founder and core group member of the Evaluation Community of India. Presently, he is a board member of the Community of Evaluators–South Asia. He has co-edited a book entitled *Voices and Values: The Politics of Feminist Evaluation*, published by Zubaan Books, New Delhi.

Awuor Ponge is an Associate Research Fellow-in-Charge of Research, Policy and Evaluation at the African Policy Centre. He is an immediate former International Development Evaluation Association (IDEAS) Board member (2014–17) representing Sub-Saharan Africa, and the African Regional Representative on the board of the International Organisation for Collaborative Outcome Management. Ponge previously served as a board member of the South African Monitoring and Evaluation Association, as well as the National Treasurer of the Evaluation Society of Kenya. He is a doctoral candidate at Kenyatta University's Department of Gender and Development Studies, where he also serves as an Adjunct Faculty.

Susanna Price, currently a Lecturer (Hon.) at the College of Asia and the Pacific, Australian National University, has worked at AusAID as an Evaluation Specialist and at the Asian Development Bank (ADB), Manila, as a Social Development Specialist in Resettlement. In 2015–16, she was the Resettlement Specialist for an ADB thematic evaluation study on safeguards. She regularly publishes on displacement, resettlement, and social impact

assessment. Recent publications include "Making a Difference? Social Assessment Policy and Praxis and its Emergence in China" (with Kathryn Robinson) and "Responses to Displacement across Asia-Pacific: Strategies for Development, Disasters and Climate Change" (with Jane Singer).

Jyotsna (Jo) Puri is Head of the Independent Evaluation Unit of the Green Climate Fund. She is also an adjunct Associate Professor at Columbia University's School of International and Public Affairs. Her areas of work include policy impact analysis on poverty reduction, environment, agriculture, health, and climate change. She has more than 22 years of experience in policy research and development evaluation at organizations including the World Bank and the United Nations, and has led evaluation-related work for the United Nations Development Programme, UNICEF, and the Global Environment Facility. Her expertise is in applying mixed methods that use rigorous quantitative methods informed by high-quality qualitative approaches. She holds a PhD and MSc in resource economics and a master's in development economics.

Jigmi Rinzin chairs the Natural Resources and Environment Committee of the National Council of Bhutan and serves on the Parliamentarians Forum for Development Evaluation in South Asia (PFDE) and the Global Parliamentarians Forum for Evaluation. Rinzin previously chaired parliamentary committees on public accounts, economics, good governance, and entitlements. He helped draft the Asia Regional Association of Public Accounts Committees Constitution, assisted international forums on public accountability and good governance, and worked on anti-corruption auditing measures in Bhutan and on financial auditing guidelines. He is a fellow Chartered Certified Accountant from the United Kingdom and holds a bachelor's of commerce degree from the University of Delhi, India.

Mallika Samaranayake is a Founding Member and the first President of the Community of Evaluators, South Asia, and a member of the Audit and Evaluation Advisory Committee of the United Nations Development Programme (UNDP). She was President of the Sri Lanka Evaluation Association; member of the Core Evaluation Team for Phase 2 of the Paris Declaration Evaluation, which won the American Evaluation Association 2012 Outstanding Evaluation Award; and member of the Steering Committee of the MFS (MedeFinancieringsStelsel) II Joint Evaluations of the Dutch Co-Financing System 2011–2015. Samaranayake has more than 40 years' service and consultancy experience with public sector, bilateral, and international nongovernmental organizations. As a participatory-development specialist, she is committed to promoting participatory approaches in development.

Marco Segone is Director of the Evaluation Office of the United Nations Population Fund (UNFPA). Previously, he was Director of the Evaluation Office at UN Women; Chair of the United Nations Evaluation Group; and Cofounder and Cochair of EvalPartners and EvalGender+. Earlier, he was responsible for the

decentralized evaluation function as well as the national evaluation capacity development portfolios at the UNICEF Evaluation Office. Also at UNICEF, he served as Regional Chief, Monitoring and Evaluation, for Europe and Central Asia; and for Latin America and the Caribbean. He has also worked in international nongovernmental organizations in Albania, Bangladesh, Pakistan, Thailand, and Uganda. He has authored dozens of publications, including *Evaluating the SDGs with a "No One Left Behind Lens," Evaluation for Equitable Development Results* and *How to Design and Manage Equity-Focused Evaluations*.

Shekhar Singh is currently Coordinator of the Research, Assessment and Analysis Group and a member of the Environment Sub-group of the Narmada Control Authority. He has been Chair/member of various government of India environmental appraisal committees and a Supreme Court commissioner for forests and related matters for the Andaman and Nicobar Islands. He has also headed the Environment and Forests Division of the Planning Commission of India, and litigated in the Supreme Court against various environmentally sensitive projects. He has taught philosophy and environmental management at various universities and institutions, and has served as a consultant to the World Bank, the United Nations Development Programme, UNESCO, the Global Environment Facility, and the Food and Agriculture Organization of the United Nations.

