The Realpolitik of Evaluation

*The Realpolitik of Evaluation* shines a light on the divergent demands for evaluation. But what explains the “gap” between what those on the “demand” side expect in terms of evaluation results, and the “supply” of information provided by evaluators? Can anything be done to narrow this gap? What works and what does not work?

Examining these questions from both the demand and the supply side, experts describe ten different global examples of the gap between demand and supply of evaluation information in different contexts. In an attempt to bridge that gap, they effectively expose the biases behind supposedly non-partisan sources of evaluation information and highlight the pros and cons of attempts to bridge the gap through the use of third parties, enhanced stakeholder involvement, and the incorporation of social science models to strengthen Theories of Change (ToC).

*The Realpolitik of Evaluation* is an important book that poses questions at multiple levels of thinking. It will be of great interest to policymakers, program implementers, and project managers.

**Markus Palenberg** is the Founder and Manager of the Institute for Development Strategy (IfDS) in Munich, Germany, where he works as researcher, evaluator, and advisor. Over the last 15 years, Markus has conducted more than 40 research and consulting assignments in the public and private sector, specializing in evaluation of global programs and evaluating and advising on results-based management. He regularly serves as peer reviewer of research publications and advisor during the design phases of global programs, has published works on evaluation methodology, and has served on boards, executive committees and advisory panels for development programs and institutions.

**Arne Paulson** pursued a career in international development, working as an economist in a number of international organizations, including the World Bank, UNCTAD (Geneva), the International Energy Agency (Paris), and the Inter-American Development Bank (IDB), from which he retired in 2007. At the IDB, he worked in all aspects of evaluation, including ex-ante economic evaluation, ongoing monitoring of projects in execution, and ex-post evaluation of completed projects financed by the IDB. At the corporate level, he reported on the overall development effectiveness of IDB operations to the bank’s board of directors and at international conferences.
Comparative Policy Evaluation
Edited by Ray C. Rist

The *Comparative Policy Evaluation* series is an interdisciplinary and internationally focused set of books that embodies within it a strong emphasis on comparative analyses of governance issues – drawing from all continents and many different nation states. The lens through which these policy initiatives are viewed and reviewed is that of evaluation. These evaluation assessments are done mainly from the perspectives of sociology, anthropology, economics, policy science, auditing, law, and human rights. The books also provide a strong longitudinal perspective on the evolution of the policy issues being analyzed.
The Realpolitik of Evaluation
Why Demand and Supply Rarely Intersect

Edited by Markus Palenberg and Arne Paulson
## Contents

*List of Figures* vii  
*List of Tables* viii  
*Notes on Contributors* ix  
*Foreword* xiii  

**RAY C. RIST**

1 **Introduction** 1  
**MARKUS PALENBERG AND ARNE PAULSON**

2 **The Results Paradox in Norwegian Development Assistance** 9  
**IDA LINDKVIST AND PER ØYVIND BASTØE**

3 **Behind the Scenes of the French Evidence-Based Policy Movement: Rise and Fall of an RCT-Driven Model of Evidence** 24  
**AGATHE DEVAUX-SPATARAKIS**

4 **Influencing the Supply of and Demand for Results Information in Government: Can Outsiders Help? A Case Study from Ireland** 42  
**RICHARD BOYLE**

5 **Evaluation One Step Removed: The Government/Not-for-Profit Performance Nexus** 60  
**DAVID GILCHRIST AND PETER WILKINS**

6 **The Role of an Intermediary Organization in a Large Evaluation: The Case of the Norwegian Health Care Reform** 77  
**OLAF RIEPER**
Contents

7 Evaluation Plans in the EU Cohesion Policy: Redefining the Relation Between the Demand for and the Supply of Evaluation 88
FRANCESCO MAZZEO RINALDI

8 Assessing the Contribution of Research to Improved Policy and Practice: An Evaluation of CIFOR’s Climate Change Research 102
BRIAN BELCHER, JOHN YOUNG, AND DANIEL SURYADARMA

9 Mind the Gaps: Integrating Human Rights and Sustainable Development Goals into Evaluation Practice 123
PEARL ELIADIS

10 Mending the Theory Gap in Evaluation: Moving Towards Theory Knitting 136
SEBASTIAN LEMIRE, CHRISTINA CHRISTIE, AND STEFFEN BOHNI NIELSEN

