



Comparative Policy Evaluation

THE POWER OF COMMISSIONING

ENABLING DEMOCRATIC EVALUATION

Edited by
Nicoletta Stame and María Bustelo



The Power of Commissioning

Far more than a contractual step, commissioning shapes what is evaluated, whose voices matter and how evidence informs democratic governance. *The Power of Commissioning: Enabling Democratic Evaluation* brings long-overdue attention to this influential yet often-overlooked dimension of evaluation practice.

Positioning commissioners as civic and political actors, the book shows how their choices, from problem framing and terms of reference to team selection, management and use can either reinforce technocratic routines or actively promote inclusion, transparency, equity and meaningful public learning. Drawing on comparative empirical cases, it offers one of the first integrated frameworks for democratic commissioning, complete with practical principles, rubrics and clear “dos and don’ts” that commissioners can apply directly in real systems. The volume also advances a pioneering competency model for commissioners and documents innovative capacity-building initiatives that treat them as knowledge brokers and democratic agents, an area largely absent from existing evaluation guidance.

Featuring contributions from leading scholars and practitioners in the field, *The Power of Commissioning* provides both conceptual depth and actionable tools. It is an essential resource for researchers, policymakers, evaluators and commissioning professionals committed to strengthening democratic practice through evaluation.

Nicoletta Stame is Emerita Professor of Social Policy at “Sapienza” University of Rome. She is a past president of the Italian Evaluation Association and of the European Evaluation Society and associate editor of *Evaluation*. Nicoletta is interested in the theory and methods of evaluation, on which she has published widely in Italian and international journals and edited books.

María Bustelo is Professor of Political Science and Public Administration at the Complutense University of Madrid (UCM), where she has directed the Master on Evaluation of Programmes and Public Policies since 2002. She is an EES past president (2012–2013) and an Executive Board member of the Spanish Evaluation National Agency (AEVAL) (2007–2012). She specializes in evaluation, gender and institutionalization.

Comparative Policy Evaluation

Edited by Andrew Koleros, Frans L. Leeuw, Ida Lindkvist, and 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.

33. Evaluation in the Post-Truth World

Edited by Mita Marra, Karol Olejniczak, Arne Paulson

34. Theories of Change in Reality: Strengths, Limitations and Future Directions

Edited by Andrew Koleros, Marie-Hélène Adrien, Tony Tyrrell

35. Artificial Intelligence and Evaluation: Emerging Technologies and Their Implications for Evaluation

Edited by Steffen Bohni Nielsen, Francesco Mazzeo Rinaldi, Gustav Jakob Petersson

36. From Algorithms to Evidence: Using GenAI in Evaluation Practice

Edited by Kerry Bruce, Valentine J. Gandhi, and Steffen Bohni Nielsen

37. The Power of Commissioning: Enabling Democratic Evaluation

Edited by Nicoletta Stame and Maria Bustelo

38. From Silos to Ecologies: Towards an Ecology of Evidence to Navigate Solutions

Edited by Sanjeev Sridharan and Frans L. Leeuw

The Power of Commissioning

Enabling Democratic Evaluation

**Edited by Nicoletta Stame and
María Bustelo**



Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2027
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2027 selection and editorial matter, Nicoletta Stame and María Bustelo;
individual chapters, the contributors

The right of Nicoletta Stame and María Bustelo to be identified as the authors of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

The Open Access version of this book, available at www.taylorandfrancis.com, has been made available under a Creative Commons Attribution-Non Commercial-No Derivatives (CC-BY-NC-ND) 4.0 international license.

Any third party material in this book is not included in the OA Creative Commons license, unless indicated otherwise in a credit line to the material. Please direct any permissions enquiries to the original rightsholder.

For Product Safety Concerns and Information please contact our EU representative
GPSR@taylorandfrancis.com. Taylor & Francis Verlag GmbH, Kaufingerstraße 24,
80331 München, Germany.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

ISBN: 9781041361237 (hbk)
ISBN: 9781041361244 (pbk)
ISBN: 9781042000968 (ebk)

DOI: 10.4324/9781042000968

Typeset in Times New Roman
by Newgen Publishing UK

Contents

<i>List of contributors</i>	<i>viii</i>
<i>Foreword</i>	<i>xiii</i>
RAY C. RIST	
<i>Preface</i>	<i>xv</i>
<i>Acknowledgments</i>	<i>xvii</i>
1 Commissioning matters: challenges for evaluation commissioners who intend to promote democracy	1
NICOLETTA STAME AND MARÍA BUSTELO	
PART I	
Commissioning (with democratic principles) matters	23
2 How to make commissioning matter in times of complex fast-changing reality	25
BURT PERRIN AND NICOLETTA STAME	
3 How evaluation commissioning processes can support or undermine democratic processes	37
PATRICIA ROGERS AND STEVE MONTAGUE	
4 Commissioning with clarity: empowering democratic spaces by drawing on the principles of deliberative democracy	49
TOM LING AND ANANDA MILLARD	
5 How evaluation commissioners and evaluators can collaboratively move equitable evaluation from a set of external actions to internal practices	63
DREW KOLEROS, MONIQUE LISTON AND RAY MCGHEE	

PART II

Experiences of commissioning in different institutional contexts 77

- 6 The influence of context for different types of commissioners of evaluation: challenges and opportunities 79

TESSIE TZAVARAS CATSAMBAS

- 7 Democracy, SDGs, evaluation and commissioners: connecting the dots 89

ROBERT LAHEY AND DOROTHY LUCKS

- 8 Learning Agendas: from studies to streams of evidence 102

KAROL OLEJNICZAK, TOMASZ KUPIEC AND YAERIN PARK

- 9 Democratic implications of commissioning market-based evaluation: insights from Italy's Cohesion Policy 116

MITA MARRA AND VIRGILIO BUSCEMI

- 10 Evaluation Management Services at the MFA of Finland 132

SARI LAAKSONEN AND MARKUS PALENBERG

- 11 The art of commissioning: lessons from commissioner–evaluator collaboration 146

IDA LINDKVIST AND ROB LLOYD

PART III

Educating the commissioner, through a dialogue between commissioners and evaluators 159

- 12 The good commissioner: competencies for enhancing democracy in commissioning processes 161

MARÍA BUSTELO, RAMÓN CRESPO AND JUAN ANDRÉS LIGERO

- 13 Training for critical evaluative thinking 177

PER ØYVIND BASTØE, ØYUNN SYRSTAD HØYDAL AND
KNUT FOSSESTØL

- 14 Academy of Evaluation: a capacity-building initiative for commissioning authorities 189

KAROL OLEJNICZAK, MARLENA STĘPIEŃ AND
PIOTR STRZĘBOSZEWSKI

PART IV

Conclusions **205**

15 Democratic commissioning: lessons from theory and practice 207

MARÍA BUSTELO AND NICOLETTA STAME

Appendix: Applying insights from the model of the Neo-Weberian
State to commission more democratic and valuable evaluations 231

ALISON POLLARD

Index 243

Contributors

Per Øyvind Bastøe has held senior positions in the Norwegian central government and in major multilateral organizations. He is known for promoting the use of evaluation to strengthen public sector learning and accountability. He now works as an independent evaluation advisor, sharing his experience across countries and institutions.

Virgilio Buscemi graduated in Philosophy, with a second-level Master's degree in Social Impact Analysis and Evaluation, and 30 years of experience in public policy evaluation. He has coordinated over 75 evaluation projects concerning programs co-financed by EU funds (Cohesion Policy and the CAP), as well as Health, Security and Legality, Cooperation, etc.

María Bustelo is Professor of Political Science and Public Administration at the Complutense University of Madrid (UCM), where she has directed the Master on Evaluation of Programmes and Public Policies since 2002. She is an EES past president (2012–2013) and an Executive Board member of the Spanish Evaluation National Agency (AEVAL) (2007–2012). She specializes in evaluation, gender and institutionalization.

Tessie Tzavaras Catsambas is founder and CEO of EnCompass LLC, a 25-year-old organization working in evaluation, learning and leadership development. She is past American Evaluation Association President. She has published books and chapters on appreciative evaluation, evaluation management, competencies and ethics. She holds an MPP from Harvard University.

Ramón Crespo has taught and worked as an independent evaluator since 2006, specializing in youth policy and community development, research and public sector program evaluations. He is in the academic team at the Master on Evaluation of Programmes and Public Policies at the Complutense University of Madrid (UCM) and lectures in evaluation training programs at several Spanish universities. Since 2023, he has served as President of APROEVAL (Spanish Association for Evaluation Professionals).

Knut Fossetøl is an evaluation researcher specializing in public-sector reform, governance and competence development. His work focuses on program theory,

hybrid governance models and learning-oriented evaluation, with extensive experience from commissioned studies in education, labour and welfare policy.

Øyunn Syrstad Høydal is a sociologist and holds a PhD in social policy. Her research on evaluation includes aspects such as evaluation and power, the co-production of knowledge between evaluators and knowledge users, and the interaction between evaluation as a symbol and as a method.

Drew Koleros is a systems change evaluator with 20+ years' experience designing mixed-methods evaluations. He specializes in theory-based, systems-oriented approaches, including actor-based theories of change and system mapping, supporting philanthropic and public-sector initiatives in sectors including climate, health and education, and translating evidence into actionable strategy with participatory learning methods.

Tomasz Kupiec is Assistant Professor at the University of Warsaw and Deputy President of the Polish Evaluation Society. He has worked in regional and central public administration, commissioning evaluations. His research focuses on the performance of evaluation systems, development of evaluation practice and organizational determinants of evaluation use.

Sari Laaksonen, MBA, has over 25 years' experience in development policy and cooperation, with over a decade in international politics and diplomacy at the UN, WTO and MFA Finland. Sari is a seasoned evaluator, evaluation team leader, and QA and MEL expert with solid experience in utilization-focused and strategic evaluation with learning and participatory approaches and futures literacy/foresight methods.

Robert Lahey was the founding head of Canada's Centre of Excellence for Evaluation. Following his public sector career that also included heading corporate evaluation functions in four federal agencies, he has been an independent advisor to many countries and international organizations on developing national monitoring and evaluation and reporting systems, ECD and the SDGs. A Fellow of the Canadian Evaluation Society (CES), he has written and presented extensively on his work to the international evaluation community.

Juan Andrés Ligeró is a sociologist specialized in Evaluation and Research Methodology. He is Co-Director of the Master on Evaluation of Programmes and Public Policies at the Complutense University of Madrid (UCM), Lecturer in evaluation at Universidad Carlos III Madrid and CEO of "Means Evaluación", a consulting agency specialized in evaluation.

Ida Lindkvist is Policy Director for Evaluation at the Norwegian Directorate for Higher Education and Skills, working on evaluation capacity building. She holds a PhD in economics and has over ten years of experience in evaluation. Her research interests include organizational learning and results-based management and its institutional effects, and she has published on global health, sustainability and the political economy of evaluation.

Tom Ling has led over 100 evaluations for many clients including the EC, UN Agencies and UK Government departments. He was an advisor to the Global Evaluation Initiative, he is a past president of the European Evaluation Society and now advisor to the Board, and in 2025 he received a Global Impact Award for Evaluation from the Africa-led M&E Bench. Tom has published widely on evaluation in theory and practice.

Monique Liston is the Founder and Chief Strategist of UBUNTU Research and Evaluation, a Milwaukee-based, Black-led learning organization. An Afrofuturist evaluator, facilitator and writer, she advances dignity-centered evaluation, collective learning and organizational transformation across philanthropy, public health, education and cultural systems, centering Black aliveness, community accountability and liberatory futures.

Rob Lloyd is the Managing Partner of Itad, a global social impact consultancy. Rob is a passionate advocate of the use of evidence and learning to improve social impact, and has spent the past two decades advising governments, philanthropy and multilateral agencies on how best to do this.

Dorothy Lucks is the Executive Director of SDF Global Pty Ltd. She has a PhD in sustainable development, is a credentialled international evaluator and evaluation Fellow with over 25 years of experience as a senior evaluator and strategic analyst with global institutions and in complex situations.

Mita Marra (MA-SAIS98; PhD-GWU03) is Professor of Economic Policy at the University of Naples in Italy. Visiting fellow at Maastricht University, George Washington University and the London School of Economics, she has served as President and Board member of the Italian and European Evaluation societies, and as Editor-in-Chief for *Evaluation and Program Planning*.

Ray McGhee is a research professional with over 20 years' experience leading research, evaluation and analysis across education, youth development and community development. He recently served as Senior Program Officer at the Robert Wood Johnson Foundation, partnering with grantees to advance health and racial equity through applied research initiatives.

Ananda Millard is an evaluator with over 25 years of experience in research, capacity development and evaluation across multiple sectors and themes in development aid and humanitarian assistance. Methodologist with extensive field experience in the Global South, specialized in data collection, analysis, synthesis and interpretation of data from hard-to-reach populations.

Steve Montague is a career evaluation practitioner, adjunct professor at Carleton University, a research associate at Ottawa University, and a fellow of the Canadian Evaluation Society. He specializes in theory-based mixed-methods evaluations, realist approaches and linking evaluative thinking to performance planning, measurement, reporting and management.

Karol Olejniczak, PhD, is the Head of the Center for Policy Design and Evaluation at SWPS University and a co-founder of the Polish research company EGO – Evaluation for Government Organizations. His work focuses on program evaluation, the use of evidence in policy design, and the application of games and behavioral insights into public policy.

Markus Palenberg is Founder and Manager of the Institute for Development Strategy (IfDS) in Munich, Germany. He works as a researcher, evaluator and advisor, specializing in comprehensive thematic and institutional evaluations of development cooperation and the advancement of evaluation methodologies.

Yaerin Park is Assistant Professor in Public Service and Public Policy at Hainan University (HNU)–Arizona State University (ASU) Joint International College (HAIC), Hainan University. Her multidisciplinary research interests include participatory budgeting (PB), collaborative governance, social equity budgeting, digital transformation of tax administrations and design thinking in public administration.

Burt Perrin is a thinker, author and independent consultant with more than 40 years' practical experience assisting governments and other organizations internationally. He has made invited presentations at settings around the world. His recent publications include "Changing bureaucracies: Adapting to uncertainty, and how evaluation can help" (edited with Tony Tyrrell) and "Theories of change: Who needs them? Or: What evaluators can learn from opera" (book chapter).

Alison Pollard has 15 years' experience of designing evaluation strategies and evaluating policies, programs and institutions for national governments, international organizations and grant-making foundations. She is committed to inclusive development, delivering evaluations that empower and build capacity, and working with others to use evidence in ways that improve people's lives.

Ray C. Rist completed his PhD in Urban Sociology in 1970 at Washington University, St. Louis. He has written 17 books and edited 19 others. He has also authored more than 150 articles. He has been a professor at four universities and worked in the US congress. He also worked for the World Bank.

Patricia Rogers is Former Professor of Public Sector Evaluation at the Royal Melbourne Institute of Technology (RMIT University), Australia; Founder of BetterEvaluation, free access knowledge platform on evaluation methods and processes; co-founder of the Footprint Evaluation Initiative, which supports the inclusion of environmental sustainability in all evaluations; Fellow of the Australian Evaluation Society.

Nicoletta Stame is Emerita Professor of Social Policy at "Sapienza" University of Rome. She is a past president of the Italian Evaluation Association and of the European Evaluation Society and associate editor of *Evaluation*. Nicoletta

is interested in the theory and methods of evaluation, on which has published widely in Italian and international journals and edited books.

Marlena Stępień is Chief Specialist at the National Evaluation Unit, Ministry of Development Funds and Regional Policy, Poland. Marlena is a civil servant, coordinator of evaluation studies in social and territorial development, responsible for event organization and the use of new technologies in evaluation projects. Marlena is a graduate of the University of Warsaw (Centre for European Regional and Local Studies – EUROREG) and Adam Mickiewicz University in Poznań (Faculty of Law and Administration).

Piotr Strzęboszewski is Head of the National Evaluation Unit, specializing in public policy evaluation and evaluation systems. For nearly 20 years, he has focused on evaluating European Cohesion Policy. His work emphasizes methodological rigor, institutional learning and evidence use in governance, advancing evaluation frameworks that support accountability, effectiveness and decision-making.

Foreword

Ray C. Rist

This book is profound and potentially groundbreaking. It calls on the evaluation community to reframe itself with a single message: to restructure as a democratic process. How? By means of *democratic commissioning*.

The authors – speaking from three continents – point to a large and important gap in the evaluation architecture. Traditionally, evaluation has concentrated on just four main topics: (1) the evaluator and his/her values or ethics; (2) the focus or design of the evaluation; (3) the methodology used for the study; and (4) the final use or dissemination of the report. What’s missing is attention to the process of democratic commissioning.

As the editors state in the introduction:

Commissioning is a relationship between two parties – the commissioner and the evaluator – that unfolds along a process: from decisions about what to evaluate, to Tors (terms of reference), selection of evaluator, budgeting, contract, up to final use. The contract is a consequence of that relationship, and its rules should not dominate the relationship itself. The object of the commissioned evaluation... is an intellectual and practical service referred to as an intervention aiming at the improvement of a situation.

They go on to analyze the three headings in the encounter between commissioning and democratic principles: exertion of agency, approaches to evaluation and evaluation systems.

Throughout the book, the role of the commissioner is elaborated as follows:

Commissioners must define priorities, ensure methodological relevance, engage stakeholders, and promote the use of evaluation results. In the end, commissioners are not passive managers, but active agents of change who influence evaluation quality, independence, and democratic contribution.

The 13 chapters by multiple authors who contributed to the book, in addition to the introduction and conclusion provided by the two editors, flesh out a discussion of “democratic” evaluation, distinguishing it from both “bureaucratic” and

“autocratic” forms of evaluation offered by MacDonald (1977). Specifically, it is a bureaucratic evaluation when the evaluator acts as a “management consultant”; autocratic when the evaluator “offers external validation of policy”; and democratic when the evaluator “acts as a broker in exchanges of information between different groups of informed citizens”.

All in all, this is a tightly reasoned book that is clear, concise and adds to our knowledge. And it absolutely addresses the profound gap in our understanding about the commissioning process.

Reference

MacDonald, Barry. 1977. “A Political Classification of Evaluation Studies in Education.” In *Beyond the Numbers Game: A Reader in Educational Evaluation*, edited by David Hamilton, David Jenkins, Christine King, Barry MacDonald, and Malcolm Parlett, 105–108. London: Macmillan.

Preface

This book was conceived in 2023, a whole geological era ago. At that time, the world was fighting against climate change, the growing social divide and the unethical use of technological innovations. The evaluation community was striving to rise to such challenges, by devising tools to empower marginalized people, and ensure the success of sustainability policies aimed at the public good. There was a global consensus on the 17 SDGs, although their difficulty of achievement was well recognized.

In that context, our attention had been devoted to the very process by which evaluation takes its rationale, that is, commissioning. Believing that democratic principles should characterize the entire evaluation process, we realized that commissioning was a highly relevant yet overlooked area. To overcome this, we concluded that studies of the institutionalization of evaluation (a process that had been accomplished almost universally) were insufficient. We wanted to look at commissioners' agency, how they collaborate with other partners and stakeholders, and what are their strategies to make commissioning a democratic process.

The result is the book that you have in your hands.

Yet, in the meantime, the world has changed. The current situation is characterized by the spread of autocratic, not to say dictatorial, political regimes, that suppress programs for improvement (like the SDGs and development aid) and, consequently, any evaluation activity.

In this new situation, is our motivation still valid? Are we an outdated nostalgic group of people with no sense of reality? Or might our understanding of the topic help us to navigate the current predicaments?

As Schwandt & Gates (*Evaluating and Valuing in Social Research*, 2021: 1) wrote, "individuals, organizations, and communities hold values implicitly and make them explicit only when challenged". These are challenging times and we need to discuss, elaborate on, and make explicit the democratic values surrounding evaluation and the commissioning processes it entails more than ever.

Indeed, we believe that our collective research into the relationship between evaluation (in all its aspects and components) and democracy, and the range of differences proposed by various typologies and rubrics, as set out in the book, have offered clues even for facing situations where democratic values are under

attack. Furthermore, as will become clear in the Conclusions (Chapter 15: Bustelo & Stame, 2027), we believe that our approach can encourage those who live in democratic societies to support those less fortunate, as well as inspire those in autocratic regimes to devise strategies for expanding their own range of action.

The place of this book in the INTEVAL collection

Many of the themes dealt with in this book are not new to the INTEVAL community. As a function exercised by public officials within an institution, or by managers within an organization, commissioning has made the object of previous books devoted to the institutionalization of evaluation.

In *The Evaluation Enterprise* (Furubo & Stame, 2019), a chapter devoted to the industry of evaluation (Bohni-Nielsen et al.) analyzes the various forces that shape the market for evaluation services: commissioning acts on the demand side alongside buyer power, supplier power, competitive rivalry, the threat of substitution and the threat of new entry. In *Changing Bureaucracies* (Perrin & Tyrrell, eds., 2020), a chapter by Raimondo and Leeuw is devoted to the bureaucratic capture of evaluation, that is, on the perverse effects of institutionalizing evaluation: “distortion” and “undermining” of evaluators work, thus “preventing evaluation systems from solving the issues for which they were created in the first place” (p. 148). In this latter book, commissioning is considered only as part of the bureaucracy that supports – and conditions – the evaluation systems. Indeed, the book is premised on the idea that evaluation can help bureaucracies adapt to the uncertainties and challenges in current policies. Its focus is on improving bureaucratic performance of bureaucracy, that is, on evaluating bureaucracy.

Our book takes a different approach. Firstly, it delves deeper on the specific tasks and characteristics of commissioning, distinguishing it from other bureaucratic and managerial roles, and exploring its connection with other stakeholders. In a sense, it emphasizes on the autonomous role that commissioners may play within the organization as a whole, on their agency. Secondly, and perhaps more importantly, it shifts the focus from evaluators’ overt search for infusing democratic principles into a collaborative process with commissioners, who might otherwise drift toward conservatism.

References

- Furubo, Jan-Eric, and Nicoletta Stame, eds. 2018. *The Evaluation Enterprise: A Critical View*. New York and London: Routledge.
- MacDonald, Barry. 1977. “A Political Classification of Evaluation Studies in Education.” In *Beyond the Numbers Game: A Reader in Educational Evaluation*, edited by David Hamilton, David Jenkins, Christine King, Barry MacDonald, and Malcolm Parlett, 105–108. London: Macmillan.
- Perrin, Burt, and Tony Tyrrell, eds. 2020. *Changing Bureaucracies: Adapting to Uncertainty, and How Evaluation Can Help*. New York: Routledge.
- Schwandt, Thomas A., and Emily F. Gates. 2021. *Evaluating and Valuing in Social Research*. New York: Guilford Press.

Acknowledgments

This volume has been conceived within a community of scholars, INTEVAL, that in its long existence had analyzed evaluation from many different perspectives. When the concept of our project emerged at the Rome meeting of 2023, it was considered with skepticism by some of them, but it got the endorsement of a large group of others, the current authors. As (a lot of) time went by, though, the topic of the book came under the radar of many more, who shared their feelings about it, from their professional and institutional point of view. Their suggestions and opinions have been incorporated into the book. We owe special thanks to the evaluator-commissioners among our INTEVAL colleagues who have answered our questions about their remarkable dual experience: Kim Forss, Hur Hassnain, Frans Leeuw, Jos Vaessen, Rob van den Berg. Their contributions have been invaluable in deepening our understanding of the challenges faced by commissioners and their views have strengthened our motivation, and at the same time oriented our research toward new directions.

As is usual, the contents of all chapters have been discussed among the group of authors in dedicated sessions at the annual meetings and in written communications. Some authors have been in contact with us all through the book process, beyond their chapters: Tessie Tzavaras Catsambas, Juan Andrès Ligeró, Tom Ling, Burt Perrin, Patricia Rogers. We acknowledge their contribution with gratitude.

Making this book available in Open Access has been possible thanks to a collective effort by the authors and the support of several institutions and individuals committed to the dissemination of knowledge.

We would like to express our sincere gratitude to the Independent Office of Evaluation of IFAD for its contribution, and especially to its Director and INTEVAL colleague, Indran Naidoo, for his generous support of one of the chapters. Among coauthors, we are also grateful to Karol Olejniczak, who secured funding from SWPS University in Warsaw for two chapters; to María Bustelo, who contributed through her research funds at the Complutense University of Madrid for two additional chapters; and to Øyunn Syrstad Høydal and Knut Fossetøl, who obtained support for their chapter through OsloMet – Oslo Metropolitan University. We also wish to thank Dorothy Lucks for contributing through her company SDF Global, as well as APROEVAL – Iberian Association of Evaluation Professionals, through

Ramón Crespo, a member of its Executive Board. We are deeply appreciative of the efforts invested in securing these resources, knowing the time, commitment, and persistence such work entails.

The remaining costs required to make the entire volume Open Access were collectively covered through personal contributions from several of the book's coauthors. We are immensely grateful for this remarkable demonstration of collegiality and shared commitment. Our heartfelt thanks go to Tessie Catsambas, Drew Koleros, Tom Ling, Monique Liston, Ananda Millard, Steve Montague, Markus Palenberg, Alison Pollard, Patricia Rogers and Nicoletta Stame.

Finally, we would like to acknowledge Routledge's flexibility and support throughout the process, particularly through our editor, Natalja Mortensen.

This book is a testament to the collaborative spirit of the evaluation community, and we are deeply thankful to everyone who helped make its Open Access publication possible.

1 Commissioning matters

Challenges for evaluation commissioners who intend to promote democracy¹

Nicoletta Stame and María Bustelo

Flexibility and diversity are preferable to the rigidity written in too many evaluation contracts.

Cronbach (1980), thesis 42

The motivation for this book

This project began due to concerns that evaluations had become routine exercises, disconnected with the primary challenges faced by programs and policies. Faced with global policies such as the Sustainable Development Goals (SDGs) and socio-ecological transformation, on the one hand, and evaluation approaches that address complexity and uncertainty on the other, we were concerned that the demand for evaluations was unable to benefit from these new sensibilities, which we consider to be linked to the democratic mission of evaluation. There was a sense of uselessness in the rigid, repetitive way evaluations were often requested – as Cronbach had already signaled earlier on – and a missed opportunity for collaboration between commissioners and evaluators (Brousselle, 2022; Rogers, 2022).

We think of evaluation and democracy as closely connected, from Chelimsky (2006)’s “purposes of evaluation in a democratic society” to Schwandt (2023)’s “democratic purpose of evaluation for the public good” (where evaluation is considered as a “civic science”, working as a “a resource for civic agency”). Evaluation is an expression of a democratic society, and is rooted in democratic values (independence, transparency, fairness, trust, inclusiveness: see below for definitions and distinctions). Following Jennifer Greene in her 2006 chapter on “Evaluation, Democracy, and Social Change”

From almost the beginning of the contemporary history of program evaluation, there have been theorists and practitioners who (...) fully embrace the intertwinement of values with evaluative practice. Moreover, these evaluators go beyond a value relative stance, which acknowledges and engages the plurality of values that inhabit evaluation contexts, to a value committed stance (...). The

2 *The Power of Commissioning*

most defensible values to promote (...) are those intrinsic to political democratic ideals, namely, social justice, equality, empowerment, and emancipation. (Greene 2006:118)

Therefore, democratic principles should underpin evaluation practice, and evaluation can clarify the values that characterize democracy in the practice. However, if we conceive evaluation to be a strong supporter of democracy, then all the actors involved in the evaluation process should be included.² Having identified the flaws in the demand for evaluation, we have decided to focus on the commissioning process to determine the scope of influence, and the space of leverage, of evaluation commissioners who intend to promote democracy. Through examples and reflections on practice, we aim to encourage their initiative.

Starting from the assumption that **commissioning matters**, we realized that not enough attention had been devoted to the way commissioning took place, not to speak of good commissioning, and we inquired into what was needed for the commissioning process to be embedded into **democratic principles**. This, by the way, led us to a never-ending interrogation about what idea of democracy could match the commissioning process we had in mind, each chapter providing definitions contextually relevant.

First, however, we must consider why commissioning is **an overlooked topic** in evaluation literature and education. Despite the foresight of Cronbach's (1980) "Ninety-five Theses" on evaluation, seven of which were expressly devoted to commissioning,³ very few scholars have devoted more than a brief mention to the topic. Rare contributions have taken commissioning seriously (Perrin & Boschen, 1991, Catsambas & Davidson, 2024, Cox & Barbrook-Johnson, 2021). Even the interesting literature on politics and evaluation (see Datta, 2011) has developed without considering the relationship between policymakers and commissioners.⁴ Similarly, few evaluation guidelines refer to commissioners, and when they do, it is mostly from the evaluator's perspective. Few professional training courses are devoted to commissioning.

Cronbach's 95 theses

- 9- Commissioners of evaluation complain that the messages from evaluations are not useful, while evaluators complain that the messages are not used.
35. "Evaluate the program" is often a vague charge because a program or a system frequently has no clear boundaries.
37. A good evaluative question invites a differentiated answer instead of leaving the program plan, the delivery of the program, and the response of clients as unexamined elements within a closed black box.
42. Flexibility and diversity are preferable to the rigidity written into many evaluation contracts.

- 48. Nothing makes a larger difference in the use of evaluations than the personal factor – the interest of officials in learning from the evaluation and the desire of the evaluator to get attention for what he knows.
- 73. In typical federal contracting, many basic research decisions are made without consulting the evaluators who will do the work.
- 76. Evaluation contracts are increasing in size, but tying many strands into a single knot is rarely the best way to get useful information.

There are different reasons that may account for this lack of attention. One is that the role of commissioning may be considered unimportant, a mere administrative task to contract the evaluation, disregarding the responsibility that commissioners have to take on, all along its process. Another is that commissioners may be considered as conditioned by the evaluation systems they are embedded in, as in the idea of “evaluation capture”, reducing the possibility of commissioners’ strategies of resilience to that trend.⁵ Both resent of a pessimistic vein.

Our perspective is different. When we talk about commissioning, we want to focus on the **commissioning process**, referring to the way in which an organization or institution (a department or someone who has been given the competence to do so) develops an idea or a frame for doing an evaluation and starts a process in which that evaluation is asked, ordered, mandated, requested. It is a relevant process of decision-making, not a simple administrative task of signing a contract. “Commissioning” is broader than “contracting” although they are often confused (and in legalistic parlance the policymaker is called “the client” and the evaluator “the contractor”). Commissioning is not only about preparing Terms of Reference (ToRs) and contracting out evaluation’s teams: it starts much earlier and goes beyond the act of contracting. In this process, different actors, such as commissioners, evaluators, evaluation users (policymakers) and other stakeholders interact, and they can do it in different ways, which can (or cannot) promote those principles that, from our perspective, make evaluation develop in its full potential and democratic mission. This differentiation between a more ample concept of commissioning and contracting, this one being only a part of the commissioning process, allows us to consider internal commissioning processes too. Therefore, we make it clear that if there is no evaluation’s outsourcing, this does not mean that there are no commissioning processes.

Commissioning usually takes place internally, within the organization; less frequently, it may be contracted out, the whole process or part of it, as we see in Chapter 10 (Laaksonen & Palenberg, 2027). If done internally, commissioners may be operatives internal to the organization who add this function to other ones, or members of evaluation units. If outsourced, commissioners are members of research centers to which the evaluation has been externalized. These differences also account for differences in commissioning personnel, that Cox & Barbrook-Johnson (2021) have identified as social scientist (interested in methods), policy

4 *The Power of Commissioning*

(organization people), procurement people, the latter having a comparatively greater impact on the fate of the evaluation (see Chapter 9: Marra & Buscemi, 2027).

Not all evaluations are contracted out. Big international agencies where evaluations are mandated have their own “independent evaluation units”, which manage the evaluation. The World Bank, for example, has a tripartite system of in-house evaluations, wholly outsourced evaluations, hybrid (in-house and outsourced) evaluations.⁶ These differences may affect the possibility for commissioners to interact with external stakeholders, discussing the assumptions of the theory of change, and in general influencing the process started with the original evaluation demand.

On the methodology used for editing this volume, apart from the literature review and the comparative analysis of the different chapters in this volume, we prepared a written questionnaire with open questions for our INTEVAL colleagues who have had experience as commissioners.⁷ Each chapter adds a methodological note about how the chapter was constructed.

What is relevant about commissioning

Commissioning is a relationship between two parties – the commissioner and the evaluator – that unfolds along a process: from decision about what to evaluate, to ToRs, selection of evaluator, budgeting, contract, up to final use (Chapter 2: Perrin & Stame, 2027). The contract is a consequence of that relationship, and its rules should not dominate the relationship itself.⁸

The object of the commissioned evaluation (and therefore of the contract in case the evaluation is contracted out) is an intellectual and practical service referred to an intervention aiming at the improvement of a situation. It is not comparable to the procurement of an infrastructure or a service, for which the quality standards are known in advance. It is an activity for the public interest, taking place among complexity and uncertainty, whose results (what works) are never totally predictable in advance through pre-determined standards and indicators. To suit these conditions, the evaluation should be situationally responsive; this would imply an attitude of commissioners to be open to interact with an evolving context and situation, allowing the evaluation team to choose, and possibly modify, approaches and methods of inquiry in order to better grasp what has been found to work better for what groups of population, in what circumstances, and why, be it expected or unexpected, and what can be potentially improved.

However, this is not the prevailing situation, which often tends to be molded on the market relationship of demand and supply, client-service: rigid, usually repetitive, ToRs define objectives, methodology and deliverables that fit simple, linear projects, not complex policies that unfold through time; tenders privilege contractors that assure the lowest price, and better financial warranties; time and budget limitations condition the search for sustainable changes. It is not uncommon the situation where “dialogue and discussions are not permitted. Everything has been decided ahead of time by the contracting agency”, such as “a 86-page EC

contract where the contractor must agree to in its entirety, and changing even a comma is grounds for disqualification” mentioned to us by Burt Perrin.⁹ In another case mentioned by Tessie Catsambas¹⁰ we have been informed of “evaluations commissioned by the US Federal Government that require you to run a national survey with no changes and have it up and running within two weeks of being awarded the contract”. As she commented: “That is not really evaluation, is it? It is just running a survey and handing it to the client”.

All this reflects a situation in which the commissioners act as pure mirrors of policymakers, feeling no autonomous responsibility. Being part of the policy organization, and at the same time managing an evaluation, commissioners may live the contradiction – that Wildavsky (1993: 220) outlined for the self-evaluating organization – between, on the one side, “administration” that implies stability and generates commitment and, on the other side, “evaluation” that suggests change and breeds skepticism. However, one should consider that while policymakers tend to act and think at the level of the project, according to a pre-established theory of change, to be translated into “goal > metric > result”, commissioners should act at the level of how the evaluated program or policy is actually implemented. And implementation is a complex and political phenomenon during which “much can go wrong” (Hill & Hupe, 2009: 7), or in a more positive vein, which “may often mean in fact a long voyage of discovery in the most varied domains, from technology to politics” (Hirschman, 1967: 32). And since most current policies are about “transformation” (addressing climate change, social inequalities, rights denials, etc.) commissioners would hopefully see beyond the linear logic of the program and feel responsible for exploring, promoting and allowing the actual changes that are in sight.