Ana Rosa Soares, Senior Evaluation Advisor at the United Nations Development Programme Independent Evaluation Office (UNDP IEO), is responsible for conducting country, regional, and global thematic and programmatic evaluations. She has 18 years of experience in international development, 13 of which have been mostly dedicated to evaluation and national evaluation capacity development. Prior to joining IEO, she was a Strategic Planning, Monitoring and Evaluation Officer for UNDP in Brazil. Before joining the United Nations, Soares was an International Development Advisor at the State University of New York (SUNY) Center for International Development and organizations in Brazil, working with bilateral and multilateral cooperation projects in Africa, Europe, and Latin America.

Ahmed Tammam is a physician with 13 years of professional experience in the area of capacity building and monitoring and evaluation (M&E) on national and regional levels. He supported the establishment of M&E systems in more than 18 countries in the Middle East and Eastern Europe. He has worked with many United Nations and other international organizations, and supported the development of the Monitoring, Evaluation, Accountability and Learning (MEAL) framework for the Middle East and Eurasia region of Save the Children International. Tammam holds a master's degree in public policy from the American University in Cairo as well as a master's degree in immunology and statistics.

Susan D. Tamondong is Vice-President of the International Development Evaluation Association (IDEAS) and Adviser to the UN Women's Global Evaluation Advisory Committee. She is an accomplished professional with wide experience

in evaluation and resettlement worldwide, including work with the World Bank, various United Nations agencies, AusAID, and the Asian Development Bank. She edited the *Oxford International Review* and has published numerous articles and chapters on resettlement and evaluation. Her most recent publication is *The Future of Evaluation: Perspectives from Developing Countries* (Palgrave MacMillan 2015). She was educated at Oxford University (U.K.), the American University (Washington, D.C.), and the University of the Philippines. She trained in evaluation at the International Program for Development Evaluation Training (IPDET) in Canada and the University of Bologna, Italy.

Michele Tarsilla has been designing and conducting development and humanitarian evaluations as well as capacity development programs funded by development organizations—including the Food and Agriculture Organization of the United Nations, UNESCO, UNICEF, the United States Agency for International Development, and the World Bank—in a variety of sectors for 15 years. His extensive fieldwork in over 40 countries and fluency in six languages have given him a firm appreciation of cultural competence in international development work, and a keen commitment to promote the use of innovative and progressive concepts and practices within the global evaluation community. Tarsilla holds a PhD in interdisciplinary evaluation from Western Michigan University and an MS in international development from Georgetown University's School of Foreign Service. A prolific blogger (Twitter: @MiEval_TuEval), he is a former International Rotary and Fulbright Scholar.

Florencia Tateossian is an Evaluation Specialist with the Independent Evaluation Office of UN Women. Her focus includes conducting gender-responsive corporate evaluations and managing the national evaluation capacity development portfolio. She coordinates EvalGender+, the global partnership to promote the demand, supply, and use of equity-focused and gender-responsive evaluations. She previously worked at the United Nations Development Programme (UNDP) Independent Evaluation Office, the World Bank, and the European Commission, as well as in government agencies both in Argentina and Armenia. A Fulbright Scholar, Tateossian holds an MA in public and international affairs from the University of Pittsburgh and a degree in political science from the University of Buenos Aires.

Vinod Thomas is Visiting Professor at the National University of Singapore and at the Asian Institute of Management, Manila; and was Director General of Independent Evaluation at the Asian Development Bank. Prior to that, he was the Director-General and Senior Vice-President of the Independent Evaluation Group at the World Bank Group. His books include *The Quality of Growth* (2000) and (with Xubei Luo) *Multilateral Banks and the Development* Process (2012) and *Climate Change and Natural Disasters* (2017). He has taught at Vassar College and the University of São Paulo, Brazil, and has addressed numerous professional and academic forums. Thomas holds a PhD and MA in economics from the University of Chicago, and a BA from St. Stephen's College, Delhi.

David Todd was Senior Evaluation Specialist in the Independent Evaluation Office of the Global Environment Facility (GEF) for several years and has since engaged in climate change evaluation work for the International Climate Fund (U.K.), Caribbean Community Climate Change Centre, and United Nations Environment Programme. He was Social Development Adviser in the European Commission Delegation in Barbados and the Eastern Caribbean as well as a member of the Evaluation Office of the U.K. Department for International Development. He is Director of International Development, Environment and Disasters (IDEDs), a consultancy company registered in Barbados, and is an active consultant.