11 Realistic Commissioning of Impact Evaluations: Getting What You Ask For? 151
JOHN MAYNE

12 Realpolitik of Demand and Supply Regarding Results Information 165
R. PABLO GUERRERO

Index 170
Figures

2.1 Average number of times the word “result” appears per page in Norwegian budget proposals for all ministries for the year 2015/2016

2.2 Average number of times the word “result” appears per page in budget proposals from the Norwegian Ministry of Foreign Affairs from 2006/2007 to 2015/2016

3.1 Graphic representation of the EBP movement in France

4.1 Theory of change for The Atlantic Philanthropies influence on government policy regarding public service reform

8.1 GCS-REDD+ Outcome Evaluation Process

8.2 GCS Theory of Change

10.1 The Theory Knitting Model

10.2 Theory use in theory-based evaluation

10.3 Theory Knitting in published theory-based evaluations
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>WA-based Best Practice Programs</td>
<td>69</td>
</tr>
<tr>
<td>7.1</td>
<td>Evaluation questions</td>
<td>95</td>
</tr>
<tr>
<td>9.1</td>
<td>Sample core international human rights instruments/treaties</td>
<td>125</td>
</tr>
<tr>
<td>9.2</td>
<td>Needs-based versus human rights-based approaches</td>
<td>128</td>
</tr>
<tr>
<td>11.1</td>
<td>Traditional impact evaluation questions</td>
<td>156</td>
</tr>
<tr>
<td>11.2</td>
<td>Meaningful impact evaluation questions</td>
<td>158</td>
</tr>
<tr>
<td>11.3</td>
<td>Meaningful accountability questions</td>
<td>161</td>
</tr>
</tbody>
</table>
Contributors

Per Øyvind Bastøe is Evaluation Director at the Norwegian Agency for Development Cooperation (Norad). Bastøe has broad experience in international development, evaluation, results management, and organizational change and has previously held senior positions in other parts of the Norwegian government administration, in the World Bank, and the Asian Development Bank. He has published several books and articles on development policy, evaluation, and organizational change.

Brian Belcher is Professor in the College of Interdisciplinary Studies at Royal Roads University (Victoria, Canada), where he holds the Tier 1 Canada Research Chair in Sustainability Research Effectiveness. Professor Belcher’s research focuses on research effectiveness, research quality, and knowledge translation, primarily in the context of natural resources management and sustainable development. He is developing and applying theory and methods to understand and improve research effectiveness, research quality, and the societal impact of research.

Richard Boyle is head of research, publishing, and corporate relations with the Institute of Public Administration in Ireland. He has worked with the IPA since 1986. His research interests focus on public service reform and developing and implementing effective performance management and evaluation systems. He has researched and written extensively on evaluation and public service reform and has worked widely with Irish central and local government, the OECD, and the World Bank. He is a former board member of the European Evaluation Society and a founder member of the Irish Evaluation Network.

Christina (Tina) Christie is Professor of Social Research Methodology and Chair of the Department of Education at UCLA. Dr. Christie specializes in social program and policy evaluation. Her research focuses on the factors and conditions that influence evaluation practice in an effort to strengthen our understanding of evaluation as a method for facilitating social change. Christie has served as a lead investigator on over 30 grants and contracts funding evaluative research on education, child maltreatment, and health behavior
Agathe Devaux-Spatarakis manages and works on evaluation projects and strategic consulting assignments for French local and national authorities as well as research institutions. She has a PhD in political science and specializes in (theory-driven) impact evaluations and the assessment of innovations and experiments. Prior to this, Agathe was the scientific manager of impact evaluations of agricultural research at Cirad and a consultant at Euréval while pursuing her doctoral degree.

Pearl Eliadis is a human rights lawyer and practitioner who has successfully led complex, global projects on national human rights institutions, capacity assessment, and women’s rights with institutional and multilateral clients. With more than 20 years’ experience, Pearl also serves as adjunct professor at McGill University’s Law Faculty and at the Max Bell School of Public Policy, where she is a member of the teaching faculty. She is a full member of the McGill Centre for Human Rights and Legal Pluralism and serves on several non-profit and advisory boards. She has published extensively on human rights and public policy.

David Gilchrist is Professor of Accounting at the University of Western Australia. He has held senior roles in the public sector, commerce, and the not-for-profit sector in Australia and has taught in a number of universities including the London School of Economics in the UK. He holds a visiting professorship at Hebei GEO University in the People’s Republic of China. Gilchrist has published widely in relation to the government and not-for-profit nexus, human services policy, and sustainability. He was a lead author of the seminal report “Australian Charities 2013,” an author of The Three Sector Solution published by the Australian National University Press, and an author of Public Sector Accounting, Accountability and Governance for Routledge.