This implies a more open attitude toward the influence/“conditionality” of context. As Catsambas (Chapter 6: Catsambas, 2027) writes: “The same factors that may be perceived as weaknesses or challenges may also be used to advantage for the government evaluation commissioner who wishes to promote democracy through evaluation. This paradox really underscores the individual power and responsibility of an evaluation commissioner to map the context and think strategically about how to promote democracy from wherever they sit”.

Our perspective and aspiration for democratic commissioning imply two main imperatives for the commissioning processes of evaluation of current programs and policies:

- 1 a **collaboration** between commissioners, evaluators and other stakeholders, so that democratic and ethical principles of inclusion, empowerment, effectiveness, and lastly use, may prevail.
- 2 a determination that the rules of the market regulating the contract (time, budget, risk aversion, etc.)¹¹ do not prevail in the relationship¹²; finding out ways of signing **flexible** contracts allowing to adjust to the uncertainties and complexities of the implementation, and to look at the findings of the evaluation without blinders.

How “commissioning matters” can match the “democratic” inspiration of evaluation

Here, we want to refer to two streams of evaluation literature about the democratic inspiration of evaluation. First is the tradition inspired by Barry MacDonald’s (1977) typology, which distinguishes democratic evaluation from bureaucratic and autocratic evaluation. This tradition originated in the field of educational evaluation and has been enriched by the work of authors such as Lee Cronbach, Ernest House, Jennifer Greene and Thomas Schwandt, who emphasize values such as deliberation, participation and inclusiveness. Second is the tradition of government evaluation pioneered by Eleonor Chelimsky’s experience at the PEMD (Program Evaluation Methodology Division) of the US-GAO (General Accounting Office). This tradition emphasizes the value of fairness in the public interest (Chelimsky 2006).¹³

Lois-Ellin Datta (2011) labeled these two ways of conceiving democracy by evaluators as “populist”¹⁴ and “public interest” and has argued for the need of an encounter between them. This is what has actually happened in the *Handbook of Evaluation*,¹⁵ and in influential works such as Julnes (2012), Schwandt (2017), Schwandt & Gates (2021). And we can say that both streams are utilized by our authors. For example, Chapter 6 (Catsambas, 2027), in the first tradition, focuses on “safe and open dialogue, engagement of marginalized voices, considerations of environmental sustainability, increased capacity for uncensored scientific discourse and free press”. Chapter 7 (Lahey & Lucks, 2027), in the second tradition, emphasizes that the Sustainable Development Goals “are about democracy – among other things – as they relate to freedom of speech, tolerance for diversity and differing opinions, equity and justice. This ensures that, where balance is required, the necessary protections are in place to guarantee the safety of the most vulnerable in the decision-making process”.

In what ways does this literature on democratic evaluation impact on commissioning? How can we conceptualize commissioning as based on democratic principles, regardless of how the term is defined? We will analyze the intersection of commissioning and democratic principles under three headings.

The first heading (“agency”) deals with the responsibility of the commissioners, their way to act in a proactive mode and exert their “agency”. It is a matter of becoming aware of the leverage they may have on the conduct of the evaluation, and of learning how to get the necessary competences (Chapter 12: Bustelo, Crespo & Ligerio, 2027). In other words, how to be a “good” evaluator (and, thus, a “good” commissioner).

The second heading (“approaches”) admits that commissioning is indeed tacitly treated in different ways according to the different approaches to evaluation. This refers to the relationship between commissioner, evaluator and other stakeholders. In the so-called “evidence-based” approaches commissioning may be seen as simply formulating a demand (does the program work?) to which evaluators supply an answer: this is the typical client-service relationship. In participatory, realistic or complexity-compatible approaches commissioning is seen as a process in which

different actors collaborate from their respective roles to inquire into the ways change may happen (Gates, 2017, Cox & Barbrook-Johnson, 2021). This refers to the need for flexibility in evaluation contracts, ToRs admitting interactive theories of change, etc.

The third heading (“evaluation systems”) refers to the context within which evaluation takes place.¹⁶ We are well aware that evaluation systems may mainly use evaluation for institutional self-reinforcement, as we have been alerted since Power’s (1997) “audit society”,¹⁷ Dahlen-Larsen’s “evaluation society” (Dahler-Larsen, 2011) and Raimondo & Leeuw’s (2020) “evaluation capture”. True, commissioners face many pressures – timing, cost, a limited market of providers, and administrative obligations to deliver evaluations. Our focus is on how evaluation systems could be more open to receive unforeseen inputs from society, and commissioners might turn conditioning rules into opportunities for promoting democratic processes (Chapter 6: Catsambas, 2027). Indeed, like all things, evaluation systems can be used for different aims: supporting democracy or supporting command and control.

First: the responsible role of commissioners: exerting their “agency”

As early as in 1991, Perrin & Boschen wrote a chapter entitled “Getting the best from consultants” in the book *Evaluation Methods Sourcebook*, edited by Arnold Love and the Canadian Evaluation Society, which guided on what they defined as the four basic stages in a commissioning process: pre-selection and defining what the commissioner wants; selection of the evaluator; managing the process while the evaluation is under way; and follow-up action and utilization. As Perrin-Stame note in their chapter about that 1991 publication,

the main message (...) was the stress on the main responsibility of the commissioner in getting a useful evaluation. Rather than dismissing the role of commissioning, the authors enhanced it, and invited commissioners to feel responsible for clarifying what they want (what are their needs), and for working with the evaluators, in a collaborative relationship, to produce something that may be used. In other words, to use their power and authority in a proper way.

(Perrin & Stame, 2027)

However, except for the recent book by Catsambas & Davidson (2024) on *Evaluation Management* that also acknowledges the role of responsible commissioning, there has been little attention to the commissioning process by the evaluation community, specifically overlooking the role played by commissioners.

Not only do we find a gap in the evaluation literature; also, a major part of the evaluation guidelines (standards, guidelines, norms or principles developed by the professional associations and other organizations),¹⁸ and consequent frameworks of evaluation competencies¹⁹ have been developed from the evaluator’s perspective. Among the rare treatment of commissioners in evaluation guidelines we can mention the Australasian Evaluation Association guidelines, the IDRC Manager’s

8 *The Power of Commissioning*

Guide to evaluation, the United Nations Evaluation Group competencies, and the UKES Guidelines for Good Practice in Evaluation, developed in 2003. The latter case is significant. These are guidelines for evaluators, for commissioners, for institutional self-evaluation, and for evaluation participants. But curiously enough, this attention to other stakeholders vanished in the “UKES Framework of Evaluation Capabilities” this same evaluation society developed in 2012, in which, as in other frameworks developed by other societies and organizations, what is presented are evaluators’ competencies or capabilities. Especially, the commissioner’s role is completely overlooked.

As we said above, the commissioning process involves various stakeholders, each one playing crucial roles singularly and in collaboration with each other. As Chapter 12 (Bustelo, Crespo & Ligeró, 2027) mentions,

although the complex and diverse contexts in which evaluations are conducted are well-recognized and addressed by evaluation theory and practitioners, much of the literature and guidance focus on the role of evaluators, who are often seen as the central figures ensuring the success of the evaluation through their expertise and interaction with commissioners and policymakers. However, as evaluators having research and practical experience, we know that no matter how well an evaluation is designed and performed, that if the commissioning process goes wrong or gets difficult, the evaluation will not manage to be fully successful.

In the same line, one of our INTEVAL colleagues interviewed in their capacity of having executed a double role of evaluators and commissioners, Kim Forss, told us:

when I started out in the evaluation business, I was lucky to encounter a few good commissioners. That was very important for the development of my career. About the same time, I also met some very bad commissioners, and I was perceptive enough to see the difference. I understood that if I wanted to have a successful long-term career, it would be very important to stay clear of incompetent commissioners and to recognize and get to work for good commissioners. I would say that is a key competence for anyone who wants to build a career in evaluation (but – to my knowledge – that is not part of any competencies framework).

In her turn, Catsambas added:

In my experience, I have worked with some excellent commissioners who think strategically about the evaluation function in their organization, and who see the strategic implications of every evaluation. As I think of them, they are tough and demanding customers, sometimes making unreasonable demands, but their intention to make evaluation an instrument of learning and change is clear, and their intelligence at navigating the politics to get there is impressive.

Other definitions of “good” commissioner are scattered all over the book.²⁰ It implies acting responsibly, acknowledging their role, authority and power, taking responsibility for enhancing evaluation use (Chapter 2: Perrin & Stame, 2027); having a strategic thinking and overview of the whole process (Chapter 12: Bustelo, Crespo & Ligerio, 2027; Chapter 3: Rogers & Montague, 2027), developing a more strategic role that goes beyond the purely technical (Chapter 14: Olejniczak, Strzęboszewski & Stępień, 2027), using and encouraging critical evaluative thinking (Chapter 13: Bastøe, Høydal & Fossetøl, 2027), and balancing accountability with learning purposes (Chapter 11: Lindkvist & Lloyd, 2027); “focusing on getting the best, in all respects”, through a “collaborative rather than an adversarial approach, and dealing with bureaucratic rules and requirements” (Chapter 2: Perrin & Stame, 2027); developing a partnership and a good collaboration with evaluators and other stakeholders, not only policymakers’ mandates, building trust among those stakeholders, including a responsibility for listening and including voices not usually heard (Chapter 4: Ling & Milliard, 2027); mastering flexibility and adaptation during the process (Chapter 13: Bastøe, Høydal & Fossetøl, 2027), that is, “taking a bespoke (made-in-measure) approach to evaluation” and “using standards, criteria and guidelines intelligently rather than mechanically” (Chapter 2: Perrin & Stame, 2027); working through transparency (Chapter 3: Rogers & Montague, 2027), in open dialogue with evaluators throughout the process (Chapter 9: Marra & Buscemi, 2027), establishing clear communication protocols that foster transparency and democratic participation throughout the evaluation process (Chapter 5: Koleros, Liston & McGhee, 2027), and investing in strategic dissemination (Chapter 11: Lindkvist & Lloyd, 2027); Overall, a “good” commissioner is the one who contributes to enhance, follow and implement democratic principles in evaluation, in a process the steps for which have been designed in Chapter 4 (Ling & Millard, 2027)

Some experienced colleagues might be somehow skeptical about the need to educate commissioners in evaluation. After all, as Forss continues, “those early ‘good’ commissioners (...) did not have any experience as evaluators, but they still did a very good job in commissioning”. However, one should recognize its potential: “Having a background as an evaluator is neither necessary nor sufficient to become a good commissioner, but it can be very useful”. That is, not only evaluators, but commissioners should also receive an education in evaluation, and one that stresses the need for their collaboration based on democratic principles. The third section of this book deals with the possibilities of specific evaluation education of commissioners.

Bustelo, Crespo & Ligerio (2027), based on their experience in the Master on Evaluation at Complutense University of Madrid, where both professional evaluators and commissioners have been educated since 2002, identify a comprehensive framework of competencies, which considers not only evaluators, but also commissioners and users of evaluations. They stem from the idea that “understanding the competencies needed by these groups is crucial for enhancing the commissioning process and, consequently, the overall effectiveness of evaluations” and discuss on a specifically active and responsible role for commissioners which can better contribute to a commissioning process that enhances democratic values.

Second: different approaches

Democratic principles may be present in various instances of commissioning: a matter of research methods, values and relationships.

First, **methods**. Confronted with evaluation questions mostly concerned with causality (does the program work? Can it be scaled up?), evaluation designs have spanned between contribution vs. attribution analysis. As all the discussion about contribution analysis and the Theory of change (Koleros et al., 2024) has shown, approaches that are limited to demonstrate whether the planned causal pathway is valid may miss other occurrences: un-expected and positive, un-expected and negative, for different groups. And, in case of recognized failure of attaining the expected goal, the fear has been expressed that a program may be phased out even if it might have had otherwise favorable results for groups not previously considered. All this past wisdom has taken a more urgent traction with the discourse about complexity, emergent traits of eco-systems and reverse causality, that are hardly considered in causal models, but are concretely experienced in the course of an evaluation. In Chapter 7, Lahey & Lucks (2027) offer a pertinent example when they note “the complexity created by the various tradeoffs across SDGs” that needs to be unpacked.

In the tight relationship outlined above, a positive breakthrough could happen if commissioning would openly acknowledge as its goal the benefits to the concerned people and situations, with a particular attention to the more marginalized among them, not the mere accomplishment of the program as planned. In Chapter 4, Ling & Millard (2027) write: “commissioners ensure that evaluation contributes to the public good, provides opportunities for enhanced public debate, and is clear about its reach and validity”. This would imply that ToRs invite to revisit the Theory of Change in interaction with the other actors (evaluator, stakeholders, operatives); allow for a choice of methods suited to the situation (and not following a hierarchy of methods, or imposing metrics and standards, that disregard the chances offered to different groups of population); and focus on change as it happens.

This point is raised in Chapter 3 by Rogers & Montague (2027) who, surprisingly, pick up a controversial view about democracy in evaluation. “Despite the claim that ‘carrying out randomized trials can help strengthen our democracy’ (Leigh, 2023) through increased transparency, [these authors] think it can actually undermine it”. And this is because the focus on the average effect, such as with Randomized Control Trials (RCTs), “ignore the differential effects that all interventions produce. Implementing what works on average often does not work well for the most marginalized individuals and communities”. What is needed is a careful attention to heterogeneity.

These considerations underscore the request for bespoke ToRs, avoiding the boilerplate effect (Chapter 6: Catsambas, 2027; Chapter 2: Perrin & Stame, 2027), and flexible contracts.

Second, and related to the above, a democratic approach is concerned with **values of inclusion**, that various authors identify as implying safe and open dialogue, representation, equity, sustainability, engagement of marginalized voices. This may refer to letting evaluated parties and beneficiaries (right holders) participate

in judgments and not only in providing information, and to considering the values of different actors and groups affected by the evaluation. In Chapter 2, Perrin & Stame (2027) refer to the wide set of actors that must be engaged (intended beneficiaries, community, civil society). All these aspects should be considered at the time of commissioning. For instance, in Chapter 3, Rogers & Montague (2027) stress the need that the commissioning process should allow time and resources for individuals and communities to be engaged in deliberative processes.

In Chapter 4, Ling & Millard (2027) provide a rubric through which it is possible to assess to what extent, at any stage of the evaluation, through inclusivity and transparency, right holders and beneficiaries have been considered in commissioning, i.e.: accessibility, engagement in data interpretation, access to results, operationalization of results.

Some authors outline the opportunity offered to democratic commissioning by the Sustainable Development Goals. In Chapter 7, Lahey & Lucks (2027) consider the SDGs as “an immense effort in global democracy” to identify diverse needs across multiple peoples and conflicting priorities. In a similar vein, in Chapter 6, Catsambas (2027) suggests that “Evaluation commissioners need to remain alert to current changes that offer new opportunities for evaluation and democracy. For example, in 2015, countries signed a UN resolution to promote evaluation of the SDGs at country level, which offered an in-the-moment opportunity for evaluation commissioners to push for evaluation of progress toward the Global Goals”.

Third, **relationship** between commissioner and evaluator. If commissioning is seen as a mere client-service relationship, the process may be affected by distrust. This may be documented in the kind of contract (full of cautionary clauses) and may be revealed by the behavior of the partners, each acting according to his/her own logic (the administrator’s and the researcher’s). However, as Lahey & Lucks (2027) suggest in Chapter 7, “building a relationship between evaluators and ‘commissioners’ of evaluation is critical in helping ensure that evaluation actually plays a meaningful role in support of the SDGs”-

Trust between commissioners and evaluators is considered the key for a successful evaluation.²¹ As Kim Forss says:

It is necessary that commissioners actually trust the teams they have appointed to do a good job. It is equally important that evaluation teams trust commissioners and feel that they can be open about problems that come, and that commissioners are able to help them solve it, or at least fundamentally understand the issues.

Moreover, trust has to be “established and maintained”. Hur Hassnain, in his turn, considers how to build trust the main lesson he learned in his experience as commissioner.

In Chapter 11, Lindkvist & Lloyd (2027) testify on a decade-long collaboration between the Norad Evaluation Department and external evaluators from ITAD. Through the lens of four major evaluations commissioned by NORAD’s Evaluation Department between 2014 and 2024, the authors (one a commissioner, the other an evaluator) highlight how commissioners exercised agency, fostered organizational

learning, and built trustful partnerships with evaluation teams, at each stage of the evaluation process – scoping, design, implementation and dissemination.

In Chapter 5, Koleros, Liston & McGhee (2027) present a specific experience where the principle of collaboration is foremost: the platform for equitable evaluation, which has been subscribed by various philanthropic societies. These authors review how critical domains such as time, cost, giving and receiving feedback, self-reflectivity, equity at every step can be conceived differently in the relationship between commissioners and evaluators: and document how “over time, the commitment to consistent virtual meetings and detailed documentation helped build trust, fostering more open and honest conversations about the evaluation’s goals and challenges”.

In Chapter 3, Rogers & Montague (2027) in their turn elaborate on how the principle of transparency, a hallmark of all democratic regimes, may affect the commissioner-evaluator relationship. “Transparency refers to openness of information about intentions, plans, rules, processes, actions and results. Transparency in evaluation includes both the products of an evaluation and the processes, including contracting and commissioning processes”. It refers to what information should be shared, methods, findings, protection of anonymity.

Third: evaluation systems

One of the ways in which the bureaucratic capture of evaluation, or evaluation drift, takes place is the subordination of commissioning to the interests of bureaucratic, administrative or inflexible preset policymaking: limits to what can be asked, short-term results and budgetary constraints, selective use of results fitting the idea of success of the decision-maker (Raimondo & Leeuw, 2020: 145–7). Inversely – as Catsambas (2027) suggests in Chapter 6 – evaluation commissioners may influence the use of evaluation to contribute to democracy if they critically consider the institutional (government and organizational) context where that evaluation takes place and develop strategies to be effective within that specific context. To grasp this possibility, attention should be given to “how evaluation commissioners perceive their assigned role and how they choose to step into that role in relation to promoting democratic processes”. This implies understanding “their risk tolerance, especially in cases where strengthening democracy involves deviating from assigned roles and job expectations, and possibly putting at risk their career and livelihood”.

On the other hand, one should also consider that most (Western) governments are now organized (or, at least, pretend to be organized) according to principles of the New Weberian State, which – as Pollard (2027) teaches us in the Appendix – incentivize “understanding and meeting citizens’ needs to support democratic accountability and responsiveness”, and may even set aside a traditional top down attitude and “focus on the value and role of citizen engagement and ‘bottom up’ influence from citizens and civil society”. If this inspiration prevails, commissioners’ role becomes critical: they can be pivotal in “supporting public administrations to understand the value, relevance and impact of results for different groups of

citizens as well as exploring design, process and context”: from conceiving the demands for the evaluation of single policies (“ask critical evaluative questions and contextualize analysis of results rather than focus on traditional results-based questions”), to building evaluation capacities in beneficiaries (as in the example of citizens panels).

An interesting case of putting the evaluative questions at the core of the evaluation thrust is offered in Chapter 8 by Olejniczak, Kupiec & Park’s (2027) account of the Learning Agendas (LA), promoted by the Office of Management and Budget (OMB) in the United States and also utilized in Poland. An LA is a tool in the hands of commissioners. It addresses directly our concern for the “repetitive exercise”. It brings together two important trends: overcoming the “tendency to focus on operational, short-term, and technical issues that occupy the current agenda for program implementation” and asking “the right questions”, based on previous knowledge, a creative way of utilizing the concept of streams of evidence (Rist & Stame, 2006). A LA is the product of an iterative process. It affects (democratic) commissioning “at two decision points and negotiation processes: (a) LA formulation” – knowledge needs and evaluation questions are identified and discussed between the evaluation unit and a plurality of actors whose voices are heard and listened to; “(b) [the content of] the evaluation study commissioned” (the chapter lists examples of LAs based on questions about citizens’ rights and other democratic goals).

When speaking of commissioning within evaluation systems, we are interested in the various aspects in which the evaluation function is organized, and their ability to favor/ hinder democratic commissioning. We may think at differences in political orientation (as presented by Pollard); differences in the institutional context, be it the administrative level (Chapter 7: Lahey & Lucks, 2027) or the nature of the institution, as between government ministries and agencies and research institutions (Chapter 13: Bastøe, Høydal & Fossetøl, 2027); or differences in the scope of the evaluations conducted (Chapter 10: Laaksonen & Palenberg, 2027; Chapter 9: Marra & Buscemi, 2027).

It is a matter of challenges (being dependent on rigid decision-making) and opportunities (exploiting the latitude of conceiving the evaluation questions and interacting with the evaluation implementation). To this purpose, in Chapter 6, Catsambas (2027) draws a table where she crosses context levels (government, United Nations, foundations²²) with aspects (organizational values, organizational priorities, placement of evaluation) that could be used by the evaluation commissioner to justify, leverage and encourage the democratic process. But such mapping – she cautions us – “is not intended to identify barriers and give up, but to assess what is feasible, determine the effort change will take, and develop the best path to achieve it”.

Taking the example of the United Nations, Catsambas notices that an obstacle to evaluation to serve democracy is the weight of bureaucracy and internal political complexities (and consequential self-protective defensiveness on the part of the evaluation unit). In contrast, she found that “there are UN/IFI evaluation offices that have furthered evaluative thinking and action by embedding methods that broadened the scope of evaluation to new underrepresented actors, invited

open dialogue, and expanded capacity in their organizations to commission and use evaluation at country level”, as in the case of the leadership of UNICEF that invested along with the Government of Finland in a global network of evaluation organizations to promote the voice of civil society in country-level evaluation.

While dealing with a particular national evaluation system, the Italian evaluation of EU structural funds of cohesion, Chapter 9 (Marra & Buscemi, 2027) illustrates the undemocratic shortcomings of keeping commissioning to a mere logic of contracting. The financial requirements for large procurement contracts with numerous deliverables have favored a market concentration on a few big centers with strong economic capacity and good political connections, which acted as a barrier to the entry of smaller, diverse and geographically representative consultancies. This produced standard evaluations, showing a scarce “ability to address core research questions, collect primary data, and analyze diverse information sources” thus raising serious concerns regarding the quality of evaluations”. On the other hand, this has not prevented that good organizational arrangements in the management of evaluations, with steering groups addressing thematical, methodological and use issues, have in some cases shown the potential for “commissioners’ capacity and willingness to foster a constructive relationship with evaluators”.

A different solution to issues of large evaluations is provided in Chapter 10 by Laaksonen & Palenberg’s (2027) case of the Ministry of Foreign Affairs (MFA) of Finland. Here too there is a concern for large central evaluations due to increasing staff shortages and lack of internal needed competence. The MFA has created a system by which the internal Development Evaluation Unit (DEU) commissions the evaluations to the Evaluation Services Management, which is composed of two consortia, who alternatively perform the evaluation. While taking the main decisions, and following its evaluation plan with continuing overall decision-making authority, the DEU is supported by other figures, issued by its interior, by the ESM and intermediary. The rationale for this elaborated system is increasing the quality of the evaluation, and supposedly also the flexibility in planning and commissioning; moreover, it intends to favor the creation of new capacities in “emerging evaluators” working inside the consortiums. At the same time, it has created a kind of duopoly in the market for evaluation.

How this book is organized

The chapters that enter into this volume alternate theoretical thinking with reflections on practice. All of them contribute to highlight some issues of our three headings.

The volume is divided into three parts, each of which is focused on a main question about the nature of commissioning, and its practical relevance:

- 1 *Commissioning (with democratic principles) matters*, about the rationale for this volume
- 2 *Experiences of commissioning in different institutional contexts*, which deals with the strategies for democratic commissioning that can be envisaged in different evaluation systems

- 3 *Educating the commissioner, through a dialogue between commissioners and evaluators*, where the specific competences for commissioners who understand the responsibility of their role are spelled out.

Part I. Commissioning (with democratic principles) matters

Chapter 2 (Perrin & Stame, 2027) lays the **foundations of the often-underestimated role of evaluation commissioners in complex, fast-changing contexts**. It explains how commissioners shape relevance, quality and democratic value by defining scope, budgets and teams, guiding implementation, assuring quality and enabling meaningful use of findings that respond to citizens' needs throughout all evaluation stages globally.

Chapter 3 (Rogers & Montague, 2027) examines **how evaluation commissioning can strengthen or undermine democracy**. Using typologies of democracy and evaluation, it highlights **transparency, meaningful stakeholder engagement and attention to heterogeneity**. The authors warn against average-effect designs and advocate learning-oriented questions, inclusive participation, and reporting differential impacts on marginalized groups.

Chapter 4 (Ling & Millard, 2027) argues, drawing on **deliberative democracy**, that **evaluation commissioning can empower democratic spaces through inclusion, dialogue and accountability**. It emphasizes intentionality amid power relations and proposes guiding questions and a rubric to ensure inclusivity, transparency and accessibility. Meaningful engagement of rights holders reframes evaluation as participatory knowledge production.

Chapter 5 (Koleros, Liston & McGhee, 2027) explores **shifting equitable evaluation from external actions to internal, shared practices between commissioners and evaluators**. It traces equity across all evaluation stages and discusses challenges commissioners face, illustrated through a case study of an equitable evaluation of the Robert Wood Johnson Foundation's housing justice work in the United States.

Part II. Experiences of commissioning in different institutional contexts

Chapter 6 (Catsambas, 2027) argues that **evaluation commissioners possess greater power than often recognized to influence democratic outcomes**. Emphasizing political, institutional and organizational context, it presents a framework to map constraints and enablers across sectors. Real-world examples illustrate ethical and strategic commissioning in fragile or politically sensitive environments.

Chapter 7 (Lahey & Lucks, 2027) **links SDG evaluation to democratic governance, highlighting the SDGs as a participatory, country-led framework**. It explores how commissioners navigate political and technical complexity across levels and sectors, arguing that SDG evaluations can engage diverse actors, inform trade-offs and support coordinated, evidence-informed and democratic decision-making.

Chapter 8 (Olejniczak, Kupiec & Park, 2027) presents **Learning Agendas as a strategic commissioning approach to bridge gaps between evidence producers and users**. It shows how Learning Agendas support democratic policymaking through participation, institutional learning and integration of expertise and values, offering an eight-step framework based on US and EU empirical cases.

Chapter 9 (Marra & Buscemi, 2027) analyzes **how market-based evaluation commissioning** in Italy's Cohesion Policy **affects democratic outcomes**. It identifies risks related to procurement, market concentration and regional disparities, and advocates inclusive commissioning practices, theory-driven collaboration and management strategies that enhance evaluation quality, utilization, informed decision-making and democratic debate.

Chapter 10 (Laaksonen & Palenberg, 2027) examines Finland's Ministry for Foreign Affairs' Evaluation Management Services model, assessing **how partial outsourcing reshapes democratic commissioning**. It discusses professionalization, knowledge management and participatory decision-making benefits, alongside trade-offs such as intermediary structures and altered commissioner roles, drawing on firsthand implementation experience.

Chapter 11 (Lindkvist & Lloyd, 2027) illustrates **commissioning as a relational, strategic practice beyond contract management**, drawing on a commissioner-evaluator collaboration in Norway. It shows how commissioners exercise agency, build trust and foster learning through all evaluation stages, offering practical lessons **to strengthen quality, use and democratic impact of evaluations**.

Part III. Educating the commissioner, through a dialogue between commissioners and evaluators

Chapter 12 (Bustelo, Crespo & Ligerio, 2027) highlights **commissioners as central yet often overlooked actors in democratic evaluation**. It frames commissioning as an exercise of agency and **identifies key competencies** – knowledge, management, interpersonal, ethical and transformative – needed to shape priorities, engage stakeholders, ensure quality and promote use, positioning commissioners as active agents of democratic change.

Chapter 13 (Bastøe, Høydal & Fossetøl, 2027) presents a Norwegian **training initiative** designed to **strengthen evaluation commissioning and use through critical evaluative thinking**. It emphasizes dialogical skills, reflection and evidence-based practice, highlighting the need for strategic anchoring and systematic planning to improve evaluation use and promote democratic values in public administration.

Chapter 14 (Olejniczak, Strzęboszewski & Stepień, 2027) analyzes Poland's Academy of Evaluation, a long-term **capacity-building initiative for commissioning authorities**. It traces its shift from technical training to **developing knowledge brokers**, showing how sustained academic programs improve commissioning quality, collaboration and networks, and offering lessons for democratic, adaptive evaluation capacity development.

The Appendix (Pollard, 2027) **applies the Neo-Weberian State model to evaluation commissioning**, arguing for participatory citizen engagement,

context-sensitive results frameworks, and professional leadership that supports delegation and agility. It emphasizes commissioners' responsibility to align evaluations with human rights, equality and democratic public value.

The concluding Chapter 15 (Bustelo & Stame, 2027) **reframes democratic commissioning through theory and practice**. It argues democratic commissioning depends on **process and agency**, not fixed rules, and provides **practical guidance**, including “dos and don’ts”, to help commissioners assess risks, seize opportunities and pursue the public good (Table 1.1).

Table 1.1 In the guise of an index of contents, to help readers navigate the book’s chapters

What commissioning is about	Chapter 2 – the power of commissioners Chapter 4 – the commissioning process Chapter 6 – the managerial tasks Chapter 8 – framing the evaluation demands Chapter 9 – contracting
Commissioning and democratic theory	Chapter 3 – democratic principles Chapter 4 – deliberative democracy Chapter 15 – democratic commissioning
Tools for democratic commissioners	Chapter 4 – a rubric of the commissioning process Chapter 12 – competencies for democratic commissioning Chapter 13 – training for critical evaluative thinking Chapter 14 – training as knowledge brokers Chapter 15 – dos and don’ts
Overcoming power unbalances	Chapter 3 – accounting for heterogeneity Chapter 5 – challenges of equitable evaluation
Contextual factors	Chapter 5 – philanthropies Chapter 6 – differences between international agencies / governments / nonprofit Chapters 8 & 10 – government Chapter 9 – government – EU Chapters 13 & 14 – public administrations Appendix – theory of government
Topical relevance	Chapter 7 – SDGs Chapter 9 – EU cohesion policy Chapters 8, 10 & 11 – international aid
Relationship between commissioners and evaluators	Chapter 5 – shared practice for equitable evaluation Chapter 10 – the Finnish experience Chapter 11 – positive collaboration in Norway Chapter 12 – both roles sharing and exchanging experiences in master’s program
Geographical area	Chapter 3 – Australia, Canada Chapters 4, 6 & 7 – International Chapter 5 – United States; Chapter 8 – United States & EU cases Chapter 9 – Italy & EU Chapter 10 – Finland Chapters 11 & 13 – Norway Chapter 12 – Spain Chapter 14 – Poland

A starting platform, not a definitive closure of the topic

Although we are proud to have entered an uncharted territory, we do not pretend to have said the last word about it. Rather, on the contrary, the more we inquire into the democratic dimension of commissioning, the more aware we become of the limits of our own treatment of the topic that need to be overcome.

First, we must deepen our understanding of the tacit reproduction of power asymmetries. Beyond acknowledging the power of commissioning, we must further interrogate how commissioners reproduce or contest structural inequities based on race, gender, or colonial legacies. We must also consider how they can link their practice to trends in decolonizing evaluation, Indigenous perspectives and intersectional analysis.

Second, in our attempt at a comprehensive treatment of commissioning within formal institutional systems, we primarily drew our empirical and conceptual bases from European, North American and Australian settings. We acknowledge that important bodies of practice and scholarship – particularly from community-based, Indigenous and Southern contexts – are underrepresented here. Therefore, we invite readers from other regions to adapt the framework to situations embedded in different administrative traditions, political cultures and community accountability logics beyond the scope of this volume.

We hope this book will be seen as a significant step in an ongoing conversation, not the final word on commissioning.

Notes

- 1 We thank Tessie Tzavaras Catsambas, Tom Ling, Burt Perrin and Patricia Rogers for savvy comments to the first version of this chapter.
- 2 Karlsson and Conner (2006: 234) speaking of the democratization of evaluation say that it “occurs in the central components in the evaluation process (...) deciding on the aims of the evaluation (...) determining the resources for the evaluation (...) selecting the evaluation questions and methods”.
- 3 See box. Thanks to Emily Gates for reminding us about Cronbach’s relevance.
- 4 Datta’s article never mentions commissioning. Moreover, that literature is premised on a negative bias against politics, therefore dismissing any possibility of independence among commissioners.
- 5 Leeuw & Pleger (2023) in their article on evaluation capture and ways of resistance are interested in the evaluator’s qualities of resilience, but do not address our concern with commissioners themselves. However, Leeuw, in his testimony as a commissioner with evaluation experience, considers behaviors of commissioners that he once held good (like detailed ToR and other ways to constrain evaluators) as expressions of evaluation capture.
- 6 See correspondence with Rob van den Berg and interview with Jos Vaessen.
- 7 Thank you very much to the selected group of Kim Forss, Hur Hassnain, Frans Leeuw, Jos Vaessen for thoughtfully answering our questionnaire about their experiences as commissioners.
- 8 See Marra & Buscemi’s Chapter 9 for understanding the flaws of limiting commissioning with contracting.

- 9 Comment by Burt Perrin to the first version of this chapter.
- 10 Comment by Tessie Tzavaras Catsambas to the first version of this chapter.
- 11 See Cox & Barbrook-Johnson (2021) for a detailed analysis of the legal rules of contracting, that are barriers to utilizing complexity-appropriate evaluation approaches.
- 12 Of course, this does not mean ignoring the risk of fraud or misuse of the public money spent for the evaluation. But that risk would be better averted if the money were spent in a purposeful evaluation, while the rigid application of an unfitted method could be a greater waste of money, not to say of a better way to hide misappropriation.
- 13 This tradition is reflected in some Evaluation guidelines. The Australasian Evaluation Society (AES) has in its Code of Ethics the principle of “Public interest” which is explained as “Members should consider the interests of the full range of stakeholders in their evaluation work, including the broader public interest, and in particular, the potential impacts of differences and inequalities in society”. The American Evaluation Association (AEA) Guiding principle E is “Common Good and Equity: Evaluators strive to contribute to the common good and advancement of an equitable and just society”.
- 14 “Populist” is a US label that would be misunderstood in a European, and perhaps world-wide, context, where it means a method of constructing a political consensus based on demagoguery and supporting the worst feelings of the masses, by a charismatic leader. In our case, we would use the word “inclusive”.
- 15 In a chapter of the *Handbook* Karlsson and Conner (2006: 229) define the public interest as opposed to specific power holder interests.
- 16 We are also aware that not in all countries the evaluation function is institutionalized enough, so the evaluations and, therefore, the commissioning process for those evaluations, might be anecdotal or not occur in sufficiently organized evaluation systems. In these cases, though, the pressures that commissioners suffer might be even worse.
- 17 Power (1997) suggests that the “rituals of verification” can take a dysfunctional quality by imposing values and have unintended negative consequences.
- 18 Apart from the codes developed by the professional associations of evaluators, see the Joint Committee on Standards of Educational Evaluation, the OECD or the UNEG standards.
- 19 See the EES, UKES, CES, or others such as IDEAS, DFID and UNEG.
- 20 In Chapter 15, we propose a re-elaboration of practical guidance based on the lessons learned by the authors in their studies.
- 21 Wildavsky (1993: 234) goes as far as declaring “evaluation as trust”.
- 22 The relationship between philanthropies and democracy has been broadly analyzed by Reich & al. in *Philanthropy in Democratic Societies* (2016).