Juha I. Uitto is Director of the Global Environment Facility (GEF) Independent Evaluation Office. He has worked in the GEF and the United Nations Environment Programme (UNDP) since 1999 as Senior Monitoring and Evaluation Specialist/Coordinator, Evaluation Advisor, and, finally, Deputy Director of UNDP's Independent Evaluation Office. Uitto spent the 1990s with the United Nations University in Tokyo as coordinator of the university's environment and sustainable development research and training programs. He has also worked with the Nordic Africa Institute and the International Fund for Agricultural Development, and has held visiting positions in universities in the United States and Japan. He was educated at the Universities of Helsinki and Lund, and holds a PhD in social and economic geography.

Rob D. van den Berg, President, International Development Evaluation Association (IDEAS) and Visiting Professor, International Development Institute, King's College (London), is a graduate of the State University of Groningen in the Netherlands. He is also a Visiting Fellow at the Centre for Development Impact at the Institute of Social Studies in Brighton, U.K. He is an adviser to the Climate Investment Funds and the Green Climate Fund, as well as a member of the Advisory Committee of Wilton Park. From 2004 to 2014, he was Director of the Independent Evaluation Office of the Global Environment Facility (GEF) in Washington, D.C.; before that, he was Director of Evaluation of the Dutch Ministry of Foreign Affairs from 1999 to 2004. He has been active in the Development Assistance Committee (DAC) Evaluation Network (Chair from 2002–04), the United Nations Evaluation Group (UNEG), and the Evaluation Cooperation Group of the International Financial Institutions.

Adinda Van Hemelrijck works as an Independent Consultant, designing, monitoring, and evaluating impact in complex development contexts (mostly rural and empowerment centered), using systemic and participatory approaches. Her work focuses on capacity building and facilitation of multistakeholder engagements to foster social innovation and transformational change. Among her clients are Oxfam, the International Institute for Environment and Development, the International Fund for Agricultural Development, and UNICEF. She worked for five years at Oxfam America, assisting regional offices and partners with rights-based impact monitoring and evaluation. Prior to this, she

worked for various European Union–based agencies on performance evaluation and learning, gender, and communication for social change.

Gwendolyn Wellmann is an independent sociologist who has worked with the extractive industry in Africa, South America, and the Philippines since 2002, and with government agencies since 1995. She specializes in sustainable development and has both implemented and evaluated development projects for clients based in the extractive industry, government agencies, and nongovernmental organizations. She is the author of two community development handbooks that cover, among other topics, stakeholder identification, grievance mechanism, local content, social impact, and community investment. Her academic work includes advising master's and PhD students at the Nelson Mandela University, Port Elizabeth, South Africa.

Evaluation for Agenda 2030
PROVIDING EVIDENCE ON PROGRESS AND SUSTAINABILITY

Rob D. van den Berg, Indran Naidoo, and Susan D. Tamondong, *editors*

Building on insights and state of the art gathered at a ground-breaking conference in Bangkok in 2015, this book presents an authoritative overview of the role of evaluation in generating evidence for informed decision making and learning to advance sustainable development. Agenda 2030 includes the Sustainable Development Goals (SDGs), and evaluation has been identified as one of the tools to understand and report on progress toward these goals. In the introductory part of the book, the challenges as well as the evaluation approaches to tackle them are discussed. An overview of how evaluation can contribute to understanding sustainable development is followed with the potential of gender- and equity-focused evaluation to "leave no one behind," and participatory evaluation to give the people a voice in their future.

The second part discusses capacities and capabilities needed for evaluation focusing on sustainable development. The third part looks at regional perspectives as evaluation becomes increasingly country-led, focusing on their development priorities. The part's chapters inform the reader on developments in Asia, Sub-Saharan Africa, small island developing states, Latin America, and the Middle East and North Africa. The fourth part examines the negative impacts of development displacement, looks at evaluative evidence on the contribution safeguards make toward sustainability, and highlights resettlement policies and practices in development. The book's fifth and final part looks at various efforts to provide evidence that an intervention, policy, or action leads to impact— not only at the level of the intervention, but at national, regional, and global levels. This part also explores new frontiers for evaluation as SDGs and the 2030 Agenda require new partnerships, new modes of operation, and innovative ways of financing that pose special problems for evaluation.

Evaluation for Agenda 2030: Providing Evidence on Progress and Sustainability is written for policy makers, practitioners, academics, researchers, and other informed audiences concerned with sustainable development and evaluative evidence, systems for monitoring and evaluation, and related capacity issues.

This book is open access under a Creative Commons license.
ISBN (paper): 978-1-9999329-0-9
ISBN (electronic): 978-1-9999329-1-6