R. Pablo Guerrero is an economist and former advisor at the World Bank. He has graduate degrees in economics and business and did postgraduate executive education at Harvard University. He is a former chief of evaluation at the Inter-American Development Bank and advisor to the Director-General of Evaluation at the World Bank. Among his contributions are the launching of the Evaluation Capacity Development Program at the World Bank, managing
the World Bank’s Secretariat for the Comprehensive Development Framework, and developing the World Bank’s Strategy for Middle-Income Countries. He has done consulting work on country strategies and finance for the World Bank and, since the 1990s, has contributed to several books in the Comparative Policy Evaluation Series: *Evaluation Capacity Development, Success in Evaluation*, and *Doing Public Good?*

**Sebastian Lemire** is an associate at Abt Associates. Sebastian has over the past 15 years designed and implemented a broad range of evaluations – nationally and internationally. As a scholar-practitioner, his interests revolve around theory-based evaluation, alternative approaches to impact evaluation, and evidence reviews. Sebastian has published numerous articles and special issues on these topics in evaluation and policy studies journals. He currently serves as Associate Editor of the *American Journal of Evaluation*.

**Ida Lindkvist** is a senior advisor in the Evaluation Department at the Norwegian agency for development co-operation (Norad), working with the evaluation of Norwegian development assistance. Lindkvist holds a PhD in economics, having previously worked as a researcher on global health and published various articles and book chapters in this area. Lindkvist has recently taken a particular interest in results-based management, results-based financing, and the political economy of evaluation.

**John Mayne** is an independent advisor on public sector performance based in Ottawa and Bangkok. Over the last 15 years he has focused on international development evaluation and results-based management. He has authored numerous articles and reports, including several on contribution analysis and theories of change, and co-edited eight books on program evaluation. In 1989 and 1995, he was awarded the Canadian Evaluation Society Award for Contribution to Evaluation in Canada. In 2006, he was made a Canadian Evaluation Society Fellow. Dr. Mayne’s current interests are on approaches in complex settings for strengthening impact evaluation and building useful theories of change.

**Steffen Bohni Nielsen** is Director-General at the Danish National Research Centre for the Working Environment. Previously, he has held positions as an Executive Director in local government, Senior Director in Ramboll Management Consulting, partner at PwC and Deputy Director at the National Board of Social Services. He has carried out an extensive number of evaluations, nationally and internationally and has worked as commissioner of evaluations. He has published articles in all major peer-reviewed evaluation journals on a variety of topics. He currently serves as Chairman of the Board of the Danish Society for Evidence-based Practices in the Human Services.

**Olaf Rieper** has an MA in sociology from the University of Copenhagen and a PhD from the Copenhagen Business School. He is currently an emeritus research director at the Danish Center for Social Science Research. He has
Conducted evaluations and research on public programs and organizations for Danish and Norwegian government agencies and for the EU Commission. The focus has been on welfare services and its organizations as well as on regional development programs. He has published on research methodology and on the evidence movement. Olaf has been working as a consultant in Ann Arbor, Michigan and in Lyon, France. He has been a co-founder of the Danish Evaluation Society and the first chairman of its board.

Francesco Mazzeo Rinaldi teaches evaluation research and programming courses in the Social and Political Science Department of the University of Catania, Italy. He is also currently Affliate Professor at the KTH, Royal Institute of Technology, School of Architecture and the Built Environment, Stockholm. In the last 15 years he carried out training, research, and consulting activities on program and policy evaluation in several regional, national, and international public organizations in the domains of environmental, social, and R&D policy.

Daniel Suryadarma conducts applied economics research in the areas of education, poverty, and social policy. His work has appeared in peer-reviewed journals, including *American Economic Journal: Applied Economics* and *Journal of Development Economics*. His research has informed policy discussions at the Indonesian Vice President’s Office, Indonesian Ministry of Education and Culture, Australian Aid Program, and the World Bank. Daniel is currently the Deputy Team Leader of the RISE Programme in Indonesia, an education research program managed by the SMERU Research Institute. Daniel is also an assessor for the Australian Research Council and a member of the Indonesia Young Academy of Sciences.

Peter Wilkins is an adjunct professor at The John Curtin Institute of Public Policy at Curtin University in Western Australia and works as an independent consultant. He has served as Western Australia’s Deputy Ombudsman and prior to that had been WA Assistant Auditor-General Performance Review. He has worked in Australia, England, Canada, and Malaysia in roles including engineer, research fellow, public sector manager, and consultant. He has published widely in the fields of accountability, governance, performance, and evaluation.