References

- Bastøe, Per Øyvind, Øyunn Syrstad Høydal and Knut Fossetøl. 2027. “Training for Critical Evaluative Thinking.” In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Brousselle, Astrid. 2022. “Planting Seeds for Change.” *Evaluation* 28 (1): 9–10.
- Bustelo, Maria, Ramón Crespo and Juan Andrés Ligeró. 2027. “The Good Commissioner: Competencies for Enhancing Democracy in Commissioning Processes.” In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.

- Bustelo, Maria, and Nicoletta Stame. 2027. "Democratic Commissioning: Lessons from Theory and Practice." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Catsambas, Tessie Tzavaras. 2027. "The Influence of Context for Different Types of Commissioners of Evaluation: Challenges and Opportunities." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Catsambas, Tessie Tzavaras, and Jane Davidson. 2024. *Evaluation Management: How to Commission and Conduct Evaluations That Matter*. Thousand Oaks, CA: Sage.
- Chelimsky, Eleanor. 2006. "The Purposes of Evaluation in a Democratic Society." In *The SAGE Handbook of Evaluation*, edited by Jan F. Shaw, Jennifer C. Greene, and Melvin M. Mark, 33–55. London: Sage.
- Cox, Jayne, and Pete Barbrook-Johnson. 2021. "How Does the Commissioning Process Hinder the Uptake of Complexity-Appropriate Evaluation?." *Evaluation* 27 (1): 32–56.
- Cronbach, Lee. 1980. "Our Ninety-Five Theses." In *Toward Reform of Program Evaluation*, edited by Lee J. Cronbach, Sueann R. Ambron, Sanford M. Dornbusch, Robert D. Hess, Robert C. Hornik, D. C. Phillips, Decker F. Walker and Stephen S. Weiner, 1–11. San Francisco: Jossey-Bass.
- Dahler-Larsen, Peter. 2011. *The Evaluation Society*. Stanford, CA: Stanford University Press.
- Datta, Lois-Ellin. 2011. "Politics and Evaluation: More than Methodology." *American Journal of Evaluation* 32 (2): 273–294.
- Gates, Emily. 2017. "Learning from Seasoned Evaluators: Implications of Systems Approaches for Evaluation Practice." *Evaluation* 23 (2): 152–171.
- Greene, Jennifer. 2006. "Evaluation, Democracy, and Social Change in a Democratic Society." In *The SAGE Handbook of Evaluation*, edited by Jan F. Shaw, Jennifer C. Greene, and Melvin M. Mark, 118–140. London: Sage.
- Hill, Michael, and Peter Hupe. 2009. *Implementing Public Policy*. 2nd ed. London: Sage.
- Hirschman, Albert O. 1967. *Development Projects Observed*. Washington, DC: Brookings Institution.
- Julnes, George, ed. 2012. "Promoting Valuation in the Public Interest: Informing Policies for Judging Value in Evaluation." In *New Directions for Evaluation*, no. 133. San Francisco: Jossey-Bass.
- Karlsson, Ove, and Ross F. Conner. 2006. "The Relationship between Evaluation and Politics." In *The SAGE Handbook of Evaluation*, edited by Jan F. Shaw, Jennifer C. Greene, and Melvin M. Mark, 225–242. London: Sage.
- Koleros, Andrew, Marie-Hélène Adrien, and Tony Tyrrell, eds. 2024. *Theories of Change in Reality: Strengths, Limitations and Future Directions*. New York: Routledge.
- Koleros, Drew, Monique Liston and Ray McGhee. 2027. "How Evaluation Commissioners and Evaluators Can Collaboratively Move Equitable Evaluation from a Set of External Actions to Internal Practices." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Laaksonen, Sari, and Markus Palenberg. 2027. "Evaluation Management Services at the MFA of Finland." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Lahey, Robert, and Dorothy Lucks. 2027. "Democracy, SDGs, Evaluation and Commissioners—Connecting the Dots." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.

- Leeuw, Frans, and Lyn Pleger. 2023. "Evaluation Capture, Evaluator Resilience, and the Need for Competencies of Evaluators." *Journal of Multidisciplinary Evaluation* 19 (46): 45–53.
- Leigh, Andrew. 2023. "Evaluating Policy Impact: Working Out What Works." *Australian Economic Review* 56 (4): 431–441.
- Lindkvist, Ida, and Rob Lloyd. 2027. "The Art of Commissioning. Lessons from Commissioner-Evaluator Collaboration." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Ling, Tom, and Ananda Millard. 2027. "Commissioning with Clarity: Empowering Democratic Spaces by Drawing on the Principles of Deliberative Democracy." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- MacDonald, Barry. 1977. "A Political Classification of Evaluation Studies in Education." In *Beyond the Numbers Game: A Reader in Educational Evaluation*, edited by David Hamilton, David Jenkins, Christine King, Barry MacDonald, and Malcolm Parlett, 105–108. London: Macmillan.
- Marra, Mita, and Virgilio Buscemi. 2027. "Democratic Implications of Commissioning Market-Based Evaluations: Insights from Italy's Cohesion Policy." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Olejniczak, Karol, Tomasz Kupiec and Yaerin Park. 2027. "Learning Agendas – From Studies to Streams of Evidence." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Olejniczak, Karol, Piotr Strzęboszewski and Marlena Stępień, 2027. "Academy of Evaluation – A Capacity-Building Initiative for Commissioning Authorities." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Perrin, Burt, and Kathryn A. Boschen. 1991. "Getting the Best from Consultants." In *Evaluation Methods Sourcebook*, edited by Arnold J. Love, 176–195. Ottawa: Canadian Evaluation Society.
- Perrin, Burt, and Nicoletta Stame. 2027. "How to Make Commissioning Matter in Times of Complex Fast-Changing Reality." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Pollard, Alison. 2027. "Applying Insights from the Model of the Neo-Weberian State to Commission More Democratic and Valuable Evaluations." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Power, Michael. 1997. *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Raimondo, Estelle, and Frans L. Leeuw. 2020. "Evaluation Systems and Bureaucratic Capture." In *Changing Bureaucracies: Adapting to Uncertainty, and How Evaluation Can Help*, edited by Burt Perrin and Tony Tyrrell, 143–160. New York: Routledge.
- Reich, Rob, Chiara Cordelli, and Lucy Bernholz. 2016. *Philanthropy in Democratic Societies*. Chicago: Chicago University Press.
- Rist, Ray, and Nicoletta Stame. 2006. *From Studies to Streams*. New Brunswick, NJ: Transaction Publishers.
- Rogers, Patricia. 2022. "Changing the Substance and Image of Evaluations." *Evaluation* 28 (1): 24–26.

22 *The Power of Commissioning*

- Rogers, Patricia, and Steve Montague. 2027. "How Evaluation Commissioning Processes Can Support or Undermine Democratic Processes." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Schwandt, Thomas A. 2017. "Professionalization, Ethics, and Fidelity to an Evaluation Ethos." *American Journal of Evaluation* 38 (4): 546–553.
- Schwandt, Thomas A. 2023. "Competencies for Evaluation as a Civic Science." *Journal of MultiDisciplinary Evaluation* 19 (46): 23–30.
- Schwandt, Thomas A., and Emily F. Gates. 2021. *Evaluating and Valuing in Social Research*. New York: Guilford Press.
- Wildavsky, Aaron. 1993. *Speaking Truth to Power*. New Brunswick, NJ: Transaction Publishers.

Part I

Commissioning (with democratic principles) matters



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

2 **How to make commissioning matter in times of complex fast-changing reality**

Burt Perrin and Nicoletta Stame

Introduction

The current movement toward more democratic policies able to address the great upheaval of climate change, huge social differences and new conflicts, at the same time of great technological innovation and new movements for social justice, puts under pressure not only evaluation approaches, but also the whole fabric of evaluation management.

What is at stake is the way evaluations can take into account and address challenges and opportunities linked to the complex reality that is changing fast during the implementation of program, providing evidence and insights that can guide future directions.

This new task requires the contribution by all parties to the evaluation: those who perform the evaluation (with all the potential approaches, methodologies and ethical guidelines) together with those who ask for an evaluation, the commissioners, who have the power to launch an evaluation and may have the last word over its use. It is a matter of responsibility on both sides, as has been stressed in a few far-sighted interventions that have called for a change in attitude largely shared by evaluators and commissioners alike (Brousselle, 2022, Rogers, 2022, Valters, 2014).

Yet, while one can get inspiration from the abundant literature on evaluation approaches, that has recently been enriched by debates spanning between Theories of Change and big data, not to forget artificial intelligence, there is a stunning silence about the role of commissioning, and its potential role to be played in the current situation. This is perhaps the legacy of an old attitude of dismissal. Commissioning seems to be viewed as separate from evaluation, with the contract seen only as a starting condition and a source of blame if issues arise.

Commissioning of evaluation comes under the area of evaluation management, and in most textbooks, it often culminates in guidelines for the preparation of a good contract, from proposal, selection of evaluator, legal arrangements for privacy and independence, deliverables. A challenge is that evaluation services are not the same as the production of goods or infrastructure, where the technique of production can be established in advance and the contractor has only to abide with them. A good evaluation contract should consider what is inherent in evaluation,

DOI: 10.4324/9781042000968-3

This chapter has been made available under a CC-BY-NC-ND license.

the immateriality of the activity of providing information, the values involved, the changes in implementation that should be expected, etc., something that is particularly cogent in the current situation.

There are a few exceptions to the traditional textbook treatment that are a good guide for undertaking the change in perspective that we are promoting.

Perrin & Boschen (1991) in a publication entitled *Getting the best from consultants*, showcased what would be the “best” kind of relationship between two parties – commissioner and evaluator who are bound together by a contract. It was a guide on how to deal with all aspects that frame the task of commissioning, along four basic stages: 1. Pre-selection: defining what the commissioner wants; 2. Selection of the evaluator (various ways and criteria); 3. Managing the process while the evaluation is under way; 4. Follow-up action and utilization.

This contribution had been prompted by the experience of the authors who had observed failure by some commissioners to think through what they want out of the evaluation study and why, problematic selection processes, minimum involvement with the evaluator, Terms of Reference (ToRs) that intrude on methodological matters, etc.

The main message of that chapter, and what strikes us for its continuous relevance, was the stress on the main responsibility of the commissioner in getting a useful evaluation. Rather than dismissing the role of commissioning, the authors enhanced it, and invited commissioners to feel responsible for clarifying what they want (their needs), and for working with the evaluators, in a collaborative relationship, to produce something that may be used. In other words, to use their power and authority in a proper way.

Thirty years later, Catsambas & Davidson (2024) in a book entitled *Evaluation Management* devote an entire chapter to “Commissioning great evaluations” that, again acknowledges the role of responsible commissioning, moving between avoiding pitfalls and promoting good practices.

Keeping the lessons of these contributions in mind, we now turn to the challenge we raised at the beginning: changes in the context where policies are implemented (complexity, uncertainty) call for new kinds of programs, seen as part of larger strategies of “transformation” (SDGs), in the “public interest”. Consequently, evaluation questions need to move from attainment of desired results by a specific program, to openness to alternatives and opportunities for change that may materialize in the getting together of multiple initiatives. Not everything can be established in advance, and commissioners should envisage new kinds of collaboration with stakeholders and larger communities, and undertake iterative processes to review theories of change and logical frameworks in order to grasp changes in the context and in the larger situation: and take responsibility for it, also through flexible contracts that allow for reassessing the situation when needed.

In the following sections we will touch on a series of issues, summarized in Table 2.1, that refer to how commissioning fares with respect to these challenges. In the first place we will return to the suggestions for good commissioning in order to identify what problems still trouble their accomplishment (I. *Overall imperatives and guiding principles for action*). Next, we will review some updated ways in which serial ToRs, wrong questions, inappropriate use of guidelines, bureaucratic

Table 2.1 Steps for commissioners to take

I. <i>Overall imperatives and guiding principles for action</i>	Acknowledge the role, authority, and power of a commissioner role Focus on getting the best, in all respects
II. <i>Good practices not always practiced so well</i>	Take a bespoke (made-in-measure) approach to evaluation Use standards, criteria, and guidelines appropriately Be sure that stakeholders are engaged in all aspects of the evaluation Ensure quality of evaluation undertakings Focus on evaluation use and influence
III. <i>Broadening the perspective</i>	Consider the bigger picture Use a positive, strength-based approach to evaluation Keep an open mind

rules for selection of evaluators, are still practiced, and propose ways for how to overcome these (II. *Good practices, albeit not always practiced so well*). Lastly (III. *Opening the perspective*), it will be a matter of how to innovate, addressing the new challenges put by the political landscape of the programs.

Steps commissioners could take to enhance the relevance, quality and use of evaluation

I. Overall imperatives and guiding principles for action

Acknowledge the role, authority and power of commissioners

It is the commissioner who arguably has more power and authority than anyone else to control what will happen with an evaluation. But commissioners often do not fully appreciate this, nor act accordingly.

A corollary of this is that, as Catsambas & Davidson (2024) discuss, the commissioner's role requires significant knowledge and expertise, responsibility, judgment and taking of decisions. This is far more than a simple administrative task that can be passed on to an underling, at least not without significant guidance, as happens all too often. And with predictable results. Are commissioners always aided in acquiring the necessary competences to be able to fulfill this role appropriately?

Given the power and authority of a commissioner to determine the direction of an evaluation, how it should be undertaken and by whom, then it is appropriate that credit or blame for how well the evaluation has been undertaken fall mainly with the commissioner, rather with others engaged in the evaluation process.

Focus on getting the best in all respects

When undertaking an evaluation, presumably one would want the best in all respects – the best possible evaluation design, undertaken in the best possible way,

undertaken by highly qualified experts, the best possible collaboration between the evaluation commissioner/manager and the evaluator, with appropriate engagement of stakeholders to ensure the relevance and use of the evaluation (Perrin & Boschen, 1991).

Again, this guidance strikes us as self-evident. But why does this so often just not happen? We will just briefly touch upon two important considerations: the importance of a collaborative rather than an adversarial approach and dealing with bureaucratic rules and requirements.

When the evaluation is viewed as a partnership, there should be a close working relationship between the evaluation consultant and the commissioner. Such a relationship can provide a good idea how the evaluation is progressing, and emerging findings and any issues at the earliest possible opportunity. It also provides opportunities for maximum input, and for discussion of how best to deal with the inevitable uncertainties as they arise.

Thus, it is helpful to be aware of the best possible approach that would result in the best possible evaluation. Invariably there may be need for compromises, as virtually all organizations have procurement rules and requirements that must be followed, even though many of these rules have not been imposed with evaluation in mind. But as we have identified in an earlier book in this series (Perrin, 2021, p. 202), “No system is perfect, and every good bureaucrat who gets things done knows how to ‘work the system’ to their advantage.... Effective bureaucrats indeed know how to get around rules and requirements, without quite contravening them”.

When characteristics of procurement rules appear problematic, might it appropriate for these procedures be the subject of an evaluation? Nevertheless, as we have discussed elsewhere (Perrin, 2021), there can be strong resistance to evaluation of internal bureaucratic processes, and even “independent” evaluation functions often seem loath to undertake such evaluations.

II. Good practices, albeit not always practiced so well

Take a bespoke (made-in-measure) approach to evaluation

Evaluations need to be made to measure in order to be able to address the specific information needs of the potential users of the specific evaluation, taking into account the context in which it is being undertaken. The mantra of utilization-focused evaluation (Patton, 2008) is: intended use for intended users.

This seems to be a no-brainer. If an evaluation is not focused around the information needs of the stakeholders, how likely is it that it will be useful? Yet this does not always happen. In what we might refer to as “serial evaluations”, evaluation approaches seem to be replicated from one situation to another. There are numerous Terms of Reference issued by the same agency that all sound much the same, mandating a very similar approach to evaluation.

A bespoke approach to evaluation has numerous implications. Firstly, this implies use of some form of a needs analysis, where stakeholders have an opportunity to identify what they could learn and use from an evaluation. This also

implies evaluation tailored around these questions and information needs, rather than something predetermined. For such an approach to be meaningful, most likely to address the real information needs of stakeholders and to be used, there would be a need for considerable engagement throughout the evaluation process with the key stakeholders.

A bespoke approach contrasts with a template, or standardized approach that treats programs as essentially alike and requires all evaluations to adhere to a very similar or even identical approach. Such a one-size-fits-all approach runs the risk of not being relevant or useful to anyone, with, at best, limited buy-in and interest to the evaluation process and in its use. Sometimes well-intentioned guidelines concerning what every evaluation should cover become, inadvertently, barriers to more targeted, and more appropriate evaluation approaches.

This does not mean that one should reinvent the wheel every time one designs an evaluation. One, of course, should learn from similar approaches that have been used elsewhere – but then, rather than taking a cookie-cutter approach, engage stakeholders in designing an evaluation approach that addresses the unique needs of the specific situation, questions and priority issues.

This might, perhaps, require more effort, and expertise, than applying a readymade design (or following a predisposed table of contents for the report). Certainly, more effort will be required up front in the evaluation planning. But a bespoke approach can help refine and focus the scope of the evaluation on the specific considerations of concern in that context, rather than requiring each evaluation to address a multitude of imposed questions that may distract from an evaluation focus. Thus, a bespoke evaluation potential can be more streamlined. And, given engagement of stakeholders in agreeing on the right questions, it is much more likely to result in use, representing increased effectiveness and value for money of the evaluation.

Use standards, criteria and guidelines intelligently rather than mechanically

Are standards, criteria and guidelines a help – or a hindrance in planning and undertaking an evaluation? In practice, they could be either, depending upon how the commissioner approaches and manages the evaluation.

There are various iterations of evaluation standards and guidelines, too numerous to list here. These can be immensely helpful, to a commissioner in planning an evaluation, and to all involved in evaluation during its various stages. For example, the OECD/DAC (2010) *Evaluation Quality Standards* identify general categories of quality that can be very helpful when planning and undertaking an evaluation and in considering deliverables and findings. Most major evaluation associations, as well as many large organizations, have identified various ethical norms and standards that would be helpful to bear in mind. Guidelines are more useful when used as guidelines, to help suggest questions and considerations that an evaluation should address.

The OECD/DAC evaluation criteria for development evaluation are a case in point. These are intended to be used intelligently rather than mechanically or

rigidly, such as highlighted in the revised *Criteria* of 2021 that states that: “When evaluating development co-operation programs and projects it is useful to *consider* [these] criteria” (OECD DAC, 2021; emphasis added).

Principle One: The criteria should be applied thoughtfully to support high-quality, useful evaluation.

Principle Two: Use of the criteria depends on the purpose of the evaluation.

(...) Different methodologies can be used when evaluating based on the criteria, recognizing that the criteria form just one part of the full spectrum of evaluation norms and standards.

(OECD/DAC, 2021: 10)

Not all criteria might be applicable, in the same way, in every evaluation context.

Unfortunately, too often DAC criteria are mandated, and used, mechanically rather than intelligently, as intended. Applying standards and criteria in a rigid manner is a sign of lazy management. This replaces thinking about what is really important and which criteria would matter most in a particular evaluation context.

Rigidly structuring evaluations around standards and criteria can lead to a sameness across many evaluations that does not do justice to any of them, yet another example of serial evaluations. For example, one of us was involved in the review of evaluations carried out by a large international agency. The evaluations reviewed had budgets ranging from \$20,000 to \$2M, with the evaluations taking between three months to multi-year. But the reports, irrespective of significant differences across the scope and foci of these evaluations, all mainly sounded much the same, likely due to a similar organization of all reports around the DAC criteria. Without otherwise being aware of the budgets and timeframes for these evaluations, the reader might be forgiven for wondering what the differences were.

It is within the power of the evaluation commissioner to ensure that evaluation make use of guidelines, standards and criteria appropriately rather than as rigid “rules”. And, that as we have suggested above, with each evaluation focusing on questions of specific importance.

Be sure that stakeholders are engaged in all aspects of the evaluation

There are many reasons why stakeholders should be engaged in the evaluation process. For one thing, for stakeholders who might be in position to act upon the evaluation in some way (or conversely, to inhibit or even block this), engagement in the planning process can enable these stakeholders to identify what they would like to get out of the evaluation, what questions or concerns are of primary relevance.

While engagement might seem most pertinent to the actual evaluation than to the role of commissioners, it is commissioners who set the tone for what is expected, or might even be considered acceptable, and it is commissioners who have lead responsibility for the early planning stages of the evaluation, as well as for later follow-up action.

There is general, if not quite universal, recognition of the need to engage key stakeholders in some way in the evaluation. But engagement, to be meaningful, needs to go beyond just providing key stakeholders with an opportunity to comment on drafts. This minimal level of engagement rarely is sufficient for stakeholders to feel any part of the evaluation, or to develop buy in.

Similarly, engagement should include, but go beyond, just those stakeholders within the part of the organization most directly affected by the evaluation. Invariably, there are other departments, and partners, who also might be affected by the evaluation, who also need to be engaged in a meaningful way. Beyond this, it is also important for the community affected by the initiative, including intended beneficiaries and maybe the larger civil society, and who also should be meaningfully engaged in the evaluation process, at a minimum, to be given an opportunity to indicate how the evaluation can take into account *their* concerns and issues that may well differ from those of the funding agency.

The commissioner has a major role to make sure that appropriate engagement does take place. All too often, involvement with a broader group of stakeholders is limited, indeed often of a token nature with no real opportunity to influence the focus of the evaluation. For example, external stakeholders may be interviewed or even invited to meetings to comment on evaluation plans, but usually after these have more-or-less been determined.

If one cares about how evaluation can support a democratic approach, then meaningful involvement of civil society is essential, and that evaluation addresses the interests and needs of civil society including potential and actual beneficiaries. But it is important to bear in mind power imbalances. Otherwise, it is inevitable that the evaluation will instead serve the interests of those with the power. In an earlier book in this series, one of us (Perrin, 2018) differentiated between the most common form of accountability – up to the funder/donor who already holds the power, rather than accountability down to the community.

Enabling meaningful involvement of the community in evaluation can be challenging, indeed threatening, for evaluation commissioners who most often are part of the system. Nevertheless, most public service organizations and governments do say that they support democracy and addressing interests and needs of the community, including those most marginalized. Various statements of evaluation standards also make this point. Thus, there is much that a commissioner can do to make true engagement of the community in evaluation possible.

Ensure quality of evaluation undertakings

The evaluation commissioner has a critical role in ensuring quality, which should be a key consideration throughout *all* phases of evaluation, right from the planning stages. A good process, involving a collegial relationship, with ongoing contact, is most likely to enable discussions, and as need be, modifications, to support quality.

For example, as we have discussed, Is there a high-quality evaluation design and Terms of Reference, with appropriate engagement of stakeholders? Has a competent evaluator, with the requisite expertise, been chosen, using a process that

prioritizes competence and a track record for ensuring a quality approach over other possible criteria, for example, price? Sure, one can always find someone willing to work for less, nominally promising the same form of outputs. But rarely would this lead to a high-quality evaluation undertaking. And without this, any evaluation will be useless, at a minimum not representing good value for money; worse, leading to faulty policies and programs. To emphasize, quality refers to *all* aspects of the evaluation, including but by no means limited to technical considerations.

An increasingly common way of ensuring quality is to use one or more forms of quality assurance. In a larger organization, informal discussion among internal staff, as well as a more structured review by other staff, could be seen as a valuable peer support process.

Increasingly, an expert quality assurance reviewer external to the organization is engaged to review and comment upon draft deliverables. Such reviews often are restricted to draft reports. But often there could be opportunities for a more constructive approach, with at least some engagement during the evaluation process that might help suggest ways of enhancing quality. While many organizations employ rigid templates, we have found that it is generally most useful to keep these to general categories rather than rigid criteria or rating systems that might not necessarily apply to all evaluations.

Again, the commissioner has a critical role to play in ensuring quality, for example, being aware of any real or potential or even perceived concerns about quality, while the evaluation still is under way and there still is time as needed to address these.

Focus on evaluation use and influence

The evaluation field is increasingly recognizing the importance of a focus on how evaluations can exert influence and be used. Evaluation is a practical undertaking, and without use of some form, there is little point to it. As utilization drives, or should drive, most evaluation, this is a critical consideration.

Most of the better-known evaluation standards and guidelines identify the importance of a focus on use as part of an evaluation, with an extensive literature on evaluation use, well beyond our scope to discuss here (e.g., Alkin & King, 2016; Alkin & King, 2017; King & Alkin, 2019; Patton, 2020).

Rarely does evaluation use happen just in response to a formal final report. This has implications for design, methods, engagement and more. This has important implications for how evaluations should be planned and undertaken – and for the role of the commissioner. In particular, the commissioner needs to build in considerations regarding utilization from the beginning of the evaluation planning process. Waiting until near the end to start thinking how an evaluation could be used invariably is too late.

Much use, and influence, happens informally throughout the evaluation process. This is one of the reasons why we have highlighted the importance of a collaborative approach and engaging stakeholders. It is a main responsibility of the evaluation commissioner to ensure that an evaluation is focused on being used.

While some form of use, even on an informal basis, hopefully starts early in the evaluation process, the commissioner needs to take a lead role, after the evaluation is formally completed and the final report accepted, to help encourage and facilitate meaningful use of the evaluation.

III. Broadening the perspective

Consider the bigger picture

An earlier book in this series: *From Studies to Streams* (Rist & Stame, 2006) observed that most useful information comes not from individual evaluation studies, but rather from a multitude of sources and rarely all at once such as in a standalone study. Nevertheless, this still largely seems to be the main way in which evaluation is being undertaken. This limits the potential for greater learning.

For example, is it really necessary for the same agency, such as seems to occur often in the development aid area, to commission many separate evaluations of different but similar programs, so often taking very similar approaches? Could there be ways of taking a bigger picture approach, where specific evaluations are complementary rather than carbon copy duplicates of one another, the serial evaluations we referred to above? This requires a different mindset, looking at ways of complementarity rather than seemingly unthinkingly copying what's being done elsewhere.

While this might seem inconsistent with the bespoke approach we proposed above, the contrary might be the case, where individual evaluations need not be copies of what has been done elsewhere, while still providing opportunities for mutual learning across different evaluations and different programs. This is consistent with a realist synthesis approach (Pawson, 2013).

To be sure, evaluation syntheses, drawing upon themes from multiple studies and sources of information, can help (e.g., chapter by Perrin in Rist & Stame, 2006). Planning in advance, with an open mind, can create even more opportunities for learning out of the box.

Use a positive, strength-based approach to evaluation

Evaluation is often viewed as focusing on the negatives, on identifying the inevitable weaknesses in most any program or endeavor. There now is increasing interest in positive thinking approaches that highlight what is going well, as well as what might be improved to result in even more positive effects (e.g., Stame, 2014; Stame & Lo Presti, 2015; Preskill & Catsambas, 2006; Perrin, 2014). Positive thinking approaches do not ignore the inevitable problems or “the bad stuff”, but frame these in a positive, constructive context that is more likely to encourage buy-in to an evaluation and action on its implications.

A positive thinking approach to evaluation makes use of positive dynamics that are more likely to lead to positive effects. This is consistent with a utilization-focused approach to evaluation, where the stakeholders are energized to discover

ways to make things better. Stame (2014) and Stame & Lo Presti (2015) list a variety of possible approaches; perhaps the best known of these is appreciative inquiry. Working within this latter approach, Catsambas & Davidson (2024) show how the positive consequences of upholding such approach extends also to commissioning.

Keep an open mind

During the course of most evaluations, things can change, in small and not-so-small ways. For example, key managers and other influential stakeholders and their interests can change, as well as policy priorities, key interests of stakeholders, and the context is likely to change in at least some way. To remain relevant, evaluations need to adjust to changing needs and priorities. Good evaluation also should take an iterative approach, making adjustments to the evaluation approach based upon emerging findings. Early evaluation feedback also may elicit questions or considerations that were not anticipated initially.

Thus meaningful evaluation should take an intelligent, emergent approach and be open to adaptation during the evaluation process, especially when unexpected consequences may offer, as Rogers (2022) noted, an “effective leverage of emerging opportunities to contribute to important wider impacts”, which implies a change in the commissioners’ approach, going beyond intended outcomes and exploring how change actually can happen. This has major implications for the commissioner’s role, going beyond checking compliance with the original theory of change to reflect new developments and changes from the initial context and assumptions.

Particularly with evaluations of complex or multiyear initiatives, the commissioner might anticipate the possible need for eventual modifications. Some evaluation designs are more adaptable than others to modification. Some evaluators are more open to feedback and to adapting initial approaches than others. Contracts could provide for some flexibility so that changes such as to methodology could be made without requiring contract amendments.

Conclusion

In this chapter, we have highlighted the critical importance of the commissioner’s role that is often underrated. There are many things a commissioner can do to help ensure that evaluations will be relevant and useful, and address the needs of citizens, in support of democracy. It is the commissioner who determines what evaluations will be done, their scope and budget, establishing terms of reference, handpicking the evaluation team, while providing guidance and supervision as needed throughout the evaluation, and with lead responsibility for ensuring utilization. Given the power and role of the commissioner in shaping all aspects of the evaluation, this role is critical, and it rarely would be credible to blame others if things are not done well.

When commissioners acknowledge the power and authority of their role, they then would be in position to play a role in enhancing the relevance, and use, of evaluation.

References

- Alkin, Marvin C., and Jean A. King. 2016. "The Historical Development of Evaluation Use." *American Journal of Evaluation* 37 (4): 568–79.
- Alkin, Marvin C., and Jean A. King. 2017. "Definitions of Evaluation Use and Misuse, Evaluation Influence, and Factors Affecting Use." *American Journal of Evaluation* 38 (3): 434–450.
- Brousselle, Astrid. 2022. "Planting Seeds for Change." *Evaluation* 28 (1): 9–10.
- Catsambas, Tessie T., and Jane Davidson. 2024. *Evaluation Management*. Thousand Oaks, CA: Sage.
- King, Jean A., and Marvin C. Alkin. 2019. "The Centrality of Use: Theories of Evaluation Use and Influence and Thoughts on the First 50 Years of Use Research." *American Journal of Evaluation* 40 (3): 431–58.
- Organisation for Economic Co-operation and Development / Development Assistance Committee (OECD/DAC). 2010. *DAC Evaluation Quality Standards*. In *OECD Evaluating Development Co-operation: Summary of Key Norms and Standards*. Paris: OECD.
- Organisation for Economic Co-operation and Development / Development Assistance Committee (OECD/DAC). 2021. *Applying Evaluation Criteria Thoughtfully*. Paris: OECD.
- Patton, Michael Quinn. 2008. *Utilization-Focused Evaluation*. Thousand Oaks, CA: Sage.
- Patton, Michael Quinn. 2020. "Evaluation Use Theory, Practice, and Future Research: Reflections on the Alkin and King AJE Series." *American Journal of Evaluation* 41 (4): 581–602.
- Pawson, Ray. 2013. *The Science of Evaluation*. London: Sage.
- Perrin, Burt. 2006. "How Evaluation Can Help Make Knowledge Management Real" In *From Studies to Streams*, edited by Ray C. Rist and Nicoletta Stame, 23–45. New Brunswick, NJ: Transaction Publishers.
- Perrin, Burt. 2014. "Think Positively! And Make a Difference through Evaluation." *Canadian Journal of Evaluation* 29 (2): 48–66.
- Perrin, Burt. 2018. "Accountability's Two Solitudes and the Questions It Raises: Accountability to Whom? By Whom? And for What?" In *The Evaluation Enterprise: A Critical View*, edited by Jan-Eric Furubo and Nicoletta Stame. London: Routledge.
- Perrin, Burt. 2021. "Conclusion: The Problematic of Bureaucracy and What This Means for Evaluation—And for Public Sector Leaders." In *Changing Bureaucracies: Adapting to Uncertainty and How Evaluation Can Help*, edited by Burt Perrin and Tony Tyrrell. London: Routledge.
- Perrin, Burt, and Kathryn A. Boschen. 1991. "Getting the Best from Consultants." In *Evaluation Methods Sourcebook*, edited by Arnold J. Love. Ottawa: Canadian Evaluation Society.
- Preskill, Hallie, and Tessie Tzavaras Catsambas. 2006. *Reframing Evaluation through Appreciative Inquiry*. Thousand Oaks, CA: Sage.
- Rist, Ray C., and Nicoletta Stame, eds. 2006. *From Studies to Streams*. New Brunswick, NJ: Transaction Publishers.

- Rogers, Patricia. 2022. "Changing the Substance and Image of Evaluations." *Evaluation* 28 (1): 24–26.
- Stame, Nicoletta. 2014. "Positive Thinking Approaches to Evaluation and Program Perspectives." *Canadian Journal of Program Evaluation* 29 (2): 67–86.
- Stame, Nicoletta, and Veronica Lo Presti. 2015. "Positive Thinking and Learning from Evaluation." In *Evaluation and Success*, edited by Steffen Bohni-Nielsen, Robert Turksema, and Peter van der Knaap. New Brunswick, NJ: Transaction Publishers.
- Valters, Craig. 2014. "Theories of Change in International Development: Communication, Learning, or Accountability?" *JSRP Paper 17*. London: Asia Foundation.

3 How evaluation commissioning processes can support or undermine democratic processes

Patricia Rogers and Steve Montague

Introduction

This chapter provides an analysis of how evaluation commissioning might be done in ways that better support democratic processes. This includes a range of activities including explicitly labeled evaluations, activities labeled as a “review” or an “analysis”, performance audits, informal, internal evaluative activity, which might include monitoring or reflection sessions, and evaluative work done by communities such as community scorecards. We also discuss the implications for evidence synthesis, especially in terms of its framing in terms of “what works” and “best buys”.

We focus on three key issues, which we see as central to evaluation and democracy. The first is ensuring transparency in the evaluation, including its processes, products, methods and governance. The second is ensuring time, resources and legitimacy for meaningful engagement of participants and community members during an evaluation. The third relates to ensuring that evaluations pay attention to heterogeneity, and report findings beyond average or aggregate effects; there are democratic implications of commissioning processes which require experimental or quasi-experimental evaluation designs and methods as these often lead to an excessive focus on average effects and disregard for differential effects on marginalized populations. We argue that achieving at least *some* basic level of these three elements – transparency, engagement and attention to heterogeneity – is important for evaluations to effectively support democratic decision-making.

Ways of thinking about how evaluation can support or undermine democracy

We draw on some different ways of thinking about what constitutes a democracy-oriented evaluation. This includes MacDonald’s (1977) early contrast of democratic evaluation (designed with community engagement and control) with autocratic evaluation and bureaucratic evaluation both of which had limited transparency with control vested in the external expert evaluator (autocratic) or the officials commissioning the evaluation (bureaucratic). We also draw on House and Howe’s work (2000a, 2000b) which identified three essential elements of democratic

DOI: 10.4324/9781042000968-4

This chapter has been made available under a CC-BY-NC-ND license.

evaluation – inclusion, dialogue and deliberation – and the further development by Fraser and Rogers (2017) which focused on transparency, inclusion and dialogue in analyzing the South African government approach to evaluation.

We also draw on Hanberger's typology (2006) which distinguished between three types of democracy. Elitist Democracy-Oriented Evaluation, which can be understood as "evaluation for the people", is designed to support elite learning, rational feedback and accountability by those charged with making decisions for the people. Participatory Democracy-Oriented Evaluation, which can be understood as "evaluation by the people", focuses on self-learning, self-determination and empowerment. Discursive Democracy-Oriented Evaluation, which can be understood as "evaluation with the people", focuses on collective learning, justification and public debate involving all legitimate stakeholders.

We see that different forms of evaluation might be seen to support democracy – from evaluations that focus on providing information to elected officials so they can make transparent, evidence-informed decisions to evaluations that explicitly support active dialogic and deliberative engagement with communities or support their self-determination. Undemocratic evaluation, by way of comparison might be absent or token. Absent evaluation can occur at a national level, for example apartheid South Africa where government did not engage in formal evaluations (Fraser and Rogers, 2017), or for an area of government investment such as fossil fuel subsidies where there is rarely any government-led evaluation (Black et al., 2023). Token evaluation includes reviews undertaken by external agencies which are done for compliance purposes only, with limited public access to the findings or scrutiny of the process.

The rest of the chapter focuses on three elements we think are needed for an evaluation to be considered democracy-oriented – transparency, engagement and attention to heterogeneity. How evaluations are commissioned – governance, initial planning, procurement, design, managing implementation, reporting and supporting use – can either support or inhibit these.

Transparency

Transparency refers to openness of information about intentions, plans, rules, processes, actions and results. Transparency in evaluation includes both the products of an evaluation and the processes, including contracting and commissioning processes.

One key question regarding transparency is the extent that there is public reporting that an evaluation is planned or has been undertaken. For example, in Australia all open "approaches to market" including evaluations and evaluation-related services are published on a publicly accessible website (tenders.gov.au/atm) which provides details about the project. Canada has a similar process called CanadaBuys and a proactive disclosure requirement where contracts for evaluations must be listed (open.canada.ca).

Transparency can include providing a public list of evaluations being planned, underway and completed. For example, South Africa at one stage published an

annual National Evaluation Plan which identified evaluations to be undertaken in that year and provided an update on the status of evaluations included in previous plans.

Transparency includes providing information about how evaluations have been done, both in terms of methods for data collection and analysis, and processes for governance and quality assurance. Gauthier et al. (2022) in their review of 96 Canadian evaluation studies found that only a small fraction of them described the governance of the evaluation process. Furthermore, very few (and fewer over time since 2015) described data collection and analysis methods.

Transparency also relates to the findings and data of evaluations – not just the high-level conclusions but the details as well, and access to enough data for independent scrutiny without breaching confidentiality commitments. In practice, few evaluations provide details of the evidence behind causal and evaluative claims and even fewer provide anonymized data sets that can be re-analyzed to verify conclusions.

There may, however, be times when full findings and data are appropriately not fully shared. In an adversarial political system, there can be damaging effects from reporting details of problems which can be misrepresented as indicating overall failure. One of the authors faced this in an evaluation where part of the evaluation had identified faults in implementation. While the issue was named in the public report, an internal report set out details of the gaps, deficiencies and delays so these could be reported in detail to staff who could address them, balancing full transparency with utility.

Evaluations that draw on detailed case studies present particular challenges to transparency. Completely anonymizing the sources and the case might be appropriate when sources are deemed confidential or protected or likely to result in breaching commitments about anonymity but it can lessen the transparency about the credibility of the sources or the relevance of the case. It has been the experience of both authors that direct sharing with case area actors can help get learning messages across which might be difficult to put in a fully public report. The point is to be as transparent as possible with those who are closest to the issue and are directly affected by it.

At a minimum, there should be transparency about the fact that an evaluation has been undertaken (except in exceptional circumstances) and sufficient information should be provided to make clear the basis for evaluative judgements and recommendations.

Engagement

The second issue relates to the extent to which commissioning processes provide scope for meaningful engagement of participants and community members.

It can be useful to think of this in terms of a continuum, drawing on the typology of five types of evaluation developed by Wehipeihana (2013) in relation to Indigenous evaluation. Evaluation done *TO* communities involves collecting data from them without involving them in any way in the decisions about the evaluation

or in using it, and to meet the objectives of other stakeholders. Evaluation done *FOR* communities is done to improve the situation for them, but with the evaluator making decisions about the evaluation without reference to their values about what is important or what constitutes credible evidence – Evaluation done *WITH* communities involves some community members in the process of evaluation, but non-Indigenous people are in control of the process. Evaluation done *BY* communities has Indigenous people in control of the process, but they are also accommodating Western values and notions of credible evidence. Evaluation *AS* community is based on community views on what is valued and what constitutes credible evidence. It does not exclude Western values or notions of credible evidence but includes them only as far as is seen to be useful. There is no automatic or presumed right of participation by non-Indigenous people or approaches, only by invitation.

The commissioning process is critically important for ensuring there are resources, authority, incentives and opportunity to support meaningful engagement. Resources include sufficient time – in terms of both timeframes that provide adequate lead-time to individuals and communities being engaged, and contracted time to ensure there can be sufficient time planning and implementing engagement processes. Resources also include sufficient funding to support the additional time required for the evaluation team to engage in consultations and to appropriately pay for other participants' time and travel expenses where required. Resources also include expertise within the evaluation team to appropriately engage, including addressing power imbalances, possibly needing an expansion of the team. Commissioning processes need to include governance processes consistent with engagement.

The reality often falls far short of this. In 1991, Robin McTaggart (1991) described his unsuccessful experiences to ensure that the contractual arrangements for the democratic evaluation he was being contracted to undertake were followed. While his proposal had clearly set out the process that would be followed, including a consensus process to finalizing the evaluation report, the standard government contract required that the contracting agency had full control over the report and could make changes to its findings, even where these were factually incorrect. In 2002, one of the authors (Rogers) had a major evaluation delayed for over a month to renegotiate the standard government contract, which waived moral rights, entitling the commissioning agency to make changes to the final report while keeping the authors' names on it.

While there is often a recognition of the need for participation in evaluation processes, all too often this is superficial or poorly done and real engagement of diverse interests and knowledges is not possible. An analysis of the South African government's national evaluation system (Fraser and Rogers, 2017) showed that despite the stated commitment to supporting democracy through evaluation, this was mostly in the form of consultation and public reporting rather than through supporting dialogue across difference and deliberation processes, which include supporting community members and stakeholder groups to take in new information and understand different viewpoints and use this to make informed decisions.

Tight timelines for commissioning and conducting evaluations make this even more difficult. For example, in Canada the norm for many organizations is to conduct evaluations within budget fiscal years – inclusive of extensive commissioning and reporting /approval processes which often reduce the time available for design, collection, analysis/synthesis and reporting. Departments are encouraged to do evaluations that cover many programs and spending dollars within fixed time periods since evaluation units and directors have been rated on the amount of departmental funding that they have “covered” via evaluation in a fiscal year. This is a metric that has been a traditional part of financial audit and has been sometimes invalidly applied to evaluation. This role confusion – leading to inappropriate management oversight of audit and evaluation functions in Canada – has been noted by Barrados, Montague and Blain (2023) and by Shepherd and Champagne (2025).

When time is tight, expected coverage is wide and resources are limited, engagement often suffers. This is especially true in terms of engagement with those most remotely connected to the program or initiative. For some programs these may be the people the program is most trying to reach. So tight resources and timelines for projects intended to cover a wide scope may be creating systematic bias against appropriate key actor engagement – let alone against overall engagement.

In Canada, these pressures have combined to severely limit engagement of any kind with key interest and rights holders. The confluence of limited budgets, tight timelines, a superficial understanding of what evaluation “does” partially due to the co-housing of evaluation units with audit units (Shepherd and Champagne, 2025), means that it has often been seen as neither practically convenient, nor culturally appropriate to engage directly with key interest holders (e.g., recipient or target groups) outside of line departments in the conduct of evaluations. This has been true even for areas such as gender-based analysis, which has been an agenda item for over a decade in Canada. Recent reviews have noted a lack of follow-through commitment to truly inclusive designs (Whynot, 2021).

Adequate time and resourcing are needed to link with relevant networks and organizations to support engagement, to help in design, access and collection of information, interpretation of findings and reporting – leading to use in various areas.

We need to ensure meaningful involvement of all key actors in an evaluation of a policy / program / initiative as the default starting position for evaluation planning and design. This should include policy, program and service delivery agents and users/beneficiaries as well as those involved in establishing the enabling environment for policies and programs to take place. This should also include consideration of who does the evaluation. In order to meaningfully address the above principles, a carefully selected, diverse and at least somewhat representative community needs to be engaged.

Indigenous communities across the world have noted problems with conventional evaluation attempts or in some cases pretenses for engagement (Evans, Bremner, Johnston, Rowe, and Saskamoose, 2020). Evans et al. (2020) have identified questions for consideration to help guide the planning and implementation of an evaluation. The questions are intended to be asked throughout an evaluation

process to ensure that the approach is community driven and prioritizes relationship building.

Three of their questions are as follows:

- *How will the community and program staff be engaged at the outset and throughout the study to ensure that they are contributing significantly to the evaluation or research? How will these relationships be developed?*
- *What are the reasons for undertaking the evaluation or research study? How does it give back to the community or address the needs of the community or program?*
- *How will the principle of reciprocity be implemented? How will you give back to those involved in the process (e.g., gift giving, capacity-building through mentorship, hiring or training of those within the community)?*

The questions clearly position evaluation as a community driven, reciprocity-seeking, relationship-focused collective learning experience. The questions also represent key principles. While Evans et al. suggested these principles in the context of Indigenous evaluation, they would seem appropriate for most if not all improvement-focused evaluations.

While not all evaluations can aspire to deliberative democratic levels of engagement, at least minimum levels of engagement are needed to ensure that community perspectives are understood and incorporated. While this would almost certainly extend the time it takes to *do* evaluations, it would improve their validity and could also reduce the summary dismissal of findings as being those of “out of touch” elites. This should in turn improve the likelihood of study acceptance and use.

Heterogeneity

While the first two issues we discuss refer to evaluation processes, this issue relates to methods and methodology. It relates to democracy being not only about majority rule but also about protecting the rights of minority groups (BII, n.d.). Our understanding of democracy is not the tyranny of the majority but a process that recognizes and protects the rights of minorities and of those most vulnerable and in need.

Evaluations and syntheses of evaluations, which aim to inform policymakers and managers about “what works”, with the intent to “scale up” “what works” or “best buys” are at serious risk of harming groups whose experience of programs are different – either by implementing programs which are ineffective or harmful for them even though they work on average, or by terminating programs that are effective for them but that don’t work on average.

Implementing what works on average often does not work well for the most marginalized individuals and communities – for example, family support programs that worked on average were found to be harmful for mothers with multiple risk factors, and the same is true for many other programs (Westhorp, 2012).

Focusing only on aggregate outcomes can be similarly problematic. For example, Canada is reported to have a very high level of internet use (Pew, 2022) but there remain significant portions of the large country which are only weakly covered or not at all covered by internet service (CCA, 2021). This creates a digital divide as more and more public services and functions are delivered online.

Despite the claim that “carrying out randomised trials can help strengthen our democracy” (Leigh, 2023) through increased transparency, we think it can actually undermine it unless they are part of an evidence system that recognizes heterogeneity in how data are collected, analyzed and reported in individual evaluations and in syntheses of evidence.

Tanasoca and Leigh (2024) have laid out the argument that explicit evidence for policy resonates with three key democratic principles: non-arbitrariness, revisability and public justification. Leigh (2023) has articulated this position:

By building a strong evidence base for programs, citizens can see that government is crafting programs based on what works, rather than blind ideology or partisan self-interest. It is no coincidence that authoritarian regimes have been the most resistant to science and evidence. Building a better feedback loop demonstrates to the public that the focus of government is on practical problem-solving. And because the results of randomised trials are intuitively easy to understand, they bring the public into the tent, allowing everyone to see what works.

We would argue that these three criteria – non-arbitrariness, revisability and public justification – are important for democratic governance but not sufficient. Explicit attention is also needed to heterogeneity to ensure the needs and rights of non-typical people are addressed.

Randomized controlled trials (RCTs) can be done in ways that address heterogeneity, by ensuring sample sizes permit subgroup analysis for pre-identified groups, and by investigating and reporting other important differential effects even where the numbers are small, and by incorporating RCTs findings with a range of other relevant information. But this is not always done – and these problems become worse when heterogeneity is lost in study summaries and in evidence syntheses.

For example, the Victorian Government’s Menu of Evidence for Children and Family Services (2025) was developed to provide “searchable and independent information on evidence-based practices and programs that have been demonstrated to improve outcomes for children, young people and families experiencing vulnerability”. It reported that the Early Head Start Program was found to be effective, and that negative effects were “not found”. This misrepresented the report findings which had identified important heterogeneity. For the families with the highest levels of disadvantage it was not only ineffective, but actually harmful, producing worse cognitive and socio-emotional outcomes for children than those in the control group.

Guidance provided for such syntheses sometimes deliberately ignores heterogeneity. The highly influential GRADE (Grading of Recommendations, Assessment, Development and Evaluation) handbook, recommended by the Cochrane Collaboration, only uses the term “strong recommendation” to refer to recommendations undifferentiated by context; recommendations that might be labeled as “conditional” or “qualified” which take into account contextual variations are labeled “weak recommendations” even where supported by strong evidence.

Commissioning processes often present a barrier to addressing heterogeneity. It is problematic when the stated purpose is to find out “what works” and then scale this up or embed it in policy without investigation of heterogeneity. The key evaluation questions, methods and designs either prescribed in the terms of reference or clearly preferred can also be a barrier to exploring the nature of effects and impacts of policies, programs and interventions on different groups. The driving question needs to be “What works to what extent for whom in what conditions and why?” rather than “What works”. This means more than simply adding additional treatment arms and subgroup analysis to RCTs. It requires paying attention to a range of contextual variations, using a realist and/or actor-based perspective and drawing on evidence synthesis approaches such as realist synthesis which can include and incorporate diverse types of evidence (Pawson et al., 2004; Pawson, 2006; Westthorp et al., 2015).

The types of methods and designs endorsed or preferred by commissioning agencies need to eschew conventional hierarchies of evidence to clearly focus on approaches and methods that suit the particular situation in terms of: the decisions to be supported and key questions to be addressed, the nature of the program/policy and intervention (archetypes, complexity etc.) and the resources and constraints for the evaluation (Rogers, 2009; Stern et al., 2012). This includes especially drawing on a wider range of designs and approaches to causal inference in impact evaluation, both for individual evaluations and for evidence syntheses (Wimbush, Montague and Mulherin, 2012; Riley et al., 2018).

At a minimum, evaluations need to be designed in ways that can report findings that include important differentiations and discuss the implications in recommendations.

Summary

Commissioning processes can either constrain or enable the ability of evaluation to serve democracy. Commitments to transparency, engagement and addressing heterogeneity should be included to some extent in all evaluations that aim to support democracy, even where this is in the form of serving the information needs of those who make decisions on behalf of the community.

As evaluations seek to provide information and useful processes to support decision-making in increasingly complex and contested context, we think there is an increasing need to go beyond this to create the enabling conditions that will support democratic deliberation.

Transparency involves openness about the evaluation's intentions, processes, and outcomes, including how evaluations are commissioned and reported. In practice, many evaluations lack adequate transparency in reporting methodologies, governance and detailed findings, making it difficult for independent scrutiny. A base level of transparency is crucial for all democratic evaluations in order to ensure that the public can see how decisions are informed by evidence rather than ideology or special interests.

Engagement focuses on meaningful participation and dialogue among diverse stakeholders, especially marginalized groups, to ensure balanced representation in evaluation processes. Commissioning processes often limit resources, time and genuine engagement, resulting in superficial participation rather than true deliberation. Successful deliberative democratic evaluations foster inclusivity and facilitate discussions that can influence decision-making, especially in complex contexts.

Attention to heterogeneity has implications for the methods and methodologies used in evaluations, emphasizing the need to focus on the diverse impacts of interventions rather than averaging effects. As such, a reliance on average outcomes (e.g., such as those produced by randomized controlled trials or elegant statistical models) can misrepresent the effectiveness of programs for specific populations, necessitating a shift toward more context-sensitive evaluation approaches and reporting. As a minimum, evaluations and syntheses of evaluations should identify how different groups are affected by programs to ensure that the needs of non-average participants, and especially marginalized communities, are not overlooked – and draw on this in making recommendations.

Table 3.1 provides a first version of a rubric presented for further discussion and development, and which might be a useful tool for framing an individual evaluation. It sets out three levels for each of these three elements. In each case we suggest that as a minimum each of the issues should be addressed at least in terms of “some”.

In order for evaluations to effectively support democracy, evaluations and synthesis of evaluations must achieve at least some level of transparency, engagement and attention to heterogeneity.

We set out below what we see as promising practices for commissioning evaluations to support democratic processes which could be adopted at various levels and in different contexts related to evaluation:

- 1 Formulate evaluation questions that promote collective learning and stakeholder engagement.

We have found that the realist question of “What works to what extent with whom under what conditions and why?” focuses evaluations on appropriate areas and inherently includes a discussion of who is reached and affected by a given set of interventions and who is not – and it also promotes a learning focus and the inclusion of all voices in the commissioning, design and delivery of evaluations.

- 2 Provide sufficient time, resources and linkage opportunities to support some level of engagement.

Table 3.1 Levels of enabling conditions

	<i>None</i>	<i>Some</i>	<i>Much</i>
Transparency	Procurement process, budget, reports and data not disclosed.	An RFP is public and the final report but little information on design, methods, process or data.	Details of the evaluation design and the workings of the evaluative judgements (how values and evidence were brought together) are explicit and some of the data are made available so the evidence can be tested by others (while protecting sources, privacy, confidentiality).
Engagement	Participants and community members are just treated as sources of data.	Perspectives of participants and community are explored and used as part of the evaluative framing and interpretation.	Full-scale deliberative engagement or self-determination.
Attention to heterogeneity	Evaluations only focus on the average effect or aggregate value and recommend that for everyone.	Identify differences but don't work through what this means in terms of contextually relevant recommendations.	Realist evaluation or realist synthesis which identifies and seeks to understand variations and explicitly pays attention to this in recommendations.

Source: Authors.

This will probably mean increasing the time to develop terms of reference, commissioning and select designs. It will probably mean the use of an iterative and generative approach to all stages of the collection, interpretation/analysis/synthesis and reporting – which can be a departure from traditional review processes.

- 3 Ensure meaningful involvement of all key actors in the evaluation process, considering who conducts the evaluation.

Given the need to focus on learning and the need to provide time for inclusive, discursive processes – it naturally infers that a wide (wider than typical) net be cast in terms of governing, commissioning, designing and delivering evaluations and involving groups and people who are directly affected by the evaluation and the programs and policies under review.

- 4 Consider the use of “community engagement-oriented” guiding principles to at minimum complement conventional evaluation principles

This might involve learning from the participatory and discursive evaluation practices used by Indigenous and other community focused groups. The emphasis should be placed on wherever possible encouraging community-driven and relationship-focused evaluation approaches.

These promising practices represent just the beginning of the emerging role for evaluation as a true collective learning function and a true support to modern democratic processes. It is likely that more fundamental structural and strategic changes will be needed to governance, commissioning and other system supports in order to ensconce evaluation within public management as a true support to democratic processes.

References

- Barrados, Maria, Steve Montague, and Jim Blain. 2023. "COVID Crisis: Time to Recalibrate Evaluation." In *Policy Evaluation in the Era of COVID-19*, edited by Pearl Eliadis, Indran A. Naidoo, and Ray C. Rist. New York: Routledge.
- Black, Simon, Antung A. Liu, Ian W. H. Parry, and Nate Vernon. 2023. *IMF Fossil Fuel Subsidies Data: 2023 Update*. Washington, DC: International Monetary Fund.
- Bureau of International Information Programs (IIP), U.S. Department of State. n.d. *Principles of Democracy: Majority Rule, Minority Rights*. www.principlesofdemocracy.org/majority.
- Council of Canadian Academies. 2021. *Waiting to Connect*. October 28. <https://cca-reports.ca>.
- Evans, Jane, Larry K. Bremner, Andrea L. K. Johnston, Gladys Rowe, and JoLee Sasakamoose. 2020. *Exploring Indigenous Approaches to Evaluation and Research in the Context of Victim Services and Supports*. Ottawa: Department of Justice Canada.
- Fraser, Dugan, and Patricia J. Rogers. 2017. "Is Government's Approach to Evaluation Deepening Democracy in South Africa?" In *Democratic Evaluation and Democracy: Exploring the Reality*, edited by Donna Podems. Charlotte, NC: Information Age Publishing.
- Gauthier, Benoit, Simon Roy, Shelley Borys, and Natalie Kischuk. 2022–23. "Bridging 2015, 2018 and 2021: How Has the Policy on Results Modified Federal Evaluation Practice?" June 14, 2022, and May 25, 2023. <https://cutt.ly/15-18-21>.
- Hanberger, Anders. 2006. "Evaluation of and for Democracy." *Evaluation* 12 (1): 17–37.
- House, Ernest R., and Kenneth R. Howe. 2000a. "Deliberative Democratic Evaluation in Practice." In *Evaluation Models*, edited by Daniel L. Stufflebeam, George F. Madaus, and Thomas Kellaghan. *Evaluation in Education and Human Services*, vol. 49. Dordrecht: Springer. https://doi.org/10.1007/0-306-47559-6_22.
- House, Ernest R., and Kenneth R. Howe. 2000b. *Deliberative Democratic Evaluation Checklist*. <https://wmich.edu/evaluation/checklists>.
- Leigh, Andrew. 2023. "Evaluating Policy Impact: Working Out What Works." *Australian Economic Review* 56 (4): 431–41. Speech, August 29. <https://ministers.dewr.gov.au/leigh/evaluating-policy-impact-working-out-what-works>.
- MacDonald, Barry. 1977. "A Political Classification of Evaluation Studies in Education." In *Beyond the Numbers Game: A Reader in Educational Evaluation*, edited by David Hamilton, David Jenkins, Christine King, Barry MacDonald, and Malcolm Parlett, 105–108. London: Macmillan.

- McTaggart, Robin. 1991. "When Democratic Evaluation Doesn't Seem Democratic." *Evaluation Practice* 12 (1): 9–21.
- Pawson, Ray. 2006. *Evidence-Based Policy: A Realist Perspective*. London: Sage Publications.
- Pawson, Ray, Trisha Greenhalgh, Gill Harvey, and Kieran Walshe. 2004. "Realist Synthesis—An Introduction." *ESRC Research Methods Programme* 2: 55.
- Pew Research Center. 2022. "Internet, Smartphone and Social Media Use." www.pewresearch.org/global/2022/12/06/internet-smartphone-and-social-media-use-in-advanced-economies-2022/.
- Riley, Barbara, Aislinn Kernaghan, Lauren Stockton, Steve Montague, Jenna Yessis, and Cameron Willis. 2018. "Using Contribution Analysis to Evaluate the Impacts of Research on Policy: Getting to 'Good Enough.'" *Research Evaluation* 27 (1): 16–27. <https://doi.org/10.1093/reseval/rvx037>.
- Rogers, Patricia J. 2009. "Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation." *Journal of Development Effectiveness* 1 (3): 217–26.
- Shepherd, Robert P., and Eric Champagne. 2025. "Delivering Results for Canadians: Improving the Conditions of Enabling Functions". *Canadian Public Administration*, 67 (4).
- Stern, Elliot, Nicoletta Stame, John Mayne, Krister Forss, Rick Davies, and Barbara Befani. 2012. *Broadening the Range of Designs and Methods for Impact Evaluations*. Department for International Development (UK).
- Tanasoca, Ana, and Andrew Leigh. 2024. "The Democratic Virtues of Randomized Trials." *Moral Philosophy and Politics* 11 (1): 113–40.
- Victorian Government Department of Families, Fairness and Housing. 2025. *Menu of Evidence: Children and Family Services*. Accessed March 3, 2025. <https://menu.dffh.vic.gov.au/>.
- Wehipeihana, Nan. 2013. "A Vision for Indigenous Evaluation." Keynote address, Australasian Evaluation Society Conference. www.slideshare.net/slideshow/a-vision-for-indigenous-evaluation-keynote-presentation-at-the/27047707.
- Westhorp, G. (2012). "Using Complexity-Consistent Theory for Evaluating Complex Systems". *Evaluation*, 18(4): 405–420.
- Westhorp, Gill, Bill Walker, and Patricia Rogers. 2015. "Dealing with Complexity in a Realist Synthesis: Community Accountability and Empowerment Initiatives." In *Dealing with Complexity in Development Evaluation: A Practical Approach*, edited by Michael Bamberger, Jos Vaessen, and Estelle Raimondo. Thousand Oaks, CA: Sage Publications.
- Whynot, Jane. 2021. *Where's Betty? Integrating Gender-Based Analysis Plus (GBA+) in Canada's Federal Government Evaluation Function*. PhD diss., University of Ottawa.
- Wimbush, Erica, Steve Montague, and Tamara Mulherin. 2012. "Applications of Contribution Analysis to Outcome Planning and Impact Evaluation." *Evaluation* 18 (3): 310–29.

4 Commissioning with clarity

Empowering democratic spaces by drawing on the principles of deliberative democracy

Tom Ling and Ananda Millard

Introduction

Commissioning evaluation is structured by power and interests. This reflects and refracts not only immediate organizational interests and administrative requirements (including the demands of policy makers) but also meso-level processes such as professionalism of evaluation and the structure of evaluation systems and macro levels including patriarchy, class and race. These power relationships are inscribed in each invitation to tender, each proposal and each evaluation. Ensuring that evaluation is not unsupportive for democratic debate requires intentionality in the commissioning process to lean evaluative activities toward pro-democratic practices and away from reinforcing vested interests.

This book discusses several ways in which evaluation can contribute to strengthening the democratic space. This chapter focuses on one critical aspect of this set of challenges; the commissioning process and environment. Although it relates to wider issues, this chapter is primarily intended to provide practical guidance for commissioning and conducting evaluation and should be of interest not only for commissioners but also policy makers, users of evaluation and evaluators themselves. Commissioning is a fundamental part of the evaluation ecosystem, and poorly managed commissioning can be a barrier to the efficiency, relevance and the contribution the evaluation function can make to democratic spaces. This chapter will examine the practical steps required to help commissioners ensure that evaluation contributes to the public good, provides opportunities for enhanced public debate, and is clear about its reach and validity. The intention is to enable an open dialogue between commissioners and evaluators about the role that different actors can play in the evaluation process.

We approach the question of how evaluation can empower democratic spaces from a position informed by deliberative democracy; that is to say that democracy should be characterized not only by the outcomes of decision-making but also by the processes through which decisions are made. There are a variety of positions on deliberative democracy¹ but we would (broadly following Fishkin²) summarize the characteristics of deliberative democracy as:

1 Decisions should be informed by information that is accurate and fair

- 2 Decisions should not be driven solely or mainly by arguments determined by one stakeholder or another
- 3 Those who are most affected by decisions should be prioritized to have their voices included in the process
- 4 Decisions should be made following a sincere effort to understand opposing views
- 5 No views should be discounted because of a dislike of the people who hold them

By “democratic space” we mean the practices and institutional settings that enable all individuals and groups to have a voice and to speak to those who hold power, and to hold decision-makers to account. This includes the rules, regulations, laws, cultures and policies that shape the lives and experiences of citizens and the groups to which they belong. This also includes both the context of state actors and the civic spaces within which individuals and groups organize and work in pursuit of individual or shared interests.

Much of the effort in the evaluation environment is currently focused on ensuring that the evaluation responds to questions identified by the commissioners and serves the commissioner’s needs. Of course, this is critically important, but we have found that far less attention is paid to how evaluations prioritize certain interests and delegitimize others, the real-world implications of this, and the role that evaluated parties, including right holders (beneficiaries), of activities or processes being evaluated play in evaluations. Commissioners of evaluation may be less sensitive to these issues; they necessarily reflect the immediate (real world) pressures and interests facing their organization and the wider relationships between their organization and right holders. Key questions such as “who is the evaluation aiming to inform?” and “can the evaluation design inform the intended audience?” are specifically relevant when assignments are focused, in part or mainly, on vulnerable and, or, hard to reach populations and, or, on programming that is hampered by contextual factors.

Whether intentional or not, evaluations are implicated in democratic processes. Evaluations meet the information needs of some but not others, they support some sorts of decisions but not others, and they empower certain actions and disempower others. In this chapter we argue, first, that there is a well-developed set of arguments and values that have focused on deliberative democracy that can be applied to the principles under-pinning the commissioning and conduct of evaluation. Second, these values should inform each stage of the sequence of questions that a commissioner of evaluations should ask. Third, we will argue that the use of rubrics provides a straightforward way to absorb these values and ensures that they are applied in a transparent manner during the implementation of the evaluation.

Who are commissioners and what questions should they ask themselves?

In common with the other chapters of this book (as described in Chapter 1: Stame & Bustelo, 2027), we use the term “commissioning” here as a shorthand for the whole process describing how the need for an evaluation is first identified, how requirements are specified and potential providers identified and then selected, how the evaluation is managed by the funder, and how results are mobilized.

Organizations commission evaluations for different reasons, but they tend to interpret unnecessarily narrowly (we believe) these aims:

- 1 improving the commissioners' organization's ability to deliver the intervention
- 2 improving the commissioner's organization's ability to engage with other stakeholders (partners)
- 3 understanding the results of the intervention (outputs, outcomes and/or impact) for learning and accountability (including accountability to right holders) focused on the intervention
- 4 identifying lessons that are transferable to future efforts (other interventions)
- 5 identifying who should use the evaluation findings and whether the findings pursued are valuable to the expected user

Of course, commissioners may decide that an evaluation should address just one, some or all of these. To support commissioners to identify objectives that are able to provide the results they are pursuing and provide added clarity to this process, we propose that the following five questions be answered during the design of any evaluation. The intention would be to expand the perceived purposes of evaluation and consequently enrich their content. These questions are identified in Figure 4.1.

We will explore each of these in turn but first:

A brief word on participatory and discursive approaches

In Rogers and Montague (2027) chapter (this volume) the distinction between participatory and discursive approaches is made. They identify an elitist approach in which the expert evaluator identifies the singular truth about the value or worth of the thing being evaluated. They distinguish this from a participatory approach according to which those who the evaluation concerns participate in the evaluative

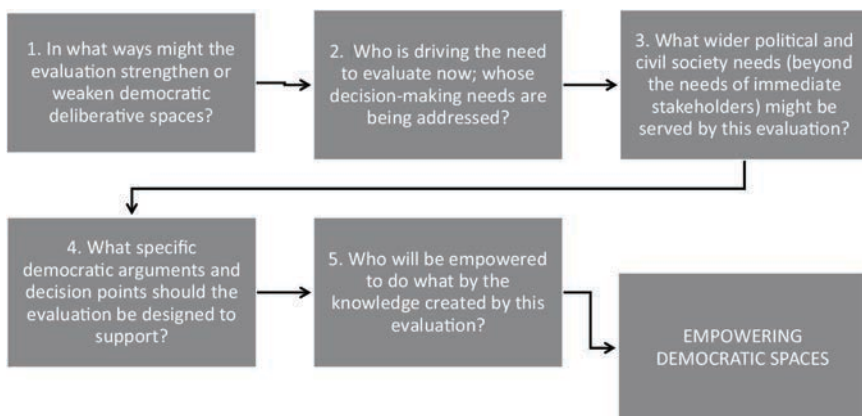


Figure 4.1 Five questions to support commissioners of evaluation in contributing to civic empowerment and democratic spaces.

process. They then distinguish both of these from a discursive approach. Discourse theory is a broad camp but in all cases, it attends to how meaning is constructed within domains of social action – and these will often differ, reflecting power relationships, how knowledge is produced, and the subjectivities of those involved. Not only will different stakeholders make sense of the world through different discourses, but it should be remembered that evaluation is itself a discourse, requiring some humility and a resistance to the elitist conception of evaluation. Accordingly, participation is important but insufficient if it does not also seek to explore, articulate and make visible the differing discourses among those involved. Our proposed approach to this is to develop a rubric that can help bring clarity and transparency to the commissioning and implementation of evaluations as described below, with a specific focus on how right holders are engaged and when.

Question 1. In what ways might the evaluation strengthen or weaken democratic deliberative spaces?

Evaluations inevitably impact on democratic deliberative spaces. Acknowledging that this happens is a key first step toward empowering democratic spaces. However, evaluation as a practice is not able to answer every democratic question; rather, it generates knowledge within recognized routines of data collection and analysis. A key element of these routines concerns “evaluability”.

Evaluability is the extent to which a program, project or activity is designed with sufficiently clear intended consequences, and sufficiently clear means to deliver these consequences, to allow a judgment to be made about whether it was successful. It also concerns the availability of data to answer the evaluation questions. If there is no clarity about what is intended, and how outcomes are to be realized, then to that extent it may not be evaluable in principle. Even if this is clear but there is no realistic possibility of collecting the data needed to arrive at such a judgment, then to that extent it is also not evaluable. The point is that there are many important democratic questions where evaluation will have little, if anything, to contribute.