John Young is the Executive Director of the International Network for Advancing Science and Policy (INASP) in Oxford, UK. He has spent the last 20 years doing research, evaluations, and advisory work on a wide range of topics including research policy linkages; policy engagement and influence; organizational and institutional development; capacity development, government service reform; monitoring evaluation and learning; and strategic communication in more than 20 developing and emerging countries. Before that he spent 20 years in Sri Lanka, Kenya, and Indonesia researching, developing, implementing, and testing new approaches to rural development and decentralized public services.
Foreword

Ray C. Rist

This is an important book for three reasons.

First, transposing refined economic terms like “supply” and “demand” to a new social science like evaluation creates more chaos and confusion than it does clarity and precision. This book documents such confusion and difficulty when it is clear that presumptions of clarity about these terms do not hold up when a transition has taken place. Paper after paper describes the gap between assumptions of what is to be understood by supply and what is presumed by “demand” when economics frames the discussion for evaluation. These terms take on different meanings when such a conversion takes place. Failing to recognize the change in meaning only breeds confusion and a lack of clarity. These papers make clear that evaluation is the discipline that suffers, not economics.

Second, these chapters also demonstrate that evaluation has not sorted out these terms in the context of its own terminology, as opposed to relying on economic understandings. A plethora of meanings is given to both terms by evaluators. Consequently, the gaps in understanding are real and not to be ignored. In these papers, one sees the confusion that evaluators generate about these concepts and the lack of congruence between these two terms. Multiple meanings are attributed to the term “supply” as well as to “demand.” But having said this, it is to be noted that the “gold standard” glossary of 75 evaluation and results-based management terms provided by the OECD in 2002 includes neither word. Consequently, this entire discussion of supply-demand has thus far emerged and evolved within the evaluation communities in just the past 17 years.

Third, due to the complexity of the issue, this book does not venture a simple recipe for how to make supply meet demand. Rather the chapters illustrate on a case by case basis what can go right and what can go wrong. In addition, the book discusses in specific contexts what can be done to close the gap. Multiple contexts are presented here, and thus no one single way forward. Different evaluation questions beget different strategies for answering them. Think of a middle-range matrix. The editors posit that finding a solution that can stand up to the “reappolitik” challenge may involve acknowledging the multiple realities of supply and demand, as well as the gap that will continue to exist into at least the near future.

This is not only an important book; it is also provocative. It merits our time and attention. At one level, it calls into question the ongoing facile use of terms
like supply and demand. We in evaluation simply need to up our game and start using language with more precision and care. At a higher level, this book calls into question a persistent fog in our thinking. It is intellectually sloppy to presume we all know what we mean by terms like “supply” and “demand.” We do not. We must demand more of ourselves. Our profession depends on us to do so.

July 2019
Washington, DC
1 Introduction

Markus Palenberg and Arne Paulson

The discussions that led to this book began with a simpler view of the evaluation world – one in which the “gap” between the demand and supply of results information (highlighted in previous works in this series), was thought to reflect a simple lack of understanding between the demanders of evaluations, and the suppliers of evaluation information, about which “facts” were needed to support credible evaluations. In that context, it was hoped that a “tool kit” could be developed that would bridge the gap between demand and supply.

This view of the world may have been unduly simplistic. As our authors demonstrate in the chapters that follow, there are in fact many types of “demand” for evaluation (some of them more aspirational than pragmatic), while the suppliers of evaluation information may sometimes focus more on their own professional interests than providing useful information for improving policies and programs. The current political climate, which even suggests that “alternative facts” can be used as a convenient substitute for actual facts when desired, merely injects additional confusion into our view of reality and further widens the evaluation gap.

The purpose of this book, therefore, is really to: shine a light on the divergent demands for evaluation (some of which may be more apparent than real); expose the biases behind some seemingly non-partisan sources of evaluation information; and highlight the pros and cons of attempts to bridge the gap between demand and supply of evaluation information through the use of third parties, enhanced stakeholder involvement, and the incorporation of social science models to strengthen Theories of Change (ToC).

Policymakers, program implementers, project managers, and evaluators all feel under increasing pressure to “demonstrate results” in order to justify their respective actions.¹ Policymakers, for example, need to demonstrate to their constituencies that current government policies are indeed on track to achieve their objectives and merit public support. Program implementers, for their part, are eager to show that the sub-programs that underlie government policies are efficient and effective. Project managers at the field level are constantly being exhorted to prove that they are successfully “doing more with less.” And evaluators – the ones who are actually in a position to supply evaluation
information to support (or deny) the claims of higher-level authorities – are under pressure to supply evaluation results.

Thus, the demand for evaluation information about “results” is tiered at different levels, but how consistent are these demands with one another, and with data that is actually available from evaluators within the required time-frame and budget? Why is frustration continuously expressed by those who demand evaluation results when professional evaluators present them with their responses? What explains the “gap” between what those on the “demand” side expect in terms of results, and the “supply” of information provided by evaluators? Can anything be done to narrow this gap? What works and what does not work?

This book examines these questions from both the demand and the supply side.

The Demand Side

Results information is required from a range of different stakeholders for a range of different purposes. One useful categorization comes from literature on results-based management (RBM) and broadly differentiates between learning, decision-making, and accountability purposes (Binnendijk, 2001; Mayne, 2007a, 2007b; Kusek & Rist, 2004; Vähämäki, Schmidt, & Molander, 2011; UNDG, 2011; Vähämäki, 2017). These purposes can be summarized as follows:

- **Learning**: results information is used for personal and organizational learning of people and institutions designing, implementing, and experiencing the effects of policies and programs. Learning also takes place during monitoring, evaluation, and research processes.
- **Decision-making**: results information is used to inform decision-making processes of people and institutions responsible for policies and programs, with the aim of improving performance in achieving results.
- **Accountability**: results information is reported by people and institutions implementing policies and programs to legitimize their choices and resource use, and for ensuring support. Results information is also reported to those people and institutions involved with implementing policies and programs, and to those affected by them, to satisfy their respective learning, decision-making, and accountability requirements.

In addition, demand for results information can also be driven by less rational notions that can be understood in the wider context of the political economy in which public policies and programs are developed and operate. Demand can be insincere and rhetorical if not associated with a push for accurate reporting on outcomes and impacts of policies and programs. Or, demand can be created to promote product marketing for specific evaluation instruments.
The Supply Side

Information on the results of policies and programs is generated in different ways, for example through monitoring, evaluations, audits, and research. Accordingly, there are different types of providers of results information. Volume, type, and quality of results information produced by these means is influenced by many factors that can be sorted into two broad groups: institutional and methodological.

Institutional Factors

Institutional factors are, for example, the degree to which evaluation, audit, monitoring, and research activities are prioritized and supported financially and politically. Lack of financial support can lead to insufficient capacity for generating results information in terms of human resources, skills, and experience, and in terms of systems and processes. With insufficient support and funds, monitoring and evaluation cannot accurately map results of policies and programs in terms of coverage and quality.

Publicly financed institutions also exhibit incentive structures and pronounced organizational cultures that can foster or limit their ability to produce and make use of results information.

Another institutional factor is timing. Public institutions and programs follow political and budget cycles that may hinder longer-term processes aimed at generating results information. In many cases, efforts for generating results information are tied – financially and time-wise – to the programs being examined, possibly leading to short-termism and lack of information on outcomes and impacts occurring after programs have ended.

Methodological Factors

Apart from technical factors, generation of results information is influenced by methodological factors. Methodological factors are related to strengths and limitations of approaches and methods with which results information is generated.

For example, results-based management (RBM) frameworks focus on indicators that go beyond activities and outputs and encounter challenges with respect to regular monitoring and causal attribution. Welfare economics approaches, for example Cost-Benefit-Analysis (CBA), have successfully been applied to interventions in many sectors but encounter difficulties when benefits are difficult to quantify. Experimental approaches to impact evaluation such as Randomized Controlled Trials (RCTs) are useful for investigating causal mechanisms in specific contexts but struggle with broader application due to their inherent inflexibility. Other evaluation methodologies such as Contribution Analysis (CA), Process Tracing (PT) or Qualitative Comparative Analysis (QCA), and mixed method approaches have been employed successfully to investigate contribution claims and to understand complex causal relationships but are not widely applied and usually cannot answer questions on how much an intervention has contributed to intended results.
Book Outline: Exploring the Gap Between Demand and Supply

The authors of this book describe ten different examples of the gap between demand and supply of evaluation information in different contexts, and attempts made to bridge that gap.

Ida Lindkvist and Per Øyvind Bastøe question how genuine the demand for evaluation results really is at higher levels and in the public at large, when it comes to Norwegian development assistance. When contrasted with actual budget expenditures on evaluation, the authors conclude that much of the supposed demand for evaluation results is actually “inflated” or “fake” demand.