In addition, we should consider not only whether the proposed evaluand is evaluable but also how the evaluand sits within the wider disposition of power and competing discourses. This is especially relevant in cases where the intended outcomes have multiple causes, and the intervention is only one (small) part of this causally dense context. The idea of “causal density” is used by Woolcock to describe programs that

are characterized by high uncertainty, which is a function of the numerous pathways and feedback loops connecting inputs, actions and outcomes, the entity’s openness to exogenous influences, and the capacity of constituent elements (most notably people) to exercise discretion (i.e., to act independently of or in accordance with rules, expectations, precedent, passions, professional norms, or self-interest).

(Woolcock 2022)

In such causally dense cases a lack of certainty should not prevent an evaluation from being commissioned, but it should encourage an open, participatory, and iterative evaluation approach where evidence is collected “on the fly” and the evaluation approach is adapted accordingly. In such cases, a well-designed evaluation will recognize that “knowledge” is constituted and carried differently, and that social inequalities will most likely shape engagement, and that multiple organizations, groups and individuals may all engage with the program in accordance with their own rules, expectations, professional norms and so forth. This will also be an evaluation that is well suited to contributing to the democratic space. The concept of evaluability reminds us that the contribution of evaluation to deliberative democracy will take place within a particular “bandwidth” but, especially in “causally dense” evaluands, this should also be sensitive to the wider power dynamics and structures of social and political life.

However, causally dense the context might be, evaluations should be commissioned with an understanding of what it would be most useful to find out, taking into account the question “useful for whom?”. Obviously, from an accountability point of view, there is a need to evaluate whether accountable persons took appropriate actions. But the opportunity for learning will be maximized where there is an understanding of what new knowledge or information would add the most value, and how the evaluation might contribute to this knowledge. To take a simple example, if a vaccine has already been rigorously assessed in clinical trials, the important new evaluation questions for a vaccination program will focus not on repeating the clinical trial but on the logistics of delivering the vaccine, the approaches to engaging communities to support uptake, measures to reach the so-called “hard to reach”, the skills and capacities required to implement the program, and so forth. The question might well be not only will the evaluation contribute new knowledge but also for whom. This is often about how the problem is framed. The problem might not be that some people are “hard to reach” but that some services are hard to reach for many intended beneficiaries.

Question 2. Who is driving the need to evaluate now? Whose decision-making needs are being addressed?

The decision about when to commission an evaluation with a particular end date is highly consequential. Complex interventions, in particular, rarely have outcomes that are linear over time and deciding when to measure outcomes matters greatly. In our experience, interventions take time to get set up, they are slow to achieve their intended outcomes as cynicism and doubt need to be overcome but then suddenly take off after a critical mass of support has been secured. Conversely, following a bout of initial enthusiasm and visible success, leadership moves on, attention is redirected, and the system reverts to business as usual. Furthermore, new approaches rarely develop in a vacuum, and the context of a complex program is likely to change, requiring further evaluation to understand its continuing relevance. Along this arc of impact, different groups may need to gain knowledge about the program at different times. However, the timing of evaluations is very

often chosen to fit with decision-making cycles that are relevant to only one stakeholder; the policy maker. Democratic debate and accountability are diminished if we fail to understand these dynamic processes.

Consider the concept map in Figure 4.2 which concerns how local innovation systems support (or not) the uptake of proven innovations in the English health and care system. It attempts to summarize the system enablers and system activities that help organizations deliver innovations that add value. The details are less important than the point that there is no obvious best moment to conduct an evaluation. Understanding how well the program met patient preferences or engaged clinicians' views might be done earlier than (say) building a whole system approach to innovation adoption. By deciding to evaluate some three years into the life of the program, some questions could be well answered, but some (for example, increasing local institutionalization of innovation) could only be weakly assessed.

Commissioners therefore need to consider the optimum time to evaluate but this is not only driven by the maturity of the intervention. In organizational life and in democratic spaces, decision-makers work to timescales that require evidence to be available at decision points unrelated to the evaluation, but this should not trump wider considerations of timing. In particular, by evaluating too early to meet decision points, the evaluation might evaluate the intervention at a point of maximum disruption (before benefits begin to flow) or, conversely, at the point of maximum enthusiasm (creating a false impression that the intervention should be scaled up). Commissioners must therefore have a deep understanding of the timing of potential outcomes as well as attend to the timing of policymaking and decision-making cycles.

Question 3. What wider political and civil society needs (beyond the needs of immediate stakeholders) might be served by this evaluation?

Polling suggests that across developed countries, most people do not believe that elected officials care about ordinary citizens.³ In the United Kingdom, 76% do not trust MPs,⁴ according to the UN global voting rates have fallen and protests have increased.⁵ This is despite the considerable resources committed to evaluating public services. There are many ways to respond to this if we wish to strengthen a belief in fairness and legitimacy (for example, citizen's juries, deliberative polls and so forth) but one simple and inexpensive solution would be to strengthen the relationships between evaluations already funded and civic life.

Commissioners could play a positive role in this respect. First, they can identify audiences for the evaluation and engage them early in the process, so they are "primed" and ready to pay attention to the evaluation results. Second, they can facilitate evaluators' access to interviewees, focus group participants, survey responders, managers and so forth to ensure that voices are heard. This often involves ensuring that stakeholders view the evaluation as important, potentially helpful and impartial. Third, given that evaluations almost never work exactly as planned, they can support adaptation and changes in direction, involving their

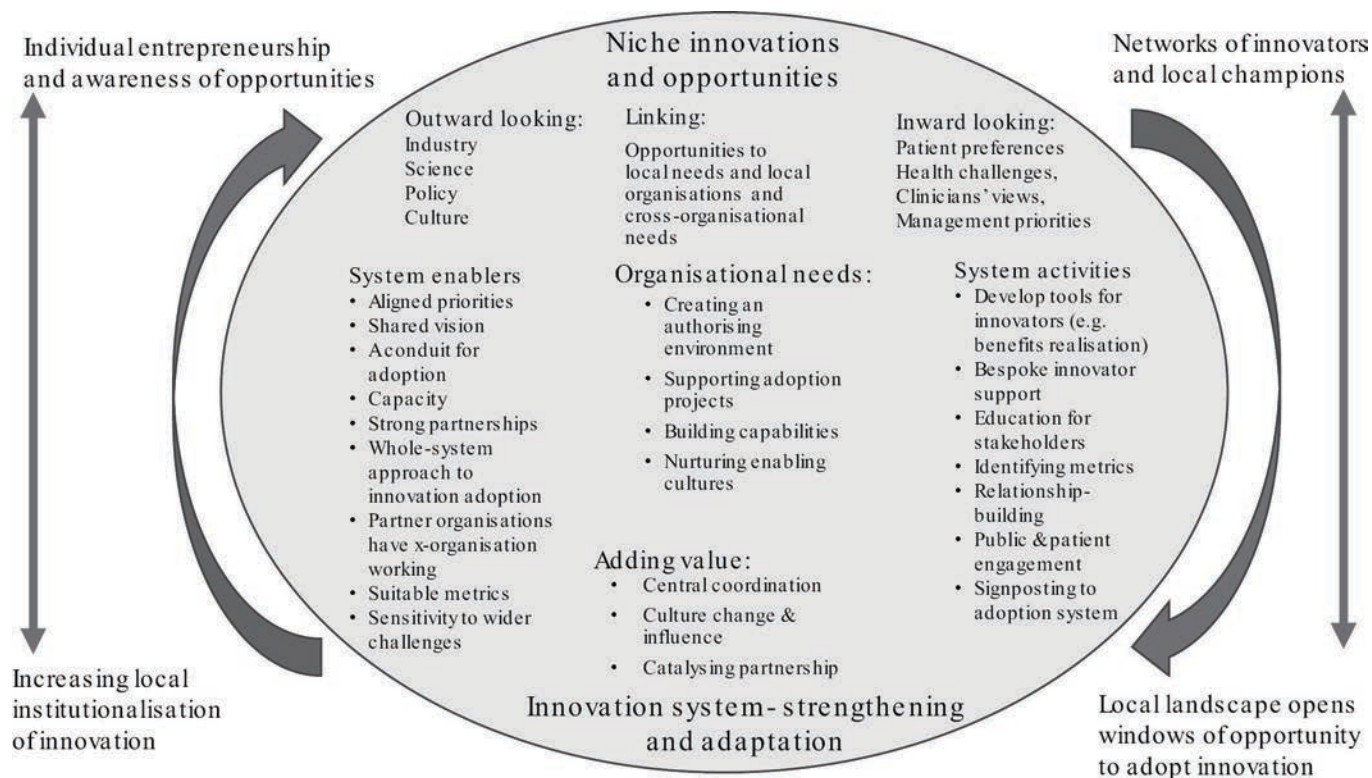


Figure 4.2 A concept map of an intervention to support the uptake of innovations in the English health and care system.

stakeholders along the way. Linked to this, they can establish “step back” moments to review lessons learned and to possibly redirect the thing being evaluated (with potential consequences for the evaluation approach). Fourth, they can help with communicating findings and mobilizing results among civic entities and informal groups. They may often understand the needs of those reading the reports better than evaluators.

Preparing the audience is only one part of an active communications strategy. How the findings are then communicated, how messaging might be segmented and targeted, and how the wider public should be informed all benefit from being addressed early in the evaluation process. A single, long, word-based document that is placed on a website is unlikely to meet all these needs. Events, toolkits, webinars, podcasts, private briefings, are just some of the communication tools to be considered.

We need to make a distinction between findings that tell us about the particular project or program being evaluated and findings that are designed to be transferable to other settings. Where the latter is important, this needs to be designed into the original invitation to tender. It is likely to include an understanding of how the new findings sit within an existing body of scientific knowledge. It should also include sufficient detail about the context of the intervention to know (a) how important context was in producing results, (b) which contextual factors were present in the intervention being evaluated and (c) how generalizable these factors are in wider settings.

It also includes an understanding, in complex interventions, of how many different agencies and decision-makers worked together to achieve the observed outcomes and how likely it is that these would behave similarly in other settings. And it includes an understanding of how service users and their communities engage with the intervention and how likely it is that other service users would behave similarly.

Question 4. What specific democratic arguments and decision points should the evaluation be designed to support?

The users of evaluation findings are rarely monolithic. For both ethical and pragmatic reasons, it is important to consider their varied needs. Commissioners are acting on behalf of others, and the better they understand these needs, the more likely it is that the evaluation findings will be useful. Key stakeholders for commissioners are (in no particular order): policy makers, implementing teams, service users, practitioners and, critically, right holders and beneficiaries.

The questions stakeholders might have will also vary, and even when they have the same question, their understanding of that question might be different (for example, each stakeholder might have a different idea about what “better”, “fairer” or more “integrated” services look like). No evaluation can answer all these questions equally and commissioners will want to prioritize some over others. Questions might be helpfully divided into “what” questions requiring largely descriptive data, “how” questions focused on causal mechanisms, and “who should do what” questions focusing on recommendations for improvement. The evaluation focus might be on impacts, processes or value for money (this should

be driven not only by what is considered to be “value” for different stakeholders but also by the distribution of costs and effort. The purpose of the evaluation might be to entrench accountability (did duty bearers do what they should have, why/not, and what should be done about this) or learning (did it work as expected, why/not, how might we do things better). The timing of the evaluation will also be relevant; is it taking place before decisions to implement are being taken, during implementation, or after the intervention is complete?

Given that resources for evaluation are always limited, and given that evaluability constrains what questions evaluation can answer, do evaluations optimally contribute to democratic debate? What the evaluation achieves will be a function of the evaluation questions asked, the methods used, and the thing being evaluated (the evaluand). Commissioners could be more clear about the decision points and democratic debates they hope to contribute to before finalizing their evaluation questions. It should also be noted that a “camel train” of one evaluation question after another may dilute evaluators’ efforts to such an extent that no questions are well-answered.

Today, many evaluations begin with a theory of change (often called “theory based evaluation”). This allows commissioners to ask “did the intervention work as intended?” and “which bits worked well and which bits did not” based on the presumed changes identified at the program design phase. It involves specifying the causal chains through which it is hoped the program, project or intervention will deliver its intended outcomes. Typically, this will lead to an assessment of the extent to which the evidence supports these claims. This, in turn can entrench accountability and support learning.

It can also provide the basis for a separate value for money exercise that arrives at a judgment about whether the outcomes and impacts justify the resources used. Where costs and consequences can be calculated independently of a theory of change, this can be conducted as a separate exercise. Democracy includes making decisions on the allocation of scarce resources and value-for-money studies should, in principle, contribute to this.

A particular challenge with this approach is that it places the emphasis on the presumption that the plan or theory was appropriate in the first place, and that pursuing an understanding of how the theory of change materialized in the real world is the most useful result for right holders/beneficiaries, for example. The practice of developing and agreeing a theory of change has enormous potential to enhance the participation of diverse stakeholders (including rights holders) in the design of the evaluation. However, this potential is often not realized in practice. A focus on what might be useful to right holders might help ensure that the results are relevant to a wider group of stakeholders.

Question 5. Who will be empowered to do what by the knowledge created by this evaluation?

As described earlier, knowledge and trusted information is fundamental to the practice of deliberative democracy. Knowing that your views are heard and, in return,

making an authentic effort to understand the views of others before making democratic choices are both fundamental. Our argument is that evaluations could contribute to this far more effectively than has most often been attempted. Reframing evaluative evidence as part of the evidence ecosystem that underpins a vibrant democratic polity would, we believe, add substantial value and would provide further justification for the resources society commits to evaluative activities. We do not think that evaluation can help address all democratic problems and nor do we think that the democratic purpose exhausts the value that evaluations bring.

This therefore brings into focus who, currently is empowered by evaluation and how might this be different. The institutional settings within which commissioning of evaluation takes place reinforces an elitist model of knowledge production and findings are communicated in a language and format that is most often exclusionary. The near universal use of executive summaries, the growing use of lay summaries, greater attention to visuals and graphics, shorter reports, searchable online documents and communications strategies have all moved things in the right direction. AI searchable evaluation reports will take this to a further level. All of this is desirable. However, problems arise much earlier and throughout the knowledge production process. One way to systematically address these issues is to use rubrics. This approach is described below.

Democratizing evaluation through a rubric model: a right holder-centric approach

In the preceding sections, we have delved into common objectives of evaluations and key considerations in their design. Now, we pivot to a critical aspect of evaluation – its democratic nature. Utilization-focused evaluation serves as our springboard, emphasizing evaluations that are actionable and beneficial for commissioners and stakeholders. This section introduces a rubric model that aids in clarifying the inclusivity and transparency of evaluation processes, particularly concerning right holders and beneficiaries. The rubrics broadly map onto our questions 1–5 listed earlier.

The rubric model

The rubric model presented below poses essential questions to enhance transparency and inclusivity in evaluations, which are intended to complement the previous questions introduced in the first part of this chapter. These questions are designed to identify and address potential gaps in engaging with right holders or beneficiaries, particularly those who are vulnerable or hard to reach. Question 5 articulates this concern.

Ensuring Inclusivity in Evaluation: The first issue of focus is the design of the evaluation. Specifically, what is being asked (question 1 presented earlier), when to ask (question 2) and possibly also questions about the added value of asking (question 3). Are right holders or beneficiaries who are difficult to reach or engage

with at the design of the assignment either determining the ToR or at least at the inception phase? And if yes to what extent? Do they have a role on determining which questions the evaluation seeks to answer? For instance, evaluations focusing on children's rights, gender equality or the rights of persons with disabilities often need to involve groups that are vulnerable or challenging to access. The model probes whether the evaluation methodology accommodates the inclusion of these groups, avoiding a narrow focus on more accessible cohorts.

Engagement in Data Interpretation and Mitigating the Echo Chamber Effect:

This aspect scrutinizes the extent to which right holders are incorporated into the data interpretation phase. It challenges the evaluation to ensure that right holders are not merely sources of data but active participants in making sense of the findings. By fostering a dialogue with right holders, the evaluation can transcend beyond the evaluators' perspectives, incorporating a broader spectrum of insights. This approach is pivotal in avoiding the echo chamber effect, where the evaluation might otherwise circulate and reinforce existing opinions and perspectives without introducing new insights or challenges. The model emphasizes a dynamic and interactive flow of information, where right holders' input serves as a critical counterbalance, ensuring that the evaluation embodies a multifaceted view rather than reflecting a narrow or self-reinforcing narrative. This element in the rubric is focused on the specific democratic arguments and decision points the evaluation is designed to support (question 4) and possibly the degree to which wider civil society needs might be served by this evaluation (question 3).

Access to Results: Does the evaluation framework ensure that right holders have access to the outcomes? While an evaluation may primarily serve the commissioner's needs, a democratic approach would also share findings with the right holders, fostering transparency and accountability. This has to do with who drives the evaluation (question 2), but also can influence the degree to which arguments and decision points the evaluation is designed to support democratic processes (question 4) and degree to which wider civil society needs might be served by this evaluation (question 3). In addition, who has access to results can also play an important role in addressing question 5: whose knowledge, in what social settings, will be enhanced by this evaluation and what will they be empowered to do. If results are not comprehensible or relevant to those with whom it is shared, then despite sharing results the impact of having done so will be limited. However, if right holders are engaged throughout, the likelihood of the value of sharing results being higher increases. Of course, even if the evaluation has been designed and commissioned in a non-transparent way, and results are shared in a meaningful way, this may still have an impact. In relation to sharing results, the increasing number of AI tools and broad and ever-expanding telecommunication networks make the sharing of information and accessible and inclusive ways increasingly easier. The value of sharing results, particularly when provision of goods and services are being evaluated, is considerable since right holders understanding what they have the right to will allow them

to increasingly hold duty bearers accountable, which in turn stands to increase the quality of results.

Operationalization of Results: This element of the rubric addresses all questions introduced earlier as to how the evaluation is designed and implemented will determine the value of the results and by extension the value of how results are operationalized. Key questions here include: Are right holders engaged in discussions about how the results are operationalized? The model emphasizes the importance of creating opportunities for right holders to influence how findings are applied, ensuring that the evaluation’s benefits are tangible and meaningful to those it aims to serve. The rubric itself can be modified from the example suggested here:

1. Accessibility and Inclusiveness

<i>Criteria</i>	<i>Not Met</i>	<i>Partially Met</i>	<i>Fully Met</i>
Inclusion of Vulnerable Groups	Vulnerable or hard-to-reach groups are not considered or included.	Some effort is made to include vulnerable groups, but there are significant gaps.	Comprehensive strategies are in place to ensure the inclusion of all relevant vulnerable or hard-to-reach groups.

2. Engagement in Data Interpretation

<i>Criteria</i>	<i>Not Met</i>	<i>Partially Met</i>	<i>Fully Met</i>
Involvement of Right Holders in Data Interpretation	Right holders are not involved in interpreting data.	Right holders have limited involvement or are only partially engaged in data interpretation.	Right holders are actively involved in interpreting the data, and their feedback is integral to the evaluation process.
Mitigation of Echo Chamber Effect	The evaluation process does not acknowledge or address the echo chamber effect.	There is some awareness of the echo chamber effect, but strategies to mitigate it are not fully implemented.	Comprehensive measures are in place to prevent the echo chamber effect, ensuring diverse sources and perspectives are considered in the evaluation.

3. Operationalization of Results

<i>Criteria</i>	<i>Not Met</i>	<i>Partially Met</i>	<i>Fully Met</i>
Engagement in Operationalization	Right holders are not involved in discussions on operationalizing results.	There is some level of engagement with right holders, but it is not comprehensive or fully integrated.	Right holders are actively engaged in discussions and decision-making on how results are operationalized.

Our rubric model is a step toward democratizing the evaluation process, highlighting the potential to shift toward a righter holder-focused approach. By addressing these critical questions, evaluators can design studies that are not only thorough and robust but also inclusive and transparent, ensuring that the voices and needs of all stakeholders, especially the most vulnerable, are acknowledged and addressed in the evaluation process.

Notes

- 1 For a helpful discussion on deliberative democracy, see: Dryzek, John S. et al. (2019).
- 2 Fishkin, James S. (2009). *When the People speak*.
- 3 Pew Research Center (February 2020). "Democratic Rights Popular Globally but Commitment to Them Not Always Strong". www.pewresearch.org/global/2020/02/27/attitudes-toward-elected-officials-voting-and-the-state/
- 4 Carnegie UK January (2022). "Lack of Trust in Politics Threatens Democracy: New Report and Poll". <https://carnegieuktrust.org.uk/news-stories/lack-of-trust-in-politics-threatens-democracy-new-report-and-poll/>
- 5 Gadgil A. (2022). "Do Governments Listen to Their Citizens?" UNDP Oslo Governance Centre. www.undp.org/policy-centre/governance/blog/do-governments-listen-their-citizens

References

- Dryzek, John S., André Bächtiger, Simone Chambers, Joshua Cohen, James N. Druckman, Andrea Felicetti, James S. Fishkin, David M. Farrell, Archon Fung, Amy Gutmann, Hélène Landemore, Jane Mansbridge, Sofie Marien, Michael A. Neblo, Simon Niemeyer, Maija Setälä, Rune Slothuus, Jane Suiter, Dennis Thompson, and Mark E. Warren. 2019. "The Crisis of Democracy and the Science of Deliberation." *Science* 363 (6432): 1144–46. <https://doi.org/10.1126/science.aaw2694>.
- Fishkin, James S. 2009. *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford: Oxford University Press.

- Stame, Nicoletta, and María Bustelo. 2027. "Commissioning Matters: Considerations and Challenges for Evaluation Commissioners Who Intend to Promote Democracy." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Woolcock, Michael. 2022. "Will It Work Here? Using Case Studies to Generate 'Key Facts' about Complex Development Programs." In *The Case for Case Studies: Methods and Applications in International Development*, edited by Jennifer Widner, Michael Woolcock, and Daniel Ortega Nieto, 87–116. Strategies for Social Inquiry. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108688253.006>.

5 **How evaluation commissioners and evaluators can collaboratively move equitable evaluation from a set of external actions to internal practices**

Drew Koleros, Monique Liston and Ray McGhee

Introduction

In the evolving landscape of social and economic sector support, US-based philanthropic institutions (hereafter referred to as “foundations”) have long been pivotal in funding organizations and crafting programs aimed at driving social change. Recently, there has been a notable shift toward focusing on achieving equitable outcomes across various sectors, from housing to health. This shift recognizes that addressing the root causes of complex societal issues, including structural racism, is essential to advancing equity.

As foundations intensify their focus on tackling systemic inequities, the adoption of “equitable evaluation” approaches has similarly gained momentum. Yet, the concept of equitable evaluation can be understood in varied ways by evaluators and evaluation commissioners, which often results in mismatched expectations and significant challenges. Central to this process is the principle of collaboration, which is essential for ensuring that both parties develop a shared vision and approach. Collaboration, when embraced as a democratic process within an evaluation, presents both challenges and opportunities: it requires negotiating differing perspectives and expectations, but it also offers the chance to build a more inclusive and aligned understanding of what equitable evaluation entails. By engaging in genuine, cooperative dialogue, commissioners and evaluators can bridge gaps in understanding and effectively advance equity in their evaluations.

This chapter explores the practical implications of equitable evaluation, emphasizing the importance of collaboration between commissioners and evaluators. It begins by discussing the evolution of equitable evaluation within the US philanthropic sector and its formalization in the Equitable Evaluation Initiative (EEI)’s Equitable Evaluation Framework – a key resource for guiding equitable evaluation practices.

The focus then shifts to understanding equitable evaluation as a shared set of internal practices rather than a series of external actions performed by evaluators alone. By examining the case study of an equitable evaluation conducted for

the Robert Wood Johnson Foundation (RWJF)'s support of the housing justice movement from 2020-2024, this chapter highlights the potential challenges that arise when commissioners and evaluators have differing interpretations of equitable evaluation. Through this case study, we aim to provide practical insights for both evaluation commissioners and evaluators, fostering a more nuanced and collaborative approach to equitable evaluation.

Centering equity in philanthropic evaluation

An evolving understanding of equity within the philanthropic sector

Centering equity in evaluation is not a novel concept within the field of evaluation. Since the 1980s, evaluators have published articles about using evaluation to create equity and conducting evaluations with equity in mind, advocating for equity concerning race, class, gender, sexuality, ability, language and many other cultural practices (Botcheva, Shih, & Huffman, 2009; Chouinard & Cousins, 2007; Greene, Millet, & Hopson, 2004; Hood, Hopson, & Frierson, 2005; Hopson, 2009; Hopson, Lucas, & Peterson, 2000; LaFrance, 2004; Kirkhart, 2005; Manswell-Butty, Daniel-Reid, & LaPoint, 2004; Reese & Vera, 2007; Smith & Jang, 2002; Mertens, 1999; Thomas & Stevens, 2004; Thompson-Robinson, Hopson, & SenGupta, 2004). The practice of culturally responsive evaluation (CRE) emerged from these efforts. CRE, as outlined by Frierson, Hood, Hughes, and Thomas (2010), is a comprehensive framework that places culture at the center of evaluation. CRE asks evaluators to decenter hegemonic ideas of success and normativity for more culturally centered ideas, norms and practices shared by those most impacted by the work of programs and initiatives, many of which are funded by the philanthropic sector. To be explicit, Hopson (2009) wrote,

CRE is a theoretical, conceptual and inherently political position that includes the centrality of and [attunement] to culture in the theory and practice of evaluation. That is, CRE recognizes that demographic, sociopolitical and contextual dimensions, locations, perspectives and characteristics of culture matter fundamentally in evaluation.

(p. 431)

CRE places special emphasis on addressing historically marginalized groups, aiming to center fairness and equity in the evaluation process. The Equitable Evaluation Framework, introduced to the sector in 2019, is the progeny of this intellectual genealogy and offers both opportunities and challenges for practicing the use of evaluation to create equity and conducting evaluations with equity in mind.

The genesis of the EEI itself is an example of collaboration as a democratic process, stemming from a collaboration between representatives from the philanthropic sector and seasoned evaluation professionals. Driven by a shared commitment to advancing equity, this collective endeavor sought to investigate the varied approaches foundations were employing to utilize evaluation

practices to promote equity. Undertaking a comprehensive research initiative, the dedicated group embarked on a journey to surface insights that could pave the way for the development of comprehensive guiding principles. These principles were envisioned not only to serve as a roadmap for integrating equity considerations into evaluations commissioned by foundations but also to foster a broader understanding of the nuanced intersections between philanthropy and evaluation in the pursuit of social justice. As the research gained traction and recognition within the philanthropic community, the initiative garnered additional financial backing from influential foundations. This infusion of support played a pivotal role in sustaining the momentum of the project and transforming it into a lasting institution. The institutionalization of this crucial work marked a significant milestone, culminating in the formal establishment of the EEI – an enduring testament to the collaborative efforts aimed at embedding equity as a cornerstone in the landscape of philanthropic evaluation practices, rooted in democratic principles.

Collaboration serves as a critical practice of equitable evaluation (EE), underscoring its dedication to inclusive and democratic evaluation practices. The Framework stresses that effective evaluation necessitates a partnership between commissioners and evaluators, with both parties actively involved in defining the evaluation's goals, methods and outcomes. This collaborative approach ensures the incorporation of diverse perspectives, leading to a richer and more comprehensive understanding of equity within the evaluation process. Through open dialogue and shared decision-making, this collaboration helps align expectations and empowers stakeholders to co-create solutions that address systemic inequities effectively. Such a cooperative approach is essential for evaluations that accurately reflect and respond to the needs of marginalized communities.

Collaboration connects to all three principles of EE. According to Principle One, evaluation and evaluative work must be dedicated to advancing equity. Effective collaboration among commissioners, evaluators and community members ensures that production, consumption and management of evaluations are aligned with equity goals. Collaborative efforts help address systemic inequities more comprehensively and contribute to meaningful progress toward equity. Principle Two emphasizes that evaluative work should be designed and implemented in ways that reflect equity values, including being multi-culturally valid and oriented toward participant ownership. Collaboration supports these values by integrating varied cultural viewpoints and ensuring that participants are actively involved in the evaluation process. Additionally, Principle Three highlights the need for evaluations to address critical questions related to historical and structural influences, the impact of strategies on different populations and the interaction between cultural contexts and systemic conditions. Collaboration brings together diverse expertise, enhancing the analysis of these complex issues and ensuring a thorough understanding of systemic drivers of inequity. Overall, collaboration helps align practices with equity principles, fostering the development of context-specific practices that advance equity effectively.

Over the last few years, EE has found its way into requests for proposals (RFPs), conference abstracts and book chapters such as this. The conversation, thanks to

the work of EEI, continues to evolve. While building awareness and increasing support for the practice is an admirable base-building goal, there are challenges with the philanthropic sector and the practice of equitable evaluation that need to be explored by the fields of evaluation and philanthropy. For instance, in their review and synthesis of the empirical literature on the evaluation of equity-focused community coalitions, Chouinard, Tovey, and Kidd (2023) found that “the cultural complexity and historical scope of each context, diversity of stakeholders and enormity of the systemic issues involved, shape and challenge the evaluation and research process in fundamental ways” (p. 50).

Challenges to practicing equitable evaluation in real-world evaluations

One important challenge that arises is when different stakeholders within an evaluation have different understandings of what it means to implement an equitable evaluation. These different understandings can be framed around the concepts of single-loop, double-loop and triple-loop learning (Argyris & Schon, 1974; Schwandt & Gates, 2021) (Figure 5.1). When equitable evaluation is implemented as an approach to single-loop learning within an evaluation, the focus is on operationalizing equitable evaluation principles through a set of actions that make a more conventional evaluation “more equitable”. This process might include reviewing the evaluation questions from an equity lens and involving affected communities in evaluation design, delivery and dissemination. Within the wider evaluation context, it is still typically foundation program staff who decide what investments will be evaluated and determine the main evaluation purposes and questions. Evaluation commissioners are then asked to hire and manage evaluators to implement the evaluation and ensure that they sufficiently operationalize equitable evaluation principles within the predefined evaluation. The information from the evaluation then travels from evaluator to commissioner to foundation program staff who then ultimately decide what will be done with the evaluation findings.

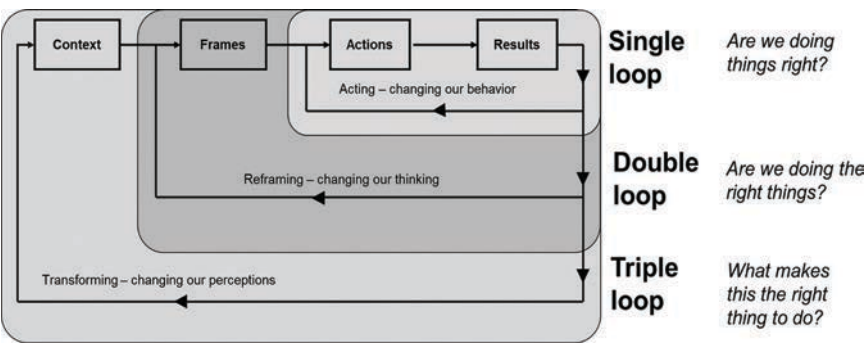


Figure 5.1 Single, double, triple loop learning.

Source: Authors own, adapted from www.mikethementor.co.uk/blog/2018/4/4/single-double-and-triple-loop-change

When equitable evaluation is implemented as an approach to double-loop or even triple-loop learning, the focus shifts from implementing a set of actions within the bounds of the evaluation to embodying a set of practices that allow for the examination of the perspectives, beliefs and assumptions of foundation staff that explain why the project was designed as it was, and their implications on how the project was ultimately delivered (e.g., which grantees received funding, for what types of activities, etc.).

In this situation, equitable evaluation is seen as a set of practices, requiring evaluation commissioners to work collaboratively with evaluators, grantees and foundation staff to examine these perspectives, beliefs and assumptions. Instead of being the “middle man” or a link in the information chain between foundation program staff and evaluators, commissioners are akin to the facilitators of a process that leads to more partnership between all evaluation stakeholders. This is consistent with how evaluators are discussed and described in more conventional evaluations (evaluators as experts) compared to system change evaluations (evaluators as facilitators of learning processes) (Schwandt & Gates, 2021). This is explored in more detail in the case study below.

Introduction to the housing justice evaluation

Housing justice component of RWJF’s community power initiative

The RWJF is one of the largest health foundations in the United States. With a mission to improve the health and well-being of everyone in America, for nearly five decades, RWJF has supported the development and promotion of health and healthcare innovations to improve the lives of all Americans. In regard to housing specifically, in recognition of the severe and persistent inequities caused by rampant and unchecked racism experienced by Black, Indigenous and People of Color in the American housing system, RWJF provided funding to select organizations committed to advancing housing justice (HJ) and building community power in 2020 with the purpose of supporting housing organizing, learning and development. This was part of a larger foundation initiative to build community power.

Housing justice evaluation

In 2021, RWJF commissioned an evaluation of their HJ investments. The RFP for the evaluation emphasized RWJF’s commitment to equitable evaluation principles and specifically called for an “equitable evaluation”, stating that RWJF “views evaluation as a tool to advance equity by applying principles of the Equitable Evaluation Framework” and referenced its adherence to EE principles in its contracting, collaboration, evaluation design, implementation and dissemination practices.

Through a competitive procurement process, the evaluation was awarded to UBUNTU Research & Evaluation, LLC (UBUNTU) and Mathematica in April 2021. UBUNTU is a Black-led evaluation firm working to intentionally

disrupt the White, male, academically hegemonic approach to program, policy and impact evaluation in the nonprofit and foundation sectors. UBUNTU's team brings lived and professional expertise in understanding power and positionality through evaluation activities, including through their own base-building work with Black-led organizations. Mathematica is a research and data analytics consultancy that operates at the intersection of data science, social science and technology, generating evidence-informed solutions that help clients and partners see clearly and act quickly. As an employee-owned company, Mathematica strives to improve public well-being by bringing the highest standards of quality, objectivity and excellence to bear on the provision of information collection and analysis to our clients.

The evaluation included three objectives: (1) to document the narratives of those who received funding to better understand how the money contributed to building the capacity and infrastructure of funded organizations, (2) to describe how changes in the capacity and infrastructure of funded organizations impacted the wider housing justice ecosystem's efforts to advance their collective housing justice goals and (3) to develop actionable insights for future funding and engagement of RWJF within the housing justice sector by identifying the ongoing capacity and infrastructure needs of the housing justice ecosystem to advance their housing justice-specific objectives and outcomes.

Centering equity within the housing justice evaluation team

As the author and social activist Grace Lee Boggs once said, "Transform yourself to transform the world". Given our evaluation team's previous experiences, we moved toward incorporating equitable evaluation first as a set of practices within our internal evaluation team. Recognizing the need for transformative change, we identified that our previous partnership, with Mathematica as the prime contractor and UBUNTU as a subcontractor, required a fundamental shift. The traditional prime-subcontractor model presented significant challenges, such as imbalanced power dynamics and inequitable resource distribution. To truly center equity in our HJET project, we needed to dismantle these existing inequities and forge a new, more collaborative approach.

One major challenge was addressing the inherent differences in our organizational cultures and practices. Mathematica, with its traditional evaluation methodologies, often operated within established frameworks and norms, while UBUNTU brought a radical, race-forward perspective that challenged conventional approaches. Aligning these differing approaches required careful negotiation and a willingness to rethink standard practices, such as contract structuring, resource allocation and the management of both internal and external communications. For example, Mathematica's standard subcontracting template included language that grants Mathematica as the prime contract holder with all rights to intellectual property (IP) developed by subcontractors over the course of a contract. This was not in line with the equitable partnership that we were trying to establish on HJET, so the Mathematica project team members worked with Mathematica's contract and

legal teams to revise the subcontract template so that all IP generated through HJET would be jointly owned by both Mathematica and Ubuntu.

The task of redefining leadership roles, agenda-setting and communication strategies highlighted the complexity of merging these distinct evaluation philosophies. There were many mistakes we made throughout this process and harm was caused in some instances between team members. But despite these early missteps, our team worked to center equity within HJET through putting the time in that was needed to build relationships among team members, a focus on learning and unlearning together and a commitment to open and honest communications within our team. For example, because Mathematica was technically the prime contract holder, in the initial start-up phases of the contract, Mathematica tended to lead on all client communications, setting meeting agendas and leading client meetings even though a team member from UBUNTU was named as the evaluation's principal investigator. It took time and discussion between our teams to recalibrate these efforts in a way that was more in line with equitable practices so that UBUNTU ultimately took over technical leadership of the evaluation and primary communications with the client, with Mathematica staff participating in a supporting role.

HJET prioritized relationship building between the team from the start in small and big ways. One small way we did this was through regular check-in questions. We would sometimes spend the first 15 or 20 minutes of our weekly one-hour team meetings discussing a check-in question that could span from identifying your favorite season to sharing what "home" meant to you. Over the weeks, this allowed us to get to know each other outside of our professional settings and start to make connections between us that improved our understanding of each other and how to work together.

Building strong relationships between Mathematica and UBUNTU was crucial for fostering effective collaboration. This involved creating a space for open discussions about the racial politics influencing our daily interactions and work environments. We engaged in conversations addressing unconscious biases and took steps to hold each other accountable for any racist, sexist or queerphobic language. Prioritizing radical honesty allowed us to cultivate an atmosphere of mutual respect and growth. This approach deepened our understanding of each other's perspectives and enhanced our ability to work together in an equitable and inclusive manner. Such open dialogue and accountability were essential for overcoming challenges and ensuring that our collaboration remained both meaningful and transformative.

We prioritized collaboration through quarterly team retreats, launched as the COVID public health emergency lifted, where we dedicated two to three days to advancing our evaluation work while building trust and deepening relationships. These retreats became essential spaces for open communication, allowing us to address difficult topics, exchange feedback and align our efforts. Between retreats, weekly team meetings maintained momentum, fostering consistent connection and collective progress. This rhythm of focused collaboration – intensive retreats paired with regular check-ins – strengthened our teamwork, encouraged mutual

learning and unlearning and centered equity in our approach, ultimately enhancing our ability to work transparently, effectively and with shared purpose.

Centering equity within the wider evaluation

As we struggled to practice equity within our internal team, we also faced challenges in moving equity toward a set of practices within the overall evaluation. This largely came from a disconnect between our evaluation team's understanding and the foundation program team's understanding of what it meant to take an equitable evaluation approach to HJET. We explore these in more detail below.

Investing the time necessary to build relationships between evaluation stakeholders. Building relationships among evaluation stakeholders required significant time and intention, a factor that was underestimated during the initial HJET evaluation design phase. Early on, our focus was on connecting with the evaluation commissioners, RWJF program staff and grantees. However, the time needed for relationship-building was not fully accounted for in the project timeline, leading to miscommunications and trust issues. From the HJET perspective, progress felt slower than RWJF expected, while the foundation was eager to see tangible outcomes after a lengthy tendering and contracting process. To address this, HJET prioritized regular meetings, thorough documentation and transparent communication. Meeting minutes, progress notes and memos were shared consistently, capturing emerging insights, adjustments and key decisions. Despite these efforts, HJET still felt we were falling short of fully meeting RWJF's expectations. However, over time, the commitment to consistent virtual meetings and detailed documentation helped build trust, fostering more open and honest conversations about the evaluation's goals and challenges.

As trust deepened, HJET uncovered internal tensions within RWJF, particularly between the program staff and the Research, Evaluation and Learning (REL) team. Program staff prioritized accountability – understanding how grantees used funds and RWJF's role in advancing housing justice – while the REL team focused on broader ecosystem-level learning and the foundation's strategic positioning within the movement. These differing priorities created internal power struggles, which became evident during critical evaluation milestones, such as finalizing evaluation questions. This misalignment placed HJET in challenging and emotionally charged situations as team members navigated conflicting expectations. However, continued transparent communication allowed the team to address these tensions openly, fostering a shared understanding of the foundation's internal dynamics. This experience highlighted the need for foundations to include internal reflections during commissioning processes, ensuring that equitable evaluation practices extend not only to external consultants but also within their own organizational structures.

Recognizing cost constraints in centering equity at every step. As the evaluation design phase unfolded, HJET recognized the budget didn't reflect the labor

intensity required to ensure equity was practiced, not just discussed. The budget primarily allowed HJET to address RWJF's evaluation questions about their housing justice work, without sufficient resources to include grantees' questions. For example, the team initially proposed a case-based design with site visits, but grantees expressed concerns that this approach could be a resource burden on their grantees. HJET also struggled with how to achieve the goals of site visits without being extractive, given the lack of established relationships with local housing justice groups.

In response to grantee feedback and budget constraints, the team shifted from deep-dive case studies to creating the Housing Justice Evaluation Fellowship Program. This fellowship provided housing justice organizers and advocates with evaluation skills to support their current work. The program included a four-day in-person workshop followed by virtual sessions over six months. Nine individuals from seven housing justice groups across the country participated. Through this fellowship, participants gained hands-on evaluation experience while continuing their organizing efforts. It also allowed us, as evaluators, to collect meaningful data, answer evaluation questions and build relationships within the movement, while enhancing evaluation capacity among local groups.

Addressing self-reflexivity with evaluation commissioners and program staff. Evaluation commissioners often hold key internal knowledge about the history, relationships and politics within their organization, shaping how programs are designed and how evaluations are commissioned. In traditional evaluations, these internal factors may not be relevant, but for equitable evaluations, in pursuit of triple-loop learning, they provide essential context. Commissioners must decide what information to share with evaluators and when to share it. From the foundation's perspective, inquiries from HJET around these topics felt out of scope, as if evaluators were asking them to "air their dirty laundry". However, from the team's perspective, this transparency was necessary to practice equitable evaluation.

At the start, program staff committed to partnering with the evaluation team and providing the necessary information, but their focus was primarily on grantees, not the foundation itself. This led to gatekeeping and limited communication with foundation staff. It took considerable time to agree on when and how HJET could engage with the program team directly. Typically, communications flowed internally within RWJF between RWJF program teams and the REL unit, before reaching external evaluators, making two-way communication difficult.

Grantees, however, had more interest in RWJF's role in the housing justice movement, asking about funding decisions and RWJF's future involvement. This shifted the evaluative focus from grantee actions to a deeper analysis of RWJF's role, requiring the team to adopt double-loop and triple-loop learning approaches. HJET sought to understand how the initiative was conceptualized and designed, which necessitated new norms of communication between program, REL and evaluation teams. These shifts led to some tension between the foundation and evaluators early on.

This last component was left largely unresolved. HJET ultimately lacked the power to fully explore and answer the grantees’ questions about RWJF’s role. Additionally, grantees expressed concerns that sharing critical feedback about the foundation might jeopardize future funding, further complicating our ability to address these questions effectively.

Implications for evaluators and evaluation commissioners

Based on this experience with HJET, several implications emerged for both evaluators and commissioners:

<i>Implication</i>	<i>Description</i>
Commit to Learning and Accountability	Establish a culture where continuous learning and mutual accountability are core principles. Encourage team members to view mistakes and conflicts as opportunities for collective growth and improvement. Facilitate open and honest dialogue where all voices are heard, ensuring that feedback is integrated into practice to advance equity effectively.
Educate Foundation Staff	Develop and implement a comprehensive educational framework on equitable evaluation principles. This involves engaging foundation staff through democratic processes such as workshops, seminars and collaborative discussions. Ensure that all staff members understand how equitable evaluation differs from traditional methods, and are prepared to support and engage in these practices from the outset.
Adapt the Role of Evaluation Commissioners	Redefine the responsibilities of evaluation commissioners to include assessing both the outcomes of grantees and the internal practices of the foundation itself. Provide training and support to help commissioners develop skills necessary for navigating and addressing internal evaluations. This shift promotes transparency and requires commissioners to engage in collaborative, reflective practices with stakeholders.
Prioritize Relationship Building	Implement democratic and collaborative strategies to build strong, productive relationships among evaluation teams, commissioners and grantees. Schedule regular, structured interactions, such as joint planning sessions and feedback meetings, to facilitate meaningful engagement. Ensure that relationship-building efforts are integrated throughout the project to create a supportive, collaborative environment.

<i>Implication</i>	<i>Description</i>
Develop a Clear Communication Strategy	Design a communication strategy that embodies democratic principles and fosters inclusivity. Define clear expectations for communication between evaluators, program officers and grantees, incorporating feedback from all stakeholders. Establish channels and protocols for regular updates and discussions to ensure transparency and effective coordination throughout the evaluation process.
Incorporate Feedback Loops	Create systematic feedback loops that enable democratic participation from all involved parties. Design mechanisms for ongoing input and dialogue, allowing evaluators, commissioners and grantees to share their perspectives and suggestions. Use this feedback to make iterative improvements to the evaluation process, ensuring it remains responsive and aligned with the needs of all stakeholders.
Ensure Equitable Resource Allocation	Apply democratic principles to the distribution of resources within the evaluation process. Engage all stakeholders in discussions about the fair allocation of resources such as funding, time and intellectual property. Ensure that resource distribution supports the equitable participation of all involved, particularly smaller or marginalized firms, to foster a collaborative and inclusive environment.
Promote Shared Decision-Making	Implement a collaborative decision-making process where all stakeholders have a voice in shaping the evaluation. Encourage democratic participation in setting evaluation agendas, defining questions and interpreting results. This approach ensures that decisions reflect diverse perspectives and promotes inclusivity and equity throughout the evaluation process.

Through the surfacing of these considerations and the use of a concrete example from a previous evaluation experience, we hope that evaluation commissioners and evaluators alike will gain some practical insights that will improve their own equitable evaluation practices.

References

- Argyris, Chris, and Donald Schön. 1974. *Theory in Practice: Increasing Professional Effectiveness*. San Francisco: Jossey-Bass.
- Botcheva, Lilia, Jessica Shih, and Lynne C. Huffman. 2009. "Emphasizing Cultural Competence in Evaluation: A Process-Oriented Approach." *American Journal of Evaluation* 30: 176–88.

- Chouinard, Jill A., and J. Bradley Cousins. 2007. "Culturally Competent Evaluation for Aboriginal Communities: A Review of the Empirical Literature." *Journal of Multidisciplinary Evaluation* 4: 40–57.
- Chouinard, Jill A., Tiffany L. S. Tovey and Kristin Kidd. 2023. "The Evaluation of Equity Focussed Community Coalitions: A Review of the Empirical Literature." *Journal of Multidisciplinary Evaluation*, 19 (45): 50–66.
- Frierson, Henry T., Stafford Hood, George Hughes, and Valerie G. Thomas. 2010. "A Guide to Conducting Culturally Responsive Evaluations." In *The 2010 User-Friendly Handbook for Project Evaluation*, edited by Joy Frechtling, 75–96. Arlington, VA: National Science Foundation.
- Greene, Jennifer C., Rodolfo A. Millet, and Rodney K. Hopson. 2004. "Evaluation as a Democratizing Practice." In *Foundations and Evaluation: Contexts and Practices for Effective Philanthropy*, edited by Marc Raverman, Naomi Constantine, and J. K. Slater, 96–118. Hoboken, NJ: Wiley.
- Hood, Stafford, Rodney Hopson, and Henry Frierson, eds. 2005. *The Role of Culture and Cultural Context: A Mandate for Inclusion, the Discovery of Truth, and Understanding in Evaluation Theory and Practice*. Greenwich, CT: Information Age.
- Hopson, Rodney K. 2009. "Reclaiming Knowledge at the Margins: Culturally Responsive Evaluation in the Current Evaluation Moment." In *The SAGE International Handbook of Educational Evaluation*, edited by Katherine E. Ryan and J. Bradley Cousins, 429–46. Thousand Oaks, CA: Sage.
- Hopson, Rodney K., Karen Lucas, and John Peterson. 2000. "HIV/AIDS Talk: Implications for Prevention Intervention and Evaluation." In *How and Why Language Matters in Evaluation*, edited by Rodney K. Hopson, 29–42. *New Directions for Evaluation*. San Francisco: Jossey-Bass.
- Kirkhart, Karen E. 2005. "Through a Cultural Lens: Reflections on Validity and Theory in Evaluation." In *The Role of Culture and Cultural Context: A Mandate for Inclusion, the Discovery of Truth, and Understanding in Evaluative Theory and Practice*, edited by Stafford Hood, Rodney Hopson, and Henry Frierson, 21–39. Greenwich, CT: Information Age.
- LaFrance, Joan. 2004. "Culturally Competent Evaluation in Indian Country." *New Directions for Evaluation* 2004 (102): 39–50.
- Manswell-Butty, Joanne, Melissa Daniel-Reid, and Valerie LaPoint. 2004. "A Culturally Responsive Evaluation Approach Applied to the Talent Development School-to-Career Intervention Program." *New Directions for Evaluation* 2004 (101): 37–49.
- Mertens, Donna M. 1999. "Inclusive Evaluation: Implications of Transformative Theory for Evaluation." *American Journal of Evaluation* 20: 1–14.
- Reese, Lissette, and Elizabeth Vera. 2007. "Culturally Relevant Prevention: The Scientific and Practical Considerations of Community-Based Programs." *The Counseling Psychologist* 35: 763–78.
- Schwandt, Thomas A., and Emily F. Gates. 2021. *Evaluating and Valuing in Social Research*. New York: Guilford Press.
- Smith, Nick L., and Sungmin Jang. 2002. "Increasing Cultural Sensitivity in Evaluation Practice: A South Korean Illustration." *Studies in Educational Evaluation* 28: 61–9.

- Thomas, Veronica G., and Frances I. Stevens, eds. 2004. "Co-constructing a Contextually Responsive Evaluation Framework: the Talent Development Model of School Reforms." In *New Directions for Evaluation* 2004 (101). San Francisco: Jossey-Bass.
- Thompson-Robinson, Marcella, Rodney Hopson, and Satish SenGupta, eds. 2004. *In Search of Cultural Competence in Evaluation. New Directions for Evaluation* 2004 (102).



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Part II

Experiences of commissioning in different institutional contexts



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

6 The influence of context for different types of commissioners of evaluation

Challenges and opportunities

Tessie Tzavaras Catsambas

Introduction

Evaluation professionals have long been interested in how to increase the use and influence of evaluation to advance the public good. When we think of evaluation professionals, we have traditionally thought about evaluators, but in fact, as this book has shown so far, evaluation commissioners have a great opportunity and hold a large share of the responsibility to make evaluation influential and to ensure it is used. Evaluation commissioners influence what gets evaluated, how that evaluation is structured, funded and staffed, who is involved, what is emphasized and the extent to which it is accessible, visible, credible and used. Evaluation commissioners are, therefore, in a great position to use evaluation to strengthen and advance democratic societies.

The ability of evaluation commissioners to influence the use of evaluation to contribute to democracy must be seen in the broader political and institutional context where that evaluation takes place – and there is, indeed, a wide variety of contexts each presenting its own, unique set of demographics, historical moment and forces at play that should be assessed and harnessed by the democratically-minded evaluation commissioner. This chapter aims to illuminate the opportunities and challenges commissioners face depending on the context in which they find themselves, and the strategies they use to promote democratic processes and outcomes within those contexts. The chapter also proposes a high-level framework that can be used by commissioners for considering the context of evaluation in relation to challenges and opportunities for promoting democracy. Finally, we consider a strategy that addresses immediate concerns and steers toward the medium-term objective of how best to manage individual evaluations, and the long-term objective of how to develop evaluation capacity in their institutions and communities to support the advancement of democracy more broadly.

The influence of context on the evaluation commissioner

The evaluation commissioner is not an external actor but is herself a product of the very context she is trying to influence. So, this chapter (and the book overall) essentially invites commissioners to step outside and look in as if they were external to

DOI: 10.4324/9781042000968-8

This chapter has been made available under a CC-BY-NC-ND license.

the context in order to become more intentional and strategic in how to act so they can exercise their influence in ways that boost democracy.

An important assumption in this discussion is that the evaluation commissioner is interested, motivated and confident to promote the public good by advancing democratic processes and strengthening democracy. Probably the first challenge for all commissioners of evaluation, but especially government commissioners, is acknowledging their role regarding the potential influence of an evaluation. For many evaluation commissioners, an assignment as “evaluation commissioner or manager” may be added onto their regular duties, and they see this “evaluation management” role as “moving the process along” – that is, managing the evaluation according to contract requirements especially within time and budget, ensuring an external evaluation team has access to the needed data sources, and avoiding any political missteps and errors. For example, in organizations where evaluation is oriented to performance assessment and linked to departmental budget allocations, evaluation commissioners’ job descriptions might be more about transparency, fiduciary responsibility and consistency across evaluations and less about representation, equity and sustainability. In this way, the evaluation context influences how evaluation commissioners perceive their assigned role and how they choose to step into that role in relation to promoting democratic processes. Thus, in developing strategies to strengthen democratic societies, we need to consider commissioners’ understanding of their role, and their risk tolerance, especially in cases where strengthening democracy involves deviating from assigned roles and job expectations and possibly putting at risk their career and livelihood. Therefore, a commissioner’s strategy for promoting democracy through evaluation needs to assess and mitigate risks both for the commissioner and for the evaluation.

Levels and aspects of context where evaluation is commissioned

“Context” is a broad term that can cover a wide range of variables at global, country and organizational levels. For the purposes of this chapter, Table 6.1 presents a high-level framework for considering the context in evaluation if our aim is to promote democracy through evaluation. The framework identifies aspects that can be used by the evaluation commissioner to justify, leverage and encourage the democratic process such as a focus on public good issues, safe and open dialogue, engagement of marginalized voices, considerations of environmental sustainability, increased capacity for uncensored scientific discourse and free press, etc.

We explore these commissioning context areas in three sectors:

- Government sector
- United Nations and international financial organizations
- Foundations and nonprofit organizations.

There are subtle but important differences across these sectors related to evaluation commissioners’ roles, as well as varied power and opportunity to exercise influence.

Table 6.1 Considerations of context elements across sectors for supporting democracy through evaluation

	<i>Government</i>	<i>United Nations</i>	<i>Foundations/Nonprofits</i>
Political system	Degree of democracy Lobbying pressures Legal framework	Diplomatic sensitivities on the executive board (made up of country representatives) Lobby interests	Degree of democracy of the organization's home country Maturity level of nonprofit sector
Organizational mission, vision, values	Relevant regulations What is prioritized?	Organizational focus Focus of government and NGO counterpart organizations in countries involved	Organizational focus Prioritized values
Current organizational priorities	Politically informed programmatic priorities at a point in time Ministerial/departmental evaluation capacity, e.g., in education, health, etc. Stakeholder public sector organizations	Current organizational strategic plan Harmonization with other UN agencies	Board priorities Current strategic plan
Placement of evaluation	Reporting and authority Links with potential evaluation users, access to information, potential for influence	Independent vs. internal Proximity to other organizational functions Drivers of the evaluation strategic plan	Proximity to leadership and other organizational functions Drivers of the evaluation strategic plan
Evaluation policy	Guidelines on priorities (e.g., learning, accountability, etc.) values and methods (e.g., preferred methods such as RCTs, requiring a gender or human rights lens, etc.)	Guidelines on priorities and process, values and methods, and evaluation use	Guidelines on priorities and process, values and methods, and evaluation use Assignment of roles and responsibilities related to evaluation (where evaluation policy is less developed)
Evaluation system	Capacity, resources, links with decision making, organizational culture related to evaluation		

Table 6.1 presents an overview of elements of context that should be considered at different levels of a system in different sectors and discussion follows.

Political system

Evaluation commissioners who work for government organizations are the most exposed to and influenced by that country's political, legal and public policy context. These commissioners may sit in government ministries or departments that hold responsibility for evaluation across the government sector, or they may sit in a government ministry or department that is explicitly tasked with evaluating programs and policies in the area overseen by that ministry or department – for example, health, education, agriculture, etc. To support democratic societies from their position, evaluation commissioners must consider pertinent government policies and legislation, the historical and current political environment, and any relevant political actors and powerful interests in the present time. In each of these areas, there may be elements that enable or hinder the use of evaluation toward democratic engagement and dialogue.

Government evaluation commissioners can exercise influence toward democracy both through specific evaluations (the way they commission, oversee evaluation implementation, and steer the dissemination of the evaluation report and briefings), and through the way they develop evaluation systems and build evaluation capacity in their ministries or departments.

In the two examples that follow, the same factors that may be perceived as weaknesses or challenges may also be used to advantage for the government evaluation commissioner who wishes to promote democracy through evaluation. This paradox really underscores the individual power and responsibility of an evaluation commissioner to map the context and think strategically about how to promote democracy from wherever they sit. The examples that follow do not provide a comprehensive description of the evaluation systems in the countries discussed but put forth selective elements to illustrate how one might analyze these contextual factors in light of the goal of promoting democracy.

Example 1. In 2019, the government of Brazil, established a new government entity to address several self-identified challenges in evaluation including limited availability of data related to several public policies, an underdeveloped culture of evaluation inside line ministries, and low evaluation capacity – human resources, alignment and training. The newly established Public Policy Monitoring and Evaluation Council (CMAP) was a new governing body of policy evaluation responsible for evaluating public policies. By 2022, many of the challenges persisted including lack of common definitions and shared understanding of monitoring and evaluation across the government (OECD, 2022). Therefore, an evaluation commissioner in one of the ministries in Brazil may face the disadvantage of little overall guidance and support in shaping the terms of reference of an evaluation. Furthermore, there may be resistance in the ministry to participate in evaluation, and even greater resistance if the evaluation calls for collaboration with other ministries. The easy way forward is for this evaluation commissioner

to take on a performative stance, push paper and stamp the evaluation “done”. Or, the pro-democracy way forward is to examine how to turn these obstacles into an opportunity using the lack of guidance as freedom to take initiative; for example, this commissioner could prioritize the Sustainable Development Goals, include in the terms of reference requirements for methods that catalyze collaboration, use personal relationships to gain consent for that collaboration, give voice to marginalized voices and equity issues in the evaluation, and then refer to the final report in design discussions for new programs (Mbava & Dahler-Larsen, 2019).

Example 2. In a different scenario, an evaluation commissioner in Chile may be working in a well-established monitoring system that aims to assess progress in program implementation against targets with performance indicators that include qualitative information on public goods and services. This evaluation commissioner also has to consider certain built-in incentives: when individual public servants meet targets in the grants they manage, they receive individual recognition in the staff performance system, and whole departments that meet annual targets are enabled to reward teamwork.¹ Perceived advantages for the evaluation commissioner in the Chilean system may include having clear understanding of goals and targets with consistent indicators across programs, as well as motivation for participating and collaborating in evaluation. Potential challenges for the commissioner who aims to promote democracy may be that established rewards for positive outcomes may skew the evaluation process toward certain findings, and thus possibly silence dissenting and critical voices, which would potentially compromise independence and impartiality in the evaluative process.

The Brazil example demonstrates how, amid limited guidance and weak evaluation culture, a commissioner might see lack of structure not as a barrier but as a space for innovation – incorporating SDG alignment, participatory methods, and equity lenses into evaluations. In contrast, the Chile example shows that a commissioner in a mature evaluation system must contend with incentives that may discourage critical findings. Even here, promoting deliberative dialogue and transparency is possible, though it requires vigilance against institutional pressures that compromise independence.

At the broadest level, government evaluation commissioners operate in political systems that fall somewhere along a continuum of “democracy” as defined by The Economist Democracy Index and “freedom” as defined by Freedom House in its annual Freedom in the World Report.

In addition, for some countries, external funding sources influence country priorities in different sectors. At international level, a commissioner working in a donor country’s agency must follow development priorities driven by political guidance in his place of work while balancing the priorities of the country where international development programs take place. Evaluation commissioners in international development constantly have to contend with this duality of external donor and country level priorities, and it is within this complexity that they need to chart their path to democracy.

At the organizational level (ministry or department) within a country’s government, evaluation commissioners need to consider the policies and capacity of

the ministry where they are working. They need to assess ministerial and sector policies as well as capacity needs to cover two separate areas: evaluation and democracy. For example, the education sector in a certain country may have high evaluation capacity (evaluations, teacher assessments, availability of reliable data, etc.), but challenges in the democracy front (curriculum censorship by the government, parent polarization, gun violence in schools).

Finally, evaluation commissioners need to remain alert to current changes that offer new opportunities for evaluation and democracy. For example, in 2015, countries signed a UN resolution to promote evaluation of progress toward the SDGs at country level, which offered an in-the-moment opportunity for evaluation commissioners in all countries to push for evaluation of progress toward the Global Goals. Another example is the opportunity that arose in 2019, when the US Congress adopted the Evidence Act, which gave an opportunity for evaluation commissioners to increase their budgets and elevate the profile of evaluation in their agencies in the United States.

To be clear, this mapping is not intended to identify barriers and give up, but to assess what is feasible, determine the effort influence and change will take, and develop the best path to achieve it. Let's also recognize that context mapping and strategy are only the first step, and then the real work begins.

Organizational context

Within organizations, the context may look different depending on where evaluation is placed in the organization, the organization's mission, values and strategy, and the evaluation policy and system in the organization. The evaluation function may be placed close to leadership or embedded in an office of programs or close to knowledge management and learning. The evaluation function may be controlled centrally or diffused in the organization. The placement of evaluation offers the evaluation commissioner different allies and channels to promote the use and influence of evaluation (Catsambas & Davidson, 2024).

In United Nations (UN) agencies and international financial institutions (IFIs) (e.g., the international banks), some evaluation offices are embedded in the organization and follow the direction of the executive, while others are "independent" and follow the direction of the Executive Board (made up of country members of the United Nations). All evaluation offices in UN organizations and IFIs have policies, budgets, and an annual plan informed by the strategic plan of the organization. Evaluation commissioners within this general structure have significant room to promote democratic processes through their evaluation work. One of the biggest obstacles to evaluation that serves democracy in UN organizations and IFIs has been the weight of bureaucracy and internal political complexities. In two large UN/IFI organizations, their evaluations have been independently assessed as difficult to understand, too long, and written with dense, academic prose; not surprisingly, there was evidence that these evaluations were not extensively used. Such inaccessible evaluations not only failed to promote democracy but were a misuse of public resources. The reason for this basic but critical failure might be due to the

pressure of strong politics in the organization, or a self-protective defensiveness on the part of the evaluation unit, but more investigation is needed to understand and improve that situation.

In contrast, there are UN/IFI evaluation offices that have furthered evaluative thinking and action by embedding methods that broadened the scope of evaluation to new underrepresented actors, invited open dialogue, and expanded capacity in their organizations to commission and use evaluation at country level. Here are some examples:

- UN leadership (UNICEF's Evaluation Office) invested along with the Government of Finland in a global network of evaluation organizations to promote the inclusion of civil society in country-level evaluation (Catsambas, 2015).
- Over the last 10 years, the World Food Programme's Office of Evaluation has continued to build evaluation capacity throughout the organization, and in the last 5 years, it included in its capacity building efforts attention to gender responsive evaluation practices in its country evaluations (World Food Programme, 2023).
- In 2023, UNFPA's Independent Evaluation Office conducted a formative evaluation of UNFPA support to adolescents and youth that was led and implemented by young and emerging evaluators and featured significant youth engagement throughout the evaluation process (UNFPA, 2023).

These cases demonstrate how commissioners can expand democratic participation, shift power dynamics and elevate underrepresented perspectives. Thus, such initiatives by UN/IFI evaluation commissioners have made tangible contributions toward greater civic participation including civil society actors, youth and women, and featured local leadership shifting the balance of power and control over the evaluation from the Global North to the Global South.

In foundations and nonprofit organizations, there is less emphasis on the evaluation policy and system, but more emphasis on understanding the priorities of the current leadership. Change in foundations is more frequent, and it is not unheard of to see an overhaul of a foundation's entire portfolio. For example, in 2023, compelled to leave Hungary by the Hungarian Government, the Soros Foundation announced withdrawal and termination of large parts of its work within the European Union with significant impact in the human rights nonprofit sector especially in Hungary; in 2019, the Salesforce Foundation was folded into the Salesforce for-profit organization closing down its evaluation office; in the 2015, the Ford Foundation channeled significant resources in a new grant making program to fight poverty. In another foundation, within a few years, the evaluation office moved from the President's Office to five steps removed. Therefore, when mapping the context in a foundation, the most important variables are mission, vision, values, and current strategy and priorities. To be influential, the evaluation commissioner will need to track closely the needs and priorities of the Foundation's leadership, and the goals of the Foundation's Board of Directors.

Building trust, anticipating change and identifying allies are key strategies for the commissioner with democratic aims.

Addressing real-world complexities

So far, we have looked at a relatively static map and list of questions to help us map the context evaluation commissioners face. In practice, many complications arise from inconsistencies between international and country priorities, country and organizational policies and priorities, as well as the complexity of fragile contexts and conflict settings. An evaluation commissioner building a context map must note these issues and the forces at play and develop thoughtful and flexible approaches to address them.

What happens when elements of context in a country are in direct opposition to those in an organization working in that country? For example, international and bilateral partners working in Uganda prioritize human rights and gender equality to varying degrees. The Anti-Homosexuality Act of 2023 of Uganda considers the promotion of homosexuality punishable by death. This Ugandan law violates multiple fundamental rights guaranteed under Uganda's own constitution and breaks commitments made by the government as a signatory to several international human rights agreements. How should a commissioner of evaluation promote democratic processes within that context? To complicate this example further, let's consider that a funder is currently supporting and promoting a localization strategy that calls for a significant percentage of its funds to be channeled to local organizations, and for government priorities to guide development assistance. How should an evaluation commissioner working in that donor organization address this conundrum? Should the donor evaluation commission prioritize human rights or localization and alignment with this anti-human-rights law? In such cases, a more detailed mapping of potential partners and actors is needed to determine the best path forward. The pro-democracy commissioner can initiate and stimulate critical reflection through consultations with colleagues and allies of democracy, and support agreements for coordinated pro-democracy programming and action.

Another situation that deserves mention is evaluation in fragile contexts and conflict settings, where promoting democracy amid hostilities is complex and risky (Catsambas & Aslam, 2024). In a personal example of an evaluation in a post-conflict country that had experienced civil war across ethnic and religious differences, the evaluators decided to conduct focus groups of women across these ethnic divisions. The evaluation commissioner consented because of the shared goal to use the evaluation to further the democratic process in that country. The focus group was peacefully completed with what seemed to be more exploration of common ground across warring sides; but it could have backfired. In fragile or conflict settings, evaluation commissioners need to balance their eagerness to promote democracy with their consideration for harm for the individuals involved. That challenging context increases the risk of good intentions not leading to the desired positive results.

Conclusion

For evaluation commissioners who wish to promote democratic processes and systems through evaluation, understanding and considering their context is an important step. A mapping of the context related to democracy offers an opportunity to step back and reflect, assess risks and opportunities, as well as identify levers of change.

Understanding context, however, is only the first step. Commissioners must translate that understanding into strategy – deciding where and how to act, whom to engage, and what values to uphold. This includes:

- Increasing the visibility and adoption of priorities in what to evaluate and methods of evaluation that promote the democratic process
- Strengthening partnership with other organizations to expand impact in democratizing societies
- An evaluation capacity-building plan for their evaluation units and their organizations for increasing their capacity to be an effective contributor to democracy.

Strategies must also account for personal and professional risks, as well as potential harm to those involved in evaluation. Commissioners often navigate political minefields; courage and prudence must coexist. Considerations of ethics and do-no-harm may serve to temper bold, risk-taking action and moderate the pace of intended change. When they put their mind to it, evaluation commissioners can find ways to sidestep political traps, work through alliances, extend their informal influence, and act incrementally, and succeed in fostering change while safeguarding their roles.

Evaluation commissioners wield more power than they often realize. When they choose to engage proactively and strategically, they can make meaningful contributions to democratic systems. By mapping context, assessing risk and planning thoughtful actions, commissioners can use evaluation to enhance civic participation, amplify marginalized voices and promote transparency.

Through deliberate engagement, even constrained contexts can become sites of democratic innovation. The path may be complex – but it is a path that commissioners are uniquely positioned to walk.

Note

- 1 Irarrazaval and Ríos (2014) and OECD (2021).

References

- Catsambas, Tessie Tzavaras. 2015. “Creating a Global Movement in Evaluation: The Story of EvalPartners.” *American Journal of Evaluation* 36 (2): 256–269. <https://doi.org/10.1177/1098214014567140>.

- Catsambas, Tessie Tzavaras, and Ghazia Aslam. 2024. "Building Evaluation Capacity." In *Research Handbook on Program Evaluation*, edited by Hallie Preskill and Darlene F. Russ. Cheltenham, UK: Edward Elgar Publishing.
- Catsambas, Tessie Tzavaras, and E. Jane Davidson. 2024. *Evaluation Management: How to Commission and Conduct Evaluations that Matter*. Thousand Oaks, CA: Sage Publications.
- Irrarrazaval, I. and Belen Ríos (2014), "Monitoreo y Evaluación de políticas Públicas", <https://doi.org/10.13140/2.1.4072.2886>
- Mbava, Nombeko P., and Peter Dahler-Larsen. 2019. "Evaluation in African Contexts: The Promises of Participatory Approaches in Theory-Based Evaluations." *African Evaluation Journal* 7 (1): Article a383. <https://doi.org/10.4102/aej.v7i1.383>.
- Organisation for Economic Co-operation and Development OECD. 2021, *Monitoring and Evaluating the Strategic Plan of Nuevo León 2015–2030: Using Evidence to Achieve Sustainable Development*, <https://dx.doi.org/10.1787/8ba79961-en>.
- Organisation for Economic Co-operation and Development (OECD). 2022. *Centre of Government Review of Brazil: Toward an Integrated and Structured Centre of Government*. Paris: OECD Publishing. <https://doi.org/10.1787/33d996b2-en>.
- UNFPA. 2023. *Formative Evaluation of UNFPA Support to Adolescents and Youth*. www.unfpa.org/sites/default/files/2023-05/Adolescents_Youth_Evaluation_Report.pdf.
- World Food Programme. 2023. *Annual Evaluation Report (2023)*. https://docs.wfp.org/api/documents/WFP-0000158757/download/?_ga=2.15346740.305154372.1732574552-924385381.1730238377.

7 Democracy, SDGs, evaluation and commissioners

Connecting the dots

Robert Lahey and Dorothy Lucks

Introduction

The 2030 Agenda – *Transforming our World* (United Nations, 2015) and the 17 Sustainable Development Goals (SDGs) – is perhaps the most ambitious global change exercise ever undertaken for all countries to take the lead on change that would address key global issues and accelerate progress on sustainable development. The process of developing the SDGs was highly participatory, and 193 countries ratified the 2030 Agenda, an immense effort in global democracy. The underlying values of democracy are incorporated into the 2030 Agenda, such as: freedom of speech (or expression); tolerance for diversity and differing opinions; respect for individuals in a decision-making process and their choices; equity – ensuring all those affected by a decision have an equal say in the decision-making process; justice – guides and protections to confirm safety in the decision-making process for the most vulnerable.

Yet, the combined impact of unforeseen challenges that SDG implementation has experienced during the first half of Agenda 2030 – the COVID-19 pandemic, climate disasters, conflict and economic downturn – is causing a lag in progress and requiring world leaders to reflect on the way forward for the SDGs. (SDG Knowledge Hub, 2020; Martin-Blanco et al., 2022; UNDESA, 2021). The current global context creates additional uncertainties and a sense of urgency to find solutions that are integrated, innovative, implementable and impactful.

Individual and systematic evaluations have value in helping to build understanding of complex issues, of engaging diverse stakeholders in important dialogue on critical issues. Evaluations provide the opportunity to explore complexities, seek multifaceted solutions to complex issues, and to better understand trade-offs across “competing” SDGs; for example, a country could face a critical balance between SDG 9 (industry, innovation and infrastructure) and SDG 13 (climate action). Evaluation practice raises the importance of engaging with multiple and diverse organizations and peoples in a democratic way that connects them, their issues and knowledge, and delivers pathways.

At present, those working in the field of SDG implementation and those working in evaluation are playing a limited role in support of SDG learning and acceleration toward achieving the SDGs. Both leaders who currently influence

DOI: 10.4324/9781042000968-9

This chapter has been made available under a CC-BY-NC-ND license.

the commissioning of evaluations, and leaders that have potential to influence commissioning of evaluation (potential commissioners), as well as evaluators could contribute more strongly to SDG-related support and decision-making so that decision-makers connect the lessons from these evaluations toward better SDG outcomes. The UN General Assembly Resolution 71/313, 2023 on country-led evaluation provides an impetus toward high-quality evaluations with greater impact for the SDGs.

This chapter explores the characteristics of decision-makers that influence and have the power to commission evaluations in the context of the SDGs. Commissioning of evaluation requires a sequence of decision-making of (i) the need for an evaluation, (ii) the scope of an evaluation, (iii) management and resourcing for the evaluation and (iv) the finalization and use of evaluation findings and recommendations. This commissioning process, particularly in complex SDG-related evaluations, involves multiple decision-makers at different levels of responsibility. For this reason, this paper uses a broad definition for an evaluation commissioner – any person that has influence over how an evaluation is identified, procured, managed, implemented and used.

Consequently, it is important to understand how SDG evaluations are commissioned, why and by whom, and how the demand for evaluation is stimulated through the roles of both commissioners and evaluators. With clearer understanding of the dynamics of decision-making on which evaluations are commissioned, commissioners and evaluation practitioners can play a more meaningful role for SDG achievement.

The context for SDG implementation and evaluation is becoming more complex

The implementation of the SDGs faces increasing complexities as the global, national, and sub-national contexts evolve. It is past the halfway mark for the 2030 Agenda and progress is lagging. There are new conflicts and imperatives emerging. Economic fragility, climate crises and polarization are increasing. There are new disparities; and a recognition of the need for greater urgency for the SDG principle of “Leave no-one behind”. A new approach to addressing the SDGs is needed. Those promoting evaluation have tended to emphasize building the supply side of evaluation expertise, but SDG evaluation also needs to focus on the demand side of the evaluation market.

There are economic, social and environmental challenges of addressing some SDGs and not others. The consequences of roll-out of a single SDG are magnified with the complex relationships across 17 SDGs. All have political consequences and impact social, cultural, economic and environmental imperatives in different ways. Some SDGs involve national and multi-country and global commitments that extend beyond national borders, requiring leaders to balance local, national and global priorities. Considering the full set of 17 SDGs represents a complicated situation for decision-makers, and one where evaluation could help. But SDG evaluations are complex, requiring work across organizational boundaries, for

decision-makers with varying objectives and agendas, and potentially drawing from a diverse resource base. Consequently, to have influence on SDG progress, the contribution of evaluation needs to be well regarded/respected by country leadership as well as at operational levels, otherwise, it will not have political space/permission or resources to operate effectively (IIED 2020).

Levels of SDG roll-out and evaluation commissioning

SDG implementation at global level extends beyond country perspectives into multi-country approaches and multilateral organizations. At the national level, SDG implementation considers macro and micro lenses, with different consequences and decisions that may need to be taken at each level. This paper focuses on the national level, given the 2030 Agenda attention to country-led evaluations.

Figure 7.1 provides a representation of how an SDG strategy/plan is aligned with a country's national goals and priorities (level 1, macro level), and eventually translated into programs, projects and activities carried out by diverse stakeholders within the country who have varying degrees of power to influence decisions on whether to conduct evaluations, and different levels of knowledge of how those evaluations might be framed (levels 2 toward micro level 5). Country-level SDG evaluations need to recognize and respect these different levels, each representing a unique set of questions that would need to be answered, and a different evaluation commissioner (or set of commissioners) to deal with.

Even where a country prioritizes only one or two SDGs within their national planning framework, there are decision-makers at various levels. For instance, in Nepal, the Monitoring and Evaluation (M&E) Act 2024 provides a national

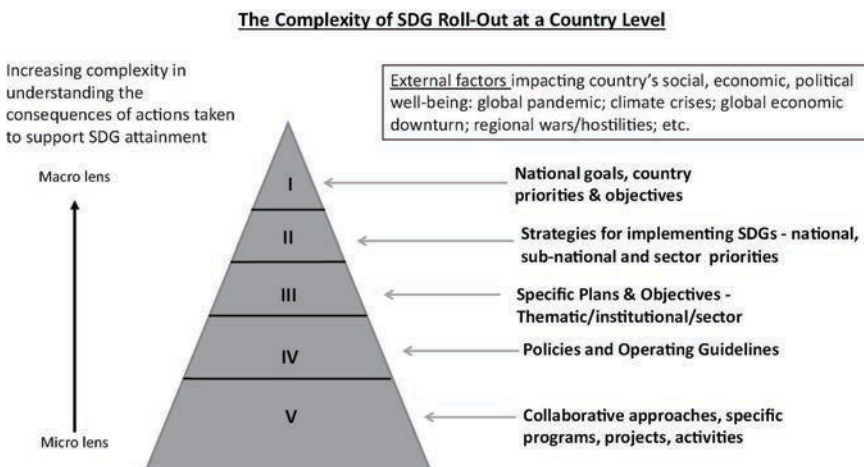


Figure 7.1 The complexity of SDG roll-out at a country level.

Source: Authors.

framework for evaluation, and three high-level national committees guide SDG implementation through a wide range of programs and projects. Alternatively, in India and Australia, evaluation and SDG activities are largely decentralized to each State government. The locus and responsibilities of decision-making are therefore critical considerations in any SDG evaluation.

For evaluation to play a meaningful role in support of the SDGs, in addition to the technical requirements of gathering, analyzing and interpreting data and information, more attention needs to be paid to those who have the power to commission evaluation and to how the flow of information, capability and resources for evaluation at national and sub-national levels can be strengthened. Evaluators must produce timely, valuable work to support good governance and sustainable development. Evaluators can use their skill set as facilitators and analysts, and work with others to help “unpack” complex situations posed by SDG trade-offs to arrive at valid and considered advice to leadership.

Evaluation relies on allocation of resources and a specific commission to conduct an evaluation. Different types of individuals or multi-stakeholder groups play a role in commissioning evaluations. The notion of an evaluation “commissioner” is an ambiguous term and can be interpreted differently. A politician can demand an evaluation, but a specific Ministry will define and manage the actual evaluation. There may also be important influencers at any level that champion the identification and commissioning of evaluations; for instance, national SDG advisors to the national leadership and an SDG Steering Committee may advise on the monitoring and evaluation of the SDGs that consider the interrelationships between SDGs.

Such influencers may be within a country or may extend to external people with national influence such as international figures, thematic experts or international organizations. For instance, during the COVID-19 pandemic, international development agencies such as the World Bank, the World Health Organization and the Food and Agriculture Organization worked with national government ministries (planning and economy, health and food systems, respectively) to evaluate the real-time effects of the pandemic on their specific COVID-related investments and how these were affecting wider SDG progress.

Figure 7.2 represents “commissioners” as those who demand an evaluation as they seek information to help in their decision-making. These are the “users” of evaluation results. It integrates two elements: the “level of authority” held by an official; and their “span of control” to influence the commissioning of an evaluation. Across each level, commissioners (and those with authority and potential to commission evaluations) may have varying roles and responsibilities in line with the macro-micro continuum, shown in Figure 7.1.

Levels I-III are more likely to take a role in influence, decision-making and high-level planning for an evaluation, approving and guiding the aim, focus and use of an evaluation. Levels IV and V are more likely to be engaged in detailed planning, procurement, management and dissemination of evaluations. All are important stakeholders in commissioning of evaluations, but have different needs to be addressed.

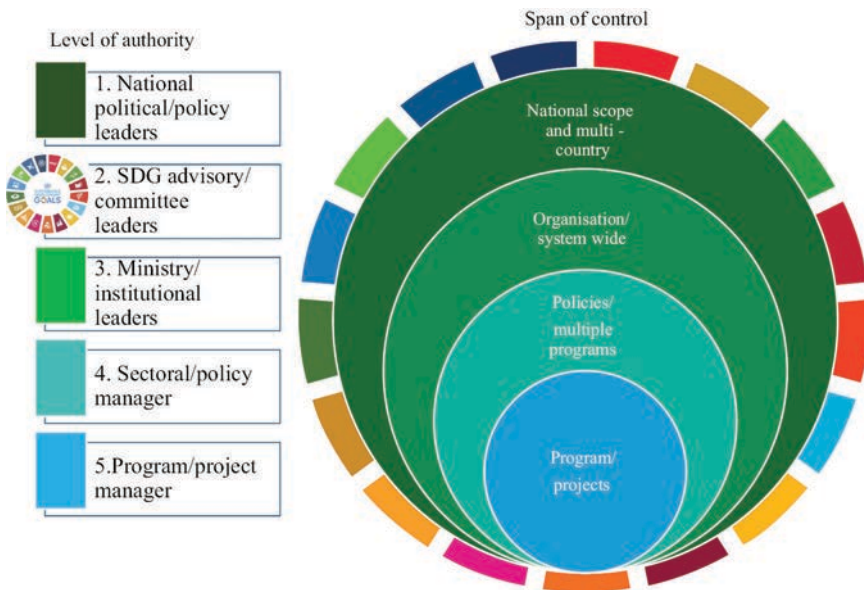


Figure 7.2 Commissioners of evaluation – level of authority and span of control.
Source: Authors.

Based on the levels in Figure 7.2, a rubric portraying different types of commissioners and their respective roles in SDG implementation and evaluation (as per Figures 7.1 and 7.2) is shown in Table 7.1. This is not a linear and bounded set of definitions but is presented as a guide.

Based on the analysis of Table 7.1, if evaluation professionals wish to stimulate demand among potential commissioners of evaluation, evaluators must build stronger relationships with the different types of commissioners. Similarly, commissioners need to consider the types of questions that need to be answered; influencing the scope of the proposed evaluation, and the profile and capacity of evaluators required. The relationship between commissioners and evaluators continues through the evaluation process and the feedback of evaluation findings and recommendations into decision-making processes.

Building demand for SDG evaluations

In the ideal, systematic evaluation is institutionalized and integrated into a country's decision-making. For this to happen, there is a need for consideration to be given to investing in building evaluation capacity on both the supply side, as well as the demand/"user" side (the latter of which would include political and government leaders, senior and mid-level bureaucrats, key strategists and planners, amongst others). An early example of this was the introduction of systematic evaluation in

Table 7.1 Potential evaluation commissioners/influencers/users – country level

<i>Commissioners</i>	<i>Roles</i>	<i>Power, responsibilities and focus for commissioning evaluation</i>
1 Senior country leadership & their chief advisors Office of the President; PMO; PCO	National goals, Country Priorities and Objectives	• Identify need for information. • Overall progress – SDGs and SDG strategy • Trade-offs and potential consequences • Broad policy options re SDG implementation • “Early warning” or analysis of specific target groups or critical areas of concern
2 Ministers/Agency heads & Advisors of National Oversight Business & civil society network leaders Key SDG committee(s)	Strategy for Implementing SDGs and Country Priorities	• Influence topics for evaluation • Specific SDGs and/or SDG nexus – effectiveness, efficiency, unintended effects, need for change, etc. • Review priorities against changing contexts • The mechanics of SDG roll-out
3 Ministers & Agency heads (Principal Secretaries, Deputy Ministers, etc.) of all relevant government ministries & agencies Heads of all relevant parastatals and similar type organizations Relevant sector business & NGO network leaders	Sector-Ministry Plans and Objectives	• Identify need and manage evaluation within relevant sector • Continued relevance of sector-specific priorities viz-a-viz country priorities • Potential trade-offs/colliding priorities across SDGs – deeper analysis • Sector-specific programs – deeper analysis of progress, performance, need for change, etc.
4 Senior officials and their key policy analysts responsible for policy within their ministry, agency, parastatal and similar type organizations Heads of relevant/sector business & NGO networks	Policies within each Sector- Ministry	• As above – particular reference to relevant policy area(s) within the ministry/agency • Assessment of policy areas – continued relevance; their contribution to specific SDG progress • Potential trade-offs/colliding priorities across ministry/sector policy areas or other SDG initiatives • Policy-specific programs – progress against targets, various aspects of performance, need for change
5 Program & Project Managers within ministries, agencies, parastatals and similar type organizations	Programs, Projects and Activities within each Sector- Ministry	• As above – particular reference to relevant programs within the policy area of the ministry/agency • Process and implementation of programs • Program performance – contribution to SDG progress • Program impacts and effects, intended and unintended; reach to various audiences • Alternatives to current programs/projects

Source: Authors.

Canada in the 1980s; and then a re-building of the evaluation system in government in the early 2000s (Lahey, 2023; Gauthier, 2022).

The reality is that, in most countries, current and potential commissioners of evaluation have been silent insofar as the SDGs are concerned, in part due to insufficient country-owned capacity and/or lack of understanding of the value, scope, methods and potential uses of evaluations for different contexts (Van den Berg et al., 2017).

A mindset change is required for more attention to be paid to the *demand* side of the market for evaluation, that is, those who have the power to commission evaluation. This is true not only of all countries, but also of development partners (UNEG, 2012). Change of course needs strong political support from the top and can be greatly aided by “champions” or leaders who already have an understanding and knowledge of evaluation and its potential for SDG achievement.

There is a potential to change this through raising the profile of what evaluation can offer and putting more focus on “capacity building” on the demand side of the evaluation sector. The intent would be to engage more directly with potential commissioners to better position evaluation through raising awareness of the varied roles, uses and utility of evaluation, and ultimately influencing the preparation and conduct of SDG-related evaluations. A key entry point to country-led evaluations must start with country leadership, the highest level of national SDG evaluation commissioner.

Commissioners of evaluation at each level need to know that evaluation can and should be used and have the capacity (both human resources and budget) to be able to benefit from SDG evaluation work in a timely way. Commissioners also need to be aware of the wider SDG vision and objectives as applied in their context, not simply reflecting on a specific SDG goal and target within their sphere of influence but considering wider synergies and potential trade-offs. Project, program, policy or sector evaluations may feed into broader, strategic evaluations with a macro set of issues to explore.

It is important to examine the cascading impacts and effects (both intended and unintended) of interrelated SDGs, and what this might imply for policy decisions and priority setting. This would have implications for scoping, timing, resourcing, management and delivery of evaluations across agencies and within the larger system.

Ultimately, this represents a way of working, both for the commissioners as well as the evaluators in a country, that focusses on issues-oriented problem-solving, requiring evaluative research and collaboration across existing institutional boundaries. This places more demand on the “learning” role of evaluation, not only an “accountability” tool to measure shorter-term outcomes (Van den Berg et al., 2021). Additionally, this could see novel types of transdisciplinarity that would combine specialists and stakeholder expertise working within a team environment that included evaluation/research/science-policy interface (Biermann et al., 2017).

Capacity for SDG evaluation – commissioners and evaluators

Country reporting on SDG progress via Voluntary National Reviews (VNRs) has demonstrated country-level capacity gaps with both monitoring and evaluation. While systematic SDG evaluation benefits from performance monitoring, it needs enabling factors supportive of a “mature” national M&E system, – transparency, freedom of information, a free media, a vibrant, engaged civil society, among others (Lahey, 2016). To be resourced and institutionalized, evaluation also needs to be understood and well regarded by country leadership.

To date, for various reasons, monitoring and evaluation of SDG progress have generally been insufficient. The Inter-Agency and Expert Group on SDG Indicators has proposed revisions and additional indicators for the 2025 Comprehensive Review in accordance with UNGA Resolution 71/313. Yet, national datasets for all SDG indicators and targets are rarely complete; and many countries have substantial data gaps due to lack of capacity or resources for monitoring. Lack of systematic SDG evaluation may reflect country capacity gaps and/or limited understanding or acknowledgement of evaluation’s utility. Consequently, the VNRs lack analytical depth, clear evidence of progress/challenges or policy focus (UNDESA 2023).

Where data is available, annual SDG Progress Reports and analyses by the Committee for Development Policy and the Global Sustainable Development Report indicate that the incremental changes achieved are inadequate for meeting expected SDG outcomes.

There is a political element to SDG decision-making to balance and synergize solutions across multiple priorities. Accordingly, evidence-based and critical thinking is essential in dealing with the complexities and trade-offs of the SDGs – and evaluation can create the conditions for that critical thinking to emerge as it helps in transforming SDG progress. Evaluations are not of themselves “transformative” in terms of contributing to SDG outcomes but, evaluators, working in collaboration with other key players, and assuming a role as “facilitator”, can be very powerful in assisting to accelerate progress on the SDGs. Evaluations enable a focus on specific issues or areas of work that can accelerate progress for specific or multiple SDGs.

In sum, building additional evaluation capacity needed to support country-led evaluation of the SDGs is especially critical (Lahey and Lucks, 2023). Some country examples where political leaders were closely associated with the institutionalization of systematic tools to measure performance and results include the United States under Presidents Clinton (1993) and Obama (2010) (Kamark, 2013); the United Kingdom under Tony Blair (2001) (Clement, 2022; Barber and Blair, 2015); and Canada under Prime Minister Trudeau (2016) (May, 2019). It reinforces the importance of “capacity building” on both the demand and the supply sides, so that “*potential*” commissioners do in fact eventually become *actual* commissioners of evaluation in practice – and recognize and set in motion the various ways that systematic evaluation can and should be used.

Evaluation capacity-building efforts are needed and important but have not proven sufficient to ensure systematic SDG evaluation, and not with the urgency

that the current global situation warrants. Lessons from the COVID crisis show that a sense of urgency is created when leadership deems it a priority. For that sense of urgency to be sustained, direction on the priority needs to come from the highest levels of leadership (Level 1 in Figures 7.1 and 7.2).

Globally, there are organizations and efforts that are helping countries determine how best to institutionalize/integrate systematic evaluation into the decision-making process and build evaluation capacity. These include the initiatives of the Global Evaluation Initiative, the work of the national Voluntary Organizations for Professional Evaluation, supported through the International Organization for Cooperation in Evaluation and EvalPartners, including EVALSDGs, EvalNet, EvalforEarth, UNEG amongst others.

These initiatives provide a bridge between evaluation professionals and national governments, albeit through different relationships and pathways. Where there is limited evaluation capacity at the national level, the efforts of UN Country Teams in supporting countries through the development of UN Sustainable Development Cooperation Frameworks (UNSDCF) provide a useful vehicle in helping identify where evaluation would be best positioned to support SDGs implementation and integration of evaluation into the country's planning, priority-setting and government policy/program delivery.

Enabling factors to accelerate country-led, SDG evaluations

The potential contribution of evaluation to decision-making is often overlooked or not even considered. Commissioners and potential commissioners often do have a seat at the table of critical SDG-related discussions and can reach and inform decision-makers of the value of evaluation. Each could raise the potential for SDG evaluations within their area of influence and span of control.

The evaluation skill set provides a systematic approach to bring together relevant stakeholders to work toward solutions to complex problems. A key and valuable role for the evaluator is as a “facilitator” to work cooperatively with areas of expertise needed to both “unpack” the complexity of addressing multiple SDG goals at a country level, and to also find workable solutions for the country, that is, solutions that would be realistic, practical, effective and sustainable.

Developing “workable solutions” to complex problems requires broader input than professional evaluators would typically provide. Increasingly, evaluations combine specific skill sets with multiple technical specialists. It is important for evaluators to add skills and ways of working with other professionals such as engineers, political scientists, economists, sociologists, public administration specialists, digital and social media specialists, amongst others. Lessons learned from the pandemic are that faster, more responsive data capture can occur, and needs to occur when needed by decision-makers. Evaluators must understand that “good enough” multi-stakeholder, multisector analysis may be required in particular situations (and can be as valuable as long-term research, although that is also needed).

Bringing decision-makers and evaluators together at a senior level can significantly accelerate critical thinking and decision-making. The examples of the SDG

evaluations that have been carried out to date provide a good example of how for example Finland and Nigeria took time to identify the most critical and urgent issues for attention as part of the evaluation process. Action was taken with the approval and support from the most senior level of commissioner in the country. Policymakers were collaboratively engaged in debate, creating ownership in the resultant evaluation recommendations.

Other enabling factors for SDG-related evaluation include transparency; accountability, access to information, and an ability and willingness of leadership to consult widely to gain an understanding of the perspective of the various population toward critical issues being raised. It requires a viable and free media that would be able to contribute to open public disclosure and debate; and, a vibrant civil society, where diverse voices can be heard. An environment that allows for public sector reform is needed. Such factors underlying an effective national evaluation system that support SDG evaluation resonate with and reinforce democratic ideals, governance and practice.

There is a range of skills that need to be built for commissioners to become more aware of the value and range of evaluation approaches and uses, what their power, role and potential line of action can be and how these can be applied in the context of the SDGs or complex issues. Advances in evaluation methods include complexity-aware, participatory and environmentally conscious approaches, along with the use of AI and digital technologies. The skills of evaluators need to be extended to align with the needs of commissioners, with the ability to manage skilled teams that can deliver complex and transdisciplinary evaluations, in a timely and efficient way.

Public debate often becomes polarized, pulling a decision-maker toward particular positions in complex situations, with insufficient formal analysis. This polarization can be exacerbated by media, increasingly, social media, where focus is placed on ideal or hoped-for end states (but with too little discussion of how to get there in specific contexts). Evaluators enable analysis that brings together scientific, political and other issues together to generate evidence-based balance in the decision-making process.

The way forward

Addressing complexity. The complexity of SDG interaction and the increased urgency to deal with “colliding priorities” create challenges and opportunities for evaluation. Ongoing SDG monitoring does not provide sufficient information to address complex issues without drilling down to capture information that sheds light on the implications of various courses of action that might be taken in support of SDG progress.

Stronger collaboration between national leadership and evaluation leaders. Buy-in is needed by country leadership so that systematic evaluation can deliver potential solutions for SDG quandaries. This involves a rethink of the role and stronger influences on the uses of evaluation to improve critical decision-making

toward SDGs impact. Specific action is required by country leadership at all levels, to build the “demand” side of the market for evaluation.

Wider and integrated partnership processes. Cooperation, partnering and time commitment for horizontal cross-sectoral studies facilitate partnering of evaluation with research, social and physical sciences and policy areas. Delivering results of systematic enquiries in a timely fashion should be drawn from lessons from the COVID-19 pandemic about the importance of “real-time” delivery. A need for cross-sectoral coordination, leadership, resourcing and commitment to a plan related to the SDGs should be mirrored in SDG evaluations. This would require bringing together potential “commissioners” of evaluation in support of effective coordination and timely delivery.

More attention to the resources required for evaluation. Due to the horizontal nature of SDG evaluation, they can be costlier, more difficult to manage and not timely. Country leadership needs to allow time and resources for the conduct of cross-sectoral evaluation studies, but the scheduling, tabling and public disclosure of results require knowledge and commitment of evaluation leaders to work closely for timely results. Country-led evaluation and an effective NMES can occur with greater efforts at evaluation capacity building and institutionalization of evaluation.

A stronger and consistent narrative on the importance of SDG evaluation by development partners at all levels. Re-energizing SDG progress over the second half of Agenda 2030 needs to address the importance of country leadership to engage and support systematic evaluation of the SDGs as a critical element of their national SDG strategy. The role of international, national and sub-national private sector actors and civil society can also play an important advocacy and practical role.

A new paradigm for capacity development and the SDGs. Seven critical areas stand out requiring more emphasis in the international dialogue and capacity development for the SDGs to harness the democratic power of leaders to identify, influence and conduct evaluations that will contribute critical lessons for SDG achievements:

- 1 Increasing understanding by national decision-makers of the roles and uses of evaluation in complex contexts, and commitment of time and resources to scoping and financing SDG evaluations.
- 2 Positioning evaluation as solution-oriented and timely with country leadership, especially potential commissioners.
- 3 Maturing of national systems from a reliance on monitoring and reporting on few indicators to a more comprehensive system where evaluation serves as a complementary tool to provide a deeper understanding of SDG performance, including the complexities of SDG trade-offs.
- 4 Building the capacity of evaluation professionals on supporting and promoting the commissioning process for evaluations, including the range of methods and tools available for effective SDG evaluations, and the benefits that such investments in evaluation can generate.

- 5 Revisiting methods for evaluation management, to become nimbler, providing timely delivery of appropriate and useful products, promoting evaluation's power lies in contribution of analytical capability, critical thinking and facilitation tools and skills to positive decision-making.
- 6 Strengthening communication skills and links between commissioners and evaluators and technical specialists and professionals/experts in other fields to arrive at solutions/recommendations that are practical and address the key SDG issues and challenges. This requires more opportunity for commissioners and evaluators to meet for scoping dialogue on where evaluation can be applied to help accelerate SDG progress.
- 7 Addressing the complexity surrounding the SDG issues and their solutions through increased collaboration between commissioners and evaluators on developing a forward-looking vision, and an ability to spur innovative thinking in SDG evaluations to determine the best way forward.

As progress is made on the increased use of systematic evaluation by decision-makers, a country moves closer to improved governance and democratic ideals. The same could be said about the SDGs. Evaluation can play a useful role in terms of facilitating SDG progress, particularly in this more complex, urgent world. Transitioning a country's national M&E system from the simple approach of monitoring and reporting to a deeper (and more complex) system of working through SDG trade-offs at a country/operational level provides vast opportunity for improvements on SDG results and democracy across the globe.

References

- Barber, Michael, and Tony Blair. 2015. *How to Run a Government*. London: The Strand Group, King's College London. June 11.
- Biermann, Frank, Norichika Kanie, and Rakhyun E. Kim. 2017. "Global Governance by Goal-Setting: The Novel Approach of the UN Sustainable Development Goals." *Current Opinion in Environmental Sustainability* 26–27: 26–31.
- Clement, Michelle. 2022. "The Art of Delivery: The Prime Minister's Delivery Unit 2001–2005." *History of Government Blog*, August 26.
- Gauthier, Benoit, Robert Lahey, and Steve Jacob. 2022. "Institutionalization of Evaluation in Canada." In *The Institutionalization of Evaluation in the Americas*, edited by Reinhard Stockmann, Wolfgang Meyer, and Laszlo Szentmarjai, 143–70. Cham, Switzerland: Palgrave.
- International Institute for Environment and Development (IIED). 2020. *Evaluation to Connect National Priorities with the SDGs: A Guide for Evaluation Commissioners and Managers*. London: IIED.
- Kamarck, Elaine. 2013. *Lessons for the Future of Government Reform*. Testimony before the U.S. House Committee on Oversight and Government Reform, Washington, DC, June 18.
- Lahey, Robert. 2023. "John Mayne and the Origins of Evaluation in the Public Sector in Canada: A Shaping of both Evaluation and the Evaluator." *Canadian Journal of Program Evaluation* 37 (3): 340–54.

- Lahey, Robert. 2016. "Why Developing Monitoring and Evaluation Capacity Is Critical to Understanding and Addressing Issues of Poverty and Inequality." In *Poverty, Inequality, and Evaluation: Changing Perspectives*, edited by Ray C. Rist, Frederic P. Martin, and Ana Maria Fernandez. Washington, DC: World Bank.
- Lahey, Robert, and Dorothy Lucks. 2023. "The Impact of the COVID-19 Pandemic on the Effective Use of Evaluation in Supporting the Sustainable Development Goals (SDGs)." In *Policy Evaluation in the Era of COVID-19*, edited by Ray Rist, Indran Naidoo, and Pearl Eliadis. London: Taylor & Francis.
- Martín-Blanco, Cathaysa, Montserrat Zamorano, Carmen Lizárraga, and Valentín Molina-Moreno. 2022. "The Impact of COVID-19 on the Sustainable Development Goals: Achievements and Expectations." *International Journal of Environmental Research and Public Health* 19 (23): 16266. <https://doi.org/10.3390/ijerph192316266>.
- May, Kathryn. 2019. "Whatever Happened to Deliverology?" *Policy Options*, May 15.
- SDG Knowledge Hub. 2020. "International Solidarity Shouldn't End with COVID-19: We Must Deliver for All." By Tijjani Muhammad-Bande and Mona Juul. International Institute for Sustainable Development. <https://sdg.iisd.org/commentary/guest-articles/international-solidarity-shouldnt-end-with-covid-19-we-must-deliver-for-all/>.
- United Nations. 2025. *Transforming Our World: The 2030 Agenda for Sustainable Development*. New York: United Nations.
- United Nations Department of Economic and Social Affairs (UNDESA). 2021. *The Impacts of the COVID-19 Pandemic on Stakeholder Engagement for the SDGs*. New York: United Nations.
- United Nations Department of Economic and Social Affairs (UNDESA). 2023. *Implementing the SDGs: Strengthening the Voluntary National Reviews (VNRs)*. CDP Background Paper no. 56. New York: United Nations.
- United Nations Evaluation Group (UNEG), Task Force on National Evaluation Capacity Development. 2012. *National Evaluation Capacity Development: Practical Tips on How to Strengthen National Evaluation Systems*. New York: UNEG.
- United Nations General Assembly. 2023. *Strengthening Voluntary National Reviews through Country-Led Evaluation*. Resolution A/RES/77/283, adopted April 26.
- Van den Berg, Rob D., Indran Naidoo, and Susan Tamondong. 2017. *Evaluation for Agenda 2030: Providing Evidence on Progress and Sustainability*. New York: UNDP and IDEAS.
- Van den Berg, Rob D., Cristina Magro, and Marie-Hélène Adrien, eds. 2021. *Transformational Evaluation for the Global Crises of Our Times*. London: Nielsen.

8 Learning Agendas

From studies to streams of evidence

*Karol Olejniczak, Tomasz Kupiec and
Yaerin Park*

Introduction

The problem of limited evaluation use has long been noted in literature and practice (e.g., Weiss, 1988). Two primary challenges emerge: the “two communities” challenge (Caplan, 1979), reflecting the gap between knowledge users (decision-makers and civil servants) and producers (evaluators and researchers), and the “streams of evidence” challenge (Rist & Stame, 2006), which highlights the difficulty in synthesizing fragmented studies into cohesive insights for decision-making.

The practice of Learning Agendas (LAs) potentially addresses both challenges. LAs are “*a strategic plan for research and evaluation activities, [...] providing a list of important questions along with plans for addressing them, while balancing interests, informational needs, and time horizons for different organizational decision-makers*” (Newcomer et al., 2022, p. 64). LAs, in theory, enable the combination of diverse evidence streams (including evaluations) to deliver relevant, timely knowledge for specific policies, strategies or programs. They also encourage meaningful engagement between evidence users (e.g., senior decision-makers, implementation teams) and providers (e.g., evaluators, researchers, data officers). The responsibility for the process is usually assigned to commissioning authorities – bodies commissioning the evaluation in public agencies.

This chapter presents early lessons from using LAs in public policy processes, drawing on pilot studies in the United States, a comparative analysis of 24 LAs developed by US federal agencies, and applications within multi-annual EU co-funded programs in Poland.

We begin by defining the LA concept, reconstructing the underlying assumptions for evidence-based policies with two conceptual tools. First, we use “if-then” causal logic to understand the Theory of Change (ToC) guiding LAs. Second, we apply the COM-B model (Capability–Opportunity–Motivation–Behavior) (Michie et al., 2014) to map the causal mechanisms assumed to influence evidence use among knowledge users. These frameworks are widely applied by policy designers (e.g., Olejniczak et al., 2020) and evaluators (e.g., Mayne, 2017).

In the empirical section, we examine recent LA initiatives, with a particular focus on advanced cases in the United States and Europe, identifying key issues in

current practices. The final section discusses how commissioning authorities can use LAs to promote democratic practices in evidence-informed policies.

This chapter is intended for authorities commissioning evaluation, such as evaluation unit staff, officers, policy analysts and senior decision-makers engaged in evidence-based capacity building. We aim to illustrate that, despite its challenges, the LA offers a promising shift by aligning evidence streams with policy-relevant questions. Practical step-by-step guidance is also provided to support effective LA implementation. We hope this chapter will foster more productive and democratic relationships between commissioning bodies, evaluators and evidence users in public policy contexts.

Learning Agendas – a conceptual background

Definition and origins of the Learning Agenda

LA comprises activities aimed at defining research questions essential for guiding an agency's short- and long-term research and evaluation initiatives (Nightingale et al., 2018). It operates under the premise that agencies can improve programs and policies by focusing learning efforts – such as evaluations and data analytics – on core challenges (Feldman, 2017). Typically, an LA is a document listing key research questions linked to data sources, learning activities and evidence types required to address those knowledge needs.

In US evaluation practice, the Office of Management and Budget (OMB) publications outlines LAs as tools for gathering evidence on effective practices, generating new knowledge and employing innovative methods to develop and deliver programs (OMB, 2019; OMB 2021a; OMB 2021b). By centering on targeted questions, LAs help agencies operate more efficiently, enhancing decision-making with well-curated evidence.

There is no fixed process for developing an LA; however, Hart et al. (2024) identifies key steps: gathering stakeholders, conducting literature reviews, identifying and prioritizing questions, developing plans to address them, conducting analyses, involving stakeholders, drawing conclusions and disseminating findings.

The concept of LA originated in the 1960s within business and education, initially referring to personal learning paths (Cangelosi & Dill, 1965). It gradually expanded into the public sector, with agencies like USAID and the Department of Labor integrating LAs into their programs. Interest in LAs surged following the 2017 US Commission on Evidence-Based Policymaking report, which sought to enhance data privacy, secure access and strengthen federal evidence-based policymaking (Hart & Newcomer, 2024). The Commission recommended that all US federal agencies develop evidence plans, now known as LAs, a requirement for the 24 largest federal agencies under the 2018 Evidence Act. The initial LA creation began in 2019 and concluded in 2022, with the documents publicly accessible on the government evaluation platform (Evaluation Officer Council, 2024).

The main assumptions and novelties of Learning Agendas

LA is built on the pragmatic premise that evidence, including evaluation findings, should serve the mission, strategy or intervention (program or policy) of a public agency. This highlights the information needs of decision-makers responsible for the successful design and delivery of these interventions – referred to here as evidence users.

While this actor-focused perspective on evidence use is not entirely new in evaluation literature (Patton & Campbell-Patton, 2021), the LA's logic places knowledge users at the center, with commissioning authorities and evaluators acting as service providers – knowledge brokers and evidence producers. This shift in perspective underpins the entire causal logic (Theory of Change) of the LA, which we reconstruct using a simple if–then framework, as illustrated in Figure 8.1.

A set of assumptions underpins each element of this ToC. The first set relates to commissioning authorities acting as facilitators in recognizing diverse evidence users linked to a specific mission or intervention. They gather input from both within the agency and its external environment, helping users anticipate and articulate their future knowledge needs. This involves defining (a) the questions they need answered, (b) the decision situations where evidence will support them and (c) the type of evidence required for decision-making. An additional assumption is that, with commissioning authority support, users can formulate insightful, operational and strategic questions that are well-focused.

The second set of assumptions concerns the actions of evidence producers. Evaluators, policy analysts, researchers, experts and internal evaluation units are expected to supply knowledge users with a continuous flow of research-based evidence based on commissioned guidelines. The produced knowledge should have three qualities: (1) timely delivery aligned with the decision-making process, (2) relevance to users' articulated questions and (3) credibility, maintaining methodological rigor and acceptable form for users.

The third, and most complex, set of assumptions addresses why knowledge users would be inclined to use evidence in their work. To explain this, we apply the COM-B model from behavioral literature, which posits that a specific behavior (B) – in this case, evidence use – is influenced by the combination of users' capacity (C), opportunity (O) and motivation (M). For capacity, evidence users should possess critical thinking skills to integrate evidence into decision-making – specifically, the ability to read, assess and interpret evidence and construct policy arguments. Opportunity within the LA context involves two elements: first, users receive timely, relevant knowledge aligned with agency operations; second, they operate in an organizational environment that values and supports evidence-based decision-making. Motivation is shaped by a sense of ownership, as users are more engaged with results addressing their own questions. Additionally, users are motivated by practical considerations, recognizing that evidence use leads to better decision-making and mitigates risks within their areas of responsibility.

The final set of assumptions in the LA concerns effective outcomes based on democratic accountability, asserting that decision-making grounded in research

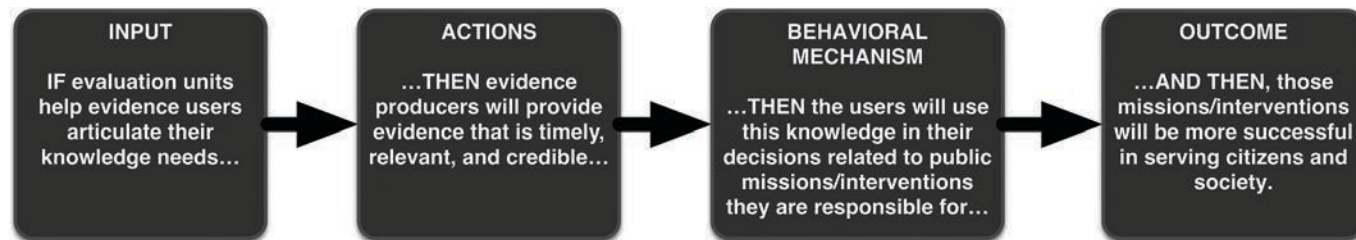


Figure 8.1 Theory of change (ToC) for Learning Agenda.

Source: Own work based on LA documents and regulations.

evidence is superior to emotional or fact-free choices. This principle underpins evaluation practice, evidence-based policymaking and the foundations of modern, science-driven societies.

Initial reflections and lessons from current practice

Overview of recent studies

The primary implementation of LAs has occurred within the US federal administration, encompassing all federal agencies and cross-agency plans. Initiated in 2019 and concluded in 2022, this substantial administrative effort has drawn academic interest, leading to studies and initiatives underpinning the reflections in this chapter. Key sources are summarized below.

The IBM Center report (Newcomer et al., 2021) analyzed pioneering US agencies that developed LAs before legal mandates, aiming to provide recommendations for evaluation officers, the OMB and Congress to enhance the effectiveness of initial federal agency LAs. The authors identified three preconditions for successful LA development: (1) consensus on mission objectives and goals, (2) commitment to the LA among leaders, staff and stakeholders, and (3) selecting an appropriate scope for the agenda (e.g., program, division, bureau or entire agency). The report also introduced the Learning Agenda Design Sprint (LADS), a seven-step model for LA development.

The Data Foundation project (Hart et al., 2024) reviewed 23 of 24 US federal LAs created under the Evidence Act, focusing on how its key concepts and OMB guidance were applied. Agencies varied in detail, structure and alignment with strategic plans, interpreting research methods and activities differently. The review recommended more flexibility and simplification for future LAs to suit each agency's needs.

The LADS approach, outlined in the IBM report, was developed and tested within the European Union's Cohesion Policy, where member states are required to establish evaluation plans. After testing the model in various organizations across the United States and Europe, it was used to create an evaluation plan in Poland for a program co-managed by the Ministry of Regional Policy and other government agencies. This initiative demonstrated LADS as a collaborative framework for multi-organizational evaluation planning, aligning with EU requirements. However, it highlighted limitations: the approach is time-intensive, requiring strong stakeholder commitment and standard design practices, such as divergent–convergent thinking and iterative testing, prove difficult to apply within the linear processes typical of public administration, particularly for complex programs and policies.

Key observations on existing Learning Agendas

Based on the review of 23 LAs and process tests across various organizational contexts, we identified areas needing consideration, improvement and further development in planning the LA process.

(1) Variety in the scope of the Learning Agendas

The starting point for an LA involves a clear reflection on its subject. Practices from US federal agencies and EU applications have shown a range of approaches to LA scope, leading to diverse logical connections between elements of analysis and synthesis and, ultimately, differing document structures. From current practice, we tentatively identify three basic options.

(2) Spectrum of evidence users

A distinctive aspect of the LA is its focus on evidence users – individuals and teams who apply the evidence generated by learning activities. However, practice has shown some confusion, with the users' perspective often conflated with both the limited group involved in LA development and the broader pool of policy stakeholders covered by the LA. Ideally, all policy stakeholders would participate in formulating questions and evidence needs. Practically, however, three groups can be identified: (1) a core group involved in LA development (representatives and key brokers); (2) direct evidence users expected to make decisions based on recommendations (e.g., decision-makers, program managers); and (3) a broad

Table 8.1 Possible options for focusing Learning Agendas

<i>Option A: Strategic goals and objectives</i>	<i>Option B: Areas of agency's actions</i>	<i>Option C: Program</i>
LA is built around the list of strategic goals, with groups of priority questions for each specific goal. The final LA document probably has a chapter for each Strategic Goal, divided into sections devoted to Priority Questions. Each PQ description covers learning activities, their timing, data and methods, use and dissemination, and challenges with mitigation strategies.	LA is built around the agency's operational areas, with priority questions generated for the specific areas the agency addresses. The final LA document is usually divided into chapters, representing each operation area. Each chapter details PQs relevant to a specific area, learning activities, data and methods, use, dissemination, and challenges with mitigation strategies.	LA is built around the program's objectives, with questions about the program. The final LA document starts with an introduction presenting the program and its goals. Subsequent chapters present priority questions specific to this program, along with learning activities, their timing, data and methods, use and dissemination, and challenges with mitigation strategies.
The nesting of LA in the Agency Strategic Plan gives a clear timeframe for planning further activities.	The agency activities area's ongoing nature can cause problems with establishing a timeline for learning activities.	Focusing LA on a specific program means that LA will closely follow the program implementation timeline.

audience of policy stakeholders who should be informed yet lack direct influence over policy.

(3) Degrees of stakeholders' engagement

Agencies varied in their approach to stakeholder engagement scale, the depth of co-creation and the level of interaction in developing different elements of the LA, likely due to differences in agency size and structure. This variation necessitates greater flexibility in the level of engagement, range of partners and number of feedback points. Examining the practices and implications of top-down (agency-led with limited stakeholders) versus bottom-up (stakeholder-driven with agency facilitation) approaches would provide valuable insights.

(4) Emerging "menu" of building blocks for Learning Agendas

Most agencies used the main building blocks of LAs – priority questions, learning activities, data and methods, challenges, and mitigation strategies – though interpretations varied widely despite OMB's initial guidance. A detailed study of these variations, including popular question types, evidence categories and commonly cited challenges, could offer a "menu" of options for participants to consider during brainstorming sessions for the next generation of LA.

(5) The challenges of developing the right questions

Formulating well-crafted research questions is essential to effective research and evaluation, yet presents notable challenges in the LA process. Ideally, evidence users, supported by the commissioning authority, would develop insightful questions. However, practice reveals several issues. First, there is often a tendency to focus on operational, short-term, technical matters, resulting in lengthy, detailed lists with few strategic questions, many of which resemble survey items rather than genuine research questions. Second, synthesizing questions from diverse users with different interests and program perspectives is a challenging task. The third issue arises in prioritizing questions effectively without disengaging original contributors. Some agencies have addressed these challenges by using a Theory of Change (ToC) framework, articulating the ToC for a program or priority, and then building questions around it.

(6) Making the progress of the design process more visible and tangible

Participants in the LA process often expressed confusion over its iterative nature, the transition from divergent thinking in creative sessions to convergent document content, and how incremental steps ultimately led to the final LA. The challenge, therefore, lies in clarifying the process and aligning it with the linear nature of

bureaucratic workflows. In essence, each step should build upon prior actions, concluding with a clear output that contributes to the final LA document.

Current unknowns in Learning Agenda practice

The implementation of LAs, especially those formalized at the US federal level, marks a shift towards a whole-of-government approach and inter-organizational collaboration in evidence-based policymaking. However, it remains too early to evaluate their impact on the effectiveness of evidence-based policy systems. At this stage, we highlight key practical aspects needing further consideration, emphasizing the importance of continued research in this field.

A practical issue lies in the relationship between LAs and Evaluation Plans (EPs). LAs appear to be broader, multi-year documents encompassing diverse evidence types, while EPs focus solely on evaluation activities within an annual timeframe. Despite these differences, evaluation units and officers are central to developing both, leading to challenges such as overlap, document redundancy and potential overburdening of policy actors. This raises the question of whether LAs might evolve into a new framework for creating EPs or serve as an overarching concept for EPs, policy analytics and other evidence-based government initiatives.

The second issue concerns the assumption of collaboration between evidence producers and users. Current practice highlights the complexities of co-production, particularly in sustaining stakeholder engagement. This process, often iterative and divergent, conflicts with the linear nature of public administration. In large, hierarchical agencies with wide stakeholder networks, involving all actors in brainstorming is challenging. LAs aim to engage both evidence users (senior decision-makers and implementation teams) and providers (evaluators, researchers and data officers) in discussions on evidence needs for improved policy design and delivery. However, practical questions remain: who should be encouraged to participate, at which stages, in what activities and to what extent – whether in creation, synthesis, prioritization or simply in feedback provision, to ensure that the process is both efficient and democratic.

The third issue concerns the evolving role of commissioning authorities, particularly evaluation units, and their relationship with evaluators. Our initial observations suggest that the long-term evidence planning in LAs shifts the dynamic between commissioners and evaluators in defining evaluation questions. In traditional practice, evaluation questions are identified during the study's conceptualization, with discussions between users, the evaluation unit and evaluators occurring concurrently as the best tender is selected. The LA approach, however, separates these interactions: (a) during LA formulation, where knowledge needs are identified and discussed between users and the evaluation unit, and (b) when an evaluation study is commissioned, where these needs are negotiated between the evaluation unit and evaluators. The practical application of this approach and its impact on the dynamics among officers, units, commissioners and evaluators remain an open question.

Ideas for using Learning Agendas in democratic-oriented commissioning of evaluation

Practical suggestions on designing Learning Agendas

In this last section, we take the perspective of the commissioning authority – personnel from evaluation units, who could be undertaking the effort to initiate and carry out LA development in their public organizations. Based on the experiences so far, we put forward eight steps in the design process, indicating options and choices available.

Step 1: Scope of Learning Agenda

The design begins with establishing the scope of the LA document. This decision is fundamental – it determines the logic of the process and structure of the final LA document. Thus far, practices indicate three distinct options, which we summarized in Table 8.1. We describe the next steps of the process from the perspective of Option A, which is the one closest to the OMB requirements. However, we indicate the peculiarities of the other options when they are relevant. We believe this decision is made between the commissioning authority – evaluation unit or officer, who is formally responsible for drafting the LA, and the agency's leadership.

The OUTPUT of this step provides clear information about agency action or priority that the LA will support with evidence. In the case of Option A, this is an established link between the Strategic Plan of the Agency and the LA.

Step 2: Users of evidence

This step aims to establish a shortlist of those who will be supported by the knowledge generated with the LA. This is again usually decided between the commissioning authority and the agency's leadership.

Users are the individuals, teams or organizations that need LA to make better decisions regarding the execution of strategic goals. It should be noted that not all stakeholders of the agency strategy are evidence users; this group is limited to those directly responsible for decisions about strategic goals execution and who rely on evidence to enhance their decision-making.

It is at this stage that decisions are made regarding the inclusivity and participatory nature of the evaluation practice. Even if the actual scope of users is narrow, the LA process makes this selection more deliberate and transparent.

The OUTPUT of this step is a shortlist of users. This material will be used in step 3 to establish who, and in what capacity, should be involved in brainstorming priority questions, as well as in step 5, when questions will be assigned to a specific audience.

Step 3: Organization of the process

At this stage, the commissioning authority developing the LA determines who should be involved in the further steps and to what extent. This includes defining engagement levels of four actor categories: evaluation and evidence staff (knowledge producers), delivery staff implementing strategic goals, decision-makers and external stakeholders.

There are two approaches. In the top-down option, a small team of evaluation staff, strategic departments, leadership and experts generates ideas, then seeks feedback from broader agency teams and stakeholders. The bottom-up approach involves gathering ideas from a wider audience at the outset, which the evaluation unit later synthesizes and prioritizes before sending back for feedback. The choice of approach in practice may determine whether the LA genuinely fulfills its promise to shift responsibility for the use of evaluation from the commissioning authority to the users and whether users take ownership of the evidence.

The OUTPUT of this step is the plan for engaging stakeholders in the LA. This is crucial for the smooth and effective development of LA content. An explanation of stakeholder engagement is also a good practice promoted in the US OMB Guidance.

Step 4: Priority questions

This step aims to establish what must be learned to support the effective execution of the agency's goals, articulated as specific questions.

In US practice, OMB recommends differentiating between operational vs. strategic and long-term vs. short-term questions. Our suggestion for brainstorming is to focus on the intent behind each question, with a short list of options: (1) questions exploring situations and trends in the policy area, (2) questions on effects and mechanisms (what works and why), (3) questions seeking policy solutions and (4) questions about implementation progress.

The questions are determined by the users selected in step 2. However, it may fall within the remit of the evaluation unit or officer to ensure that the entire portfolio of questions contributes to breaking the monopolies of problem definition and issue formulation, which is a principle of democratic evaluation (Hanberger, 2006).

The OUTPUT of this step should be a list of priority questions related to a specific strategic goal. This fulfills the requirement of the Evidence Act on a list of policy-relevant questions that the agency intends to address with evidence.

Step 5: Time frames for evidence delivery

This step identifies who will be interested in the answers to the questions and the deadlines for obtaining these answers. For each priority question, the design team should address two aspects. First, they should identify potential evidence users, distinguishing between those who "need to know" (answers crucial for their

learning and work) and those for whom it is “nice to know” (valuable but not necessary). Second, designers must establish deadlines or time frames for addressing specific priority questions, either in terms of short-term, long-term and ongoing questions or specific dates on the calendar. Monthly perspectives are typically used in Option C (LA for Program).

The OUTPUTS of this stage include short information for each priority question, specifying the major user of the answers (evidence) and the expected timeline of the delivery of the evidence.

Step 6: Learning activities

The goal is to determine the methods and data for learning to build the evidence base for priority questions. We suggest two rounds of brainstorming: the first round identifies major learning activities, using the OMB’s knowledge production methods, including policy analysis, program evaluation, performance measurement and foundational fact-finding. Follow-up sessions will focus on data sources, detailed methods and analytical approaches required to develop evidence for each priority question, which can be conducted by the knowledge producers’ group.

The OUTPUT of this stage is a list of (a) learning activities, (b) data, and (c) methods and analytical approaches for all priority questions of the LA.

Step 7: Dissemination activities

In this step, the designers of the LA focus on clarifying what methods and channels they intend to use to share evidence with knowledge users. When thinking about the delivery of evidence to users, it is worth considering decision-makers’ profiles and their preferred forms and means of communication. The evidence could be communicated in multiple ways, such as briefings, meetings, structured team discussions and other forms of presentations.

While the primary focus is on the users selected in step 2, the evaluation unit or officer also has a responsibility to consider communication forms and means accessible to other stakeholders. This ensures that the evidence serves not only the learning needs of decision-makers but also informs a broader audience.

The OUTPUT of this stage is paragraphs about dissemination activities.

Step 8: Challenges and their mitigation strategies

The final step involves reflecting on challenges the agency might encounter in implementing the LA and strategies to mitigate them. Potential challenges in evidence building include the complexity of the policy topic and its regulatory environment, methodological issues in researching the policy, resource and skill constraints within the ministry or agency, and data availability, quality and sharing issues.

The OUTPUT of this stage is typically a brief reflection on challenges specific to priority questions, but challenges can also be horizontal, affecting multiple

priority questions identified by the agency. Lessons learned from these challenges may inform the process of developing future LAs.

Broader opportunities for promoting democratic practices

We close the chapter with a broader reflection on how bodies commissioning evaluations can use LAs to promote democratic practices. We identify the following potential advances.

LAs provide a framework for engaging diverse stakeholders, including citizens and marginalized groups, in collaborative reflection. By involving these groups in setting research questions and priorities, commissioners foster transparency and democratize knowledge creation, enriching policy decisions. The LA process promotes learning by encouraging decision-makers to identify knowledge gaps and consider actions informed by new insights. Commissioners guide this reflective process, challenging the *status quo*, addressing strategic blind spots, and incorporating varied perspectives. Ultimately, these approaches empower stakeholders, making them active participants in democratic decision-making and fostering deeper understanding and accountability in public programs. While not all evaluations or evidence are produced through a participatory approach, drafting an LA inherently is, thereby democratizing evidence-based decision-making. Through this process, decision-makers engage in discussions and gain insights into other stakeholders' knowledge needs or views on what should be known, valued or measured regarding specific goals, policy areas or programs. Conversely, policy stakeholders gain awareness of the evidence generated on particular goals, how it informs decisions, and what evidence they may request to be shared.

LAs foster an ongoing process of institutional learning, enabling organizations to evolve based on evidence and shifting societal needs. This continuous feedback loop helps public institutions remain adaptable and responsive, aligning policies with current public priorities. Commissioning authorities can use LAs to embed this democratic resilience, ensuring that insights from past actions drive future policy improvements. This approach cultivates a culture of reflection and adaptation within public organizations, contributing to a more dynamic and responsive democracy.

Finally, in developing LAs, commissioners serve as mediators between technical experts, strategic decision-makers and the public. By aligning expert insights with public input, they ensure that policies are both scientifically grounded and democratically legitimate. This balance integrates technical evidence without eclipsing public values and preferences, fostering policies that reflect societal priorities. Thus, LAs ensure that expert knowledge supports, rather than dictates, democratic decision-making.

In summary, the LA seems a promising effort to address the postulate formulated almost 20 years ago (Rist & Stame, 2006) – developing meaningful streams of studies to support evidence-based policies. LAs offer commissioners a potentially powerful, although time-consuming and effortful, method to enhance democratic practices in daily evidence-informed decision-making.

References

- Cangelosi, Vincent E., and William R. Dill. 1965. "Organizational Learning: Observations toward a Theory." *Administrative Science Quarterly*, 175–203.
- Caplan, Nathan. 1979. "The Two-Communities Theory and Knowledge Utilization." *American Behavioral Scientist* 22 (3): 459–70.
- Evaluation Officer Council. 2024. "Us Evidence Plans."
- Feldman, Andrew. 2017. *Strengthening Results-Focused Government. Strategies to Build on Bipartisan Progress in Evidence-Based Policy*. Washington D.C.: Brookings Institution.
- Hanberger, Anders. 2006. "Evaluation of and for Democracy." *Evaluation* 12 (1): 17–37.
- Hart, Nicholas, and Kathryn Newcomer. 2024. "Sustaining Momentum for Evidence-Informed Policymaking: The Case of the U.S. Government." In *Evaluation in the Post-Truth World*, edited by Mita Marra, Karol Olejniczak, and Arne Paulson, 134–49. New York, NY: Routledge.
- Hart, Nicholas, Sara Stefanik, Christopher Murrell, and Karol Olejniczak. 2024. *Blueprints for Learning. A Synthesis of Federal Evidence-Building Plans under the Evidence Act*. Washington D.C.: The Data Foundation.
- Mayne, John. 2017. "Theory of Change Analysis: Building Robust Theories of Change." *Canadian Journal of Program Evaluation* 32 (2): 155–73.
- Michie, Susan, Robert West, Rona Campbell, Jamie Brown, and Heather Gainforth. 2014. *ABC of Behaviour Change Theories. An Essential Resource for Researchers, Policy Makers and Practitioners*. London: Silverback Publishing.
- Newcomer, Kathryn, Karol Olejniczak, and Nicholas Hart. 2021. *Making Federal Agencies Evidence-Based: The Key Role of Learning Agendas*. Washington, DC: The IBM Center for the Business of Government.
- Newcomer, Kathryn, Karol Olejniczak, and Nicholas Hart. 2022. "Learning Agendas: Motivation, Engagement, and Potential." *New Directions for Evaluation* 173: 63–83.
- Nightingale, Demetra Smith, Keith Fudge, and Will Schupmann. 2018. *Evidence Toolkit: Learning Agendas*. Washington, DC: Urban Institute and Evidence-Based Policymaking Collaborative.
- Office of Management and Budget (OMB). 2019. *M-19-23: Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance*. White House. Retrieved from www.whitehouse.gov/wp-content/uploads/2019/07/m-19-23.pdf.
- Office of Management and Budget (OMB). 2021a. *M-21-27: Evidence-Based Policymaking: Learning Agendas and Evaluation Plans*. Washington D.C.: White House. www.whitehouse.gov/wp-content/uploads/2021/06/M-21-27.pdf
- Office of Management and Budget (OMB). 2021b. Circular A-11. Washington, DC: White House.
- Olejniczak, Karol, Paweł Śliwowski, and Frans Leeuw. 2020. "Comparing Behavioral Assumptions of Policy Tools: Framework for Policy Designers." *Journal of Comparative Policy Analysis* 22 (6): 498–520.
- Patton, Michael Quinn, and Charmagne E. Campbell-Patton. 2021. *Utilization-Focused Evaluation*. Newbury Park, CA: SAGE Publications.

- Rist, Ray C., and Nicoletta Stame, eds. 2006. *From Studies to Streams: Managing Evaluative Systems*. New Brunswick, NJ: Transaction Publishers.
- U. .S. Congress. 2018. P.L. 115-435, Foundations for Evidence-Based Policymaking Act of 2018. (approved: January 14, 2019). H.R. 4174. 115th Congress, 1st Session.
- Weiss, Carol H. 1988. "Evaluation for Decisions: Is Anybody There? Does Anybody Care?" *Evaluation Practice* 9 (1): 5–19.

9 Democratic implications of commissioning market-based evaluation

Insights from Italy's Cohesion Policy

Mita Marra and Virgilio Buscemi

Introduction

Over the past two decades, the institutionalization of evaluation has gained momentum, propelled by the global performance management movement and the European Cohesion Policy (Furubo et al., 2002; Jacob et al., 2015). National administrative reforms and EU Structural Funds have supported the development of performance regimes across EU member states, including Italy (European Commission, 2007; Martin et al., 2016), while simultaneously expanding the space for market-based evaluation services.

Within this evolving landscape, economics and public policy research on governance and state reform has explored how public agencies procure technical assistance services to boost productivity and administrative capacity—especially within the framework of European Cohesion Policy (Barca, 2009; Allen, 2013; Rodríguez-Pose, 2013; Rodríguez-Pose et al., 2015; Polverari et al., 2020; Bachtler et al., 2024). While these studies offer important insights into public administration, they tend to overlook the specific role of evaluative and analytical services for policy design and assessment, often treating them as indistinct from broader technical assistance to public agencies. Parallel research on government–business relations has highlighted enduring barriers to the development of market-based evaluation services, including political interference, budget limitations, tight deadlines, limited evaluation cultures, cautious expectations, and skill gaps among commissioners (Le Grand, 2007; Schneider et al., 2016; Mazzucato & Collington, 2023). These analyses typically assume compliance with transparency and procurement regulations but often neglect the democratic implications of outsourcing. Key dimensions—such as the inclusiveness of contractor selection (Henrik, 2019), the rationale behind these choices, and the accountability mechanisms tied to evaluation results—remain largely unexplored.

A more specific body of literature on evaluation institutionalization and performance management has underscored the role of evaluation in fostering evidence-informed policymaking across both EU and North American contexts (Furubo et al., 2002; Trochim, 2009; Jacobs et al., 2015; Rosenstein, 2015; Martin et al., 2016; Stockmann et al., 2020; Christie & Lemire, 2019; Marra, 2018, 2019, 2021). Some scholars have drawn attention to bureaucratic and procurement-related challenges

DOI: 10.4324/9781042000968-11

This chapter has been made available under a CC-BY-NC-ND license.

in commissioning evaluations (Pattyn & Brans, 2015; Cox & Barbrook-Johnson, 2020; Schneider et al., 2016; Broer et al., 2017; Perrin & Tyrrell, 2021), advocating for a shift from a client–service model to more collaborative partnerships with flexible contracts and greater commissioner engagement (Gates, 2017). Yet, as pointed out elsewhere (Marra & Buscemi, 2026), these studies rarely unpack the commissioning process into its core components, nor do they clearly distinguish commissioning from contracting or outsourcing practices.

In relation to the evaluation marketplace, recent studies have mapped the sector in the United States (Nielsen et al., 2018; Lemire et al., 2018), Canada (Lahey et al., 2018), and the United Kingdom (Davies et al., 2018), while earlier scholarship provided insights on privately organized evaluation services in Europe (Leeuw & Rozendal, 1994; Leeuw et al., 1999; Leeuw, 2009), North America (Maynard, 2000), Australia (McTaggart et al., 1991), and globally (Jacob et al., 2015). These studies raise concerns about professional standards and ethical conduct (Nielsen et al., 2018; Peck, 2018; Lemire et al., 2018). Nonetheless, the core aspects of evaluation commissioning—such as evaluator selection, contract adjudication, commissioner–contractor dynamics, and the use of evaluative findings—remain understudied, particularly from the perspective of commissioners navigating practical constraints and learning opportunities.

Building upon our analysis of market-based procurement practices in Italy's Cohesion Policy (Marra & Buscemi, 2026), this chapter examines the commissioning of evaluation services and its democratic implications. We draw on a dataset detailing Italy's Structural and Investment (ESI) Funds evaluation expenditures during the 2007–2013 and 2014–2020 programming periods (Marra & Buscemi, 2026). In addition, we draw on a series of interviews with evaluation contractors to shed light on the funding trends for procuring evaluation, how evaluators are selected, the types of outputs procured by national and regional authorities within the commissioning system of the ESI Funds, the nature of commissioner–contractor relationships, and, finally, the mechanisms used to manage the evaluation process.

By engaging with ongoing research in evaluation policy and market development, this chapter contributes to a more nuanced understanding of the commissioning approach, methods, and relationships, which is relevant to scholars of economics, public policy, and democratic governance. It also offers practical insights for evaluation practitioners and commissioners aiming to align their roles and practices with democratic principles and enhance the legitimacy, relevance, and utility of evaluation.

The chapter is organized as follows: We first outline the study design and data sources used to examine market-based evaluation services in Italy. We then present the study findings, before discussing the key democratic implications of commissioning processes and offering suggestions for their reform.

Method and data

Our study combines qualitative and quantitative data to examine (i) the procurement of evaluation services of Italy's Cohesion Policy during the 2007–2013 and 2014–2020

programming phases at national and regional levels and (ii) the development of privately organized evaluation services, focusing on market dynamics, contracts, and the quality of evaluation knowledge produced. The analysis identifies the democratic implications inherent in the funding of ESI evaluations, focusing on the national and regional distribution of expenditure, the number and value of contracts awarded, and key contractors in both programming cycles. It also highlights methodological challenges that may compromise the quality and democratic nature of evaluations.

We used two primary data sources: the existing database from the Department of Cohesion Policy, covering outsourced and in-house evaluation assignments (2018–2021), and a novel and more comprehensive database of official invitations to bid issued by public agencies (2009–2022). We compiled this with the assistance of Lattanzio KIBS SpA, an Italian consulting firm. The latter includes contracts for evaluation studies commissioned within the 2007–2013 and 2014–2020 European Structural and Investment Funds (Marra & Buscemi, 2026).

We analyzed financial data on outsourced evaluation services, which often span multiple years, and identified contractors during both programming phases. Although the 2007–2013 data could be underestimated, the overall number and value of contracts cover both programming phases, while in-house spending data is available only for 2018–2021. Additionally, we conducted ten interviews with evaluation contractors to explore contracting and commissioner–evaluator relationships and gain insights into market-based evaluation procurement and contracting. Despite limitations of both quantitative and qualitative data used, this analysis offers an original and comprehensive overview of Italy’s Cohesion Policy evaluation commissioning to date that highlights relevant dynamics ongoing in other countries and evaluation systems throughout the world.

The commissioning system of market-based evaluation services and its democratic implications

As part of the broader commissioning system, we examine three key aspects related to the procurement of market-based evaluation services in Italy: funding trends with a national/regional government breakdown of evaluation expenditures and an analysis of the types of evaluation products procured; market development showing the configuration of private contractors; and procurement and contract management highlighting the approaches used for procuring evaluations and the mechanisms for managing contracts. By showing how evaluation services are structured, financed, and managed, the following analysis foregrounds the democratic implications of commissioning practices, particularly in relation to accountability, transparency, access to evaluative knowledge, and the capacity of evaluation to inform public decision-making (Marra & Buscemi, 2026)..

Funding trends, geographical distribution, and procured services

Available data show that, over the past two programming cycles, the number of evaluation contracts has grown, shifting from over 77 in 2007–2013 to over 88

Table 9.1 Number and value of evaluation tenders and contracts in Italy

<i>Programming phase</i>	<i>No. of tenders</i>	<i>Auction base of all contracts expressed in euros</i>	<i>Value of all adjudicated contracts expressed in euros</i>	<i>Average amount per contract expressed in euros</i>
2007–2013	77	47,053,720	33,128,735	430,243
2014–2020	109	76,702,503	53,710,157	492,754
TOTAL	186	123,756,223	86,838,892	466,876

Source: Marra and Buscemi's (2026) elaboration on evaluation tenders issued and Cohesion Policy Department data (2024).

in 2014–2020 (see Table 9.1). The number of contracts increases to 109 if we add ex ante evaluation services to the count. In financial terms, we reckon a 63 percent increase in the total amount of committed resources between 2007–2013 and 2014–2020.

Yet, if we compare these evaluation spending data with the overall allocation of resources to Cohesion Policy in Italy over the two programming phases, the increasing trend observed in the expenditure for evaluation results is quite modest in percentage terms. Specifically, the investment in evaluation has increased by 34 percent, rising from 33 million euros to almost 54 million euros from the 2007–2013 to 2014–2020 programs. By contrast, the overall volume of resources invested in Italy within the Structural Funds Programs has risen from 75 million euros in 2007–2013 to almost 140 million euros in 2014–2020, accounting for an increase of over 46 percent. Against this backdrop, less than 0.044 percent of resources were allocated to evaluation in 2007–2013, and only 0.038 percent in 2014–2020.

In the same time span, the average amount of an evaluation contract has ranged from 430 thousand to almost 500 thousand euros. Figure 9.1 shows that the value of evaluation contracts has increased in the 2014–2020 programming phase and has moved to the class of contracts ranging between 100 and 500 thousand euros.

Figure 9.2 presents how, in 2014–2020, compared to 2007–2013, the number of specific evaluation assignments has increased, while the value of contracts awarded to independent evaluators has polarized. This trend raises the stakes for selecting evaluators and managing contract operations, with important implications for commissioners' accountability for results.

Through a finer analysis, most evaluation contracts have clustered between 100 and 200 thousand euros in both the 2007–2013 and 2014–2020 phases. The latter phase has shown an almost one-third increase in the number of evaluation assignments, shifting from 16 in 2007–2013 to 23 contracts in 2014–2020. There is also a significant increase in contracts within the cluster, ranging between 300 and 400 thousand euros in the 2014–2020 phase compared to the previous period.

During the second phase of Structural Fund programming (2014–2020), there has been a considerable surge in demand for evaluation, accompanied by a corresponding growth of investment in the evaluation function outsourced to private organizations and independent contractors and an increase in the contract