At the other end of the spectrum, Agathe Devaux-Spatarakis describes a situation in which the “suppliers” of evaluation information took control of evaluation planning for certain experimental social programs in France and succeeded in forcing project managers and program implementers into accepting Randomized Control Testing (RCT) as the only legitimate evaluation tool. RCTs however require controlled and stable conditions which was at odds with the need of program managers to adjust programs during implementation. In the end, most results information generated by RCTs was not used meaningfully.

Can an intermediary organization help in bridging the gap in expectations on the demand side, and what is practically feasible on the supply side? Richard Boyle reports on the experience of the non-profit Atlantic Philanthropies in the Republic of Ireland, where it aimed at prompting government engagement with results information in a more systematic way. These efforts met with some success until the charity began to wrap up its operations. Will those efforts continue to bear fruit without outside support? Or will government budget cuts of necessity reinforce the demand for evaluation results to demonstrate that remaining funds are wisely spent?

A somewhat different take on third-party involvement with government interventions and their evaluations is provided by David Gilchrist and Peter Wilkins in their chapter on the use of not-for-profit and voluntary organizations to deliver human services to support social policy objectives in South Australia. There were several reasons for adopting this approach, including the fact that it was seen as a cheaper source of service delivery, more flexible, and – because those organizations are closer to the communities in which they work – have more opportunities for identifying needs and demands, and for assessing the relative impact of programs. On the other hand, their analysis notes that the voluntary and not-for-profit organizations have been challenged with insufficient resources, increasing pressure to perform more with less in the face of changing funder priorities, and impacting their ability to pursue their mission. Despite these challenges, a key body in the public services sector concluded that a lot had been learned about the procurement of outcome-based services and the importance of involving service users as well as providers in the co-design and co-evaluation of services. The authors remain hopeful that voluntary organizations and governments will be keen to see the process through.
Olaf Rieper’s chapter provides a detailed review of how an intermediary organization was usefully incorporated into the evaluation itself, of the Norwegian health care reform program in 2012–2015. The author, who was chairman of the intermediary organization (the Research Council of Norway) at the time, is in a unique position to describe the inner workings of the council, its Evaluation Steering Committee and advisory groups, and the organizational and cultural aspects which contributed to a seamless evaluation process with no significant gaps observed between what was expected, and what was obtained, from the evaluation. Although it is too soon to judge some of the long-term results of the health reform program, and despite the difficulty of separating out the impact of the health reform from other parallel initiatives, the author concludes that the Norwegian health reform evaluation provides a good practice example of how the involvement of an intermediary organization can harmonize expectations between demand and supply in the evaluation process.

Other chapters also emphasized the importance of stakeholder involvement in the design and evaluation of interventions. Francesco Mazzeo Rinaldi, for example, describes an evaluation done for the research and development (R&D) component of an EU-sponsored regional development program in Sardinia, Italy. The organizers of that evaluation included institutions at the national, regional, and municipal levels, various EU funds, and experts in the fields selected for evaluation, academics, and individuals from the public and private sector. The result was a highly complex evaluation setup designed to assure stakeholder participation, not only with regard to the planning and implementation aspects, but also in terms of the evaluation of the program. The author concludes that the complex arrangements helped reduce unrealistic expectations about the evaluations. Unfortunately, the failure to pay sufficient attention to the “supply” side of evaluation information – notably the lack of baseline data and inadequate data collection during implementation – made it impossible to answer even the most basic evaluation questions.

Brian Belcher, John Young, and Daniel Suryadarma report on another complex evaluation effort emphasizing stakeholder participation. Noting the high demand from funders, research users, research managers, and researchers themselves for better evaluations for improved accountability, decision-making and learning purposes, they describe and analyze the evaluation of CIFOR’s Climate Change Research program. The chapter discusses the factors driving demand for results information and analyzes how its supply can be further improved. Given that multiple actors are involved and that the “results chain” incorporates long time-lags and numerous feedback loops, quantitative evaluation methods were judged not to be appropriate for many of the kinds of projects and programs that CIFOR undertakes. The evaluation consequently focused on demonstrating results in terms of: influencing international dialogue and behavior of national policymakers; use of research results by stakeholders; and frequent citation of CIFOR publications, books, and policy papers – each of which underscored the plausibility of their Theories of Change (ToC) model. The authors observed that the process of developing the ToC was, in itself, a valuable learning experience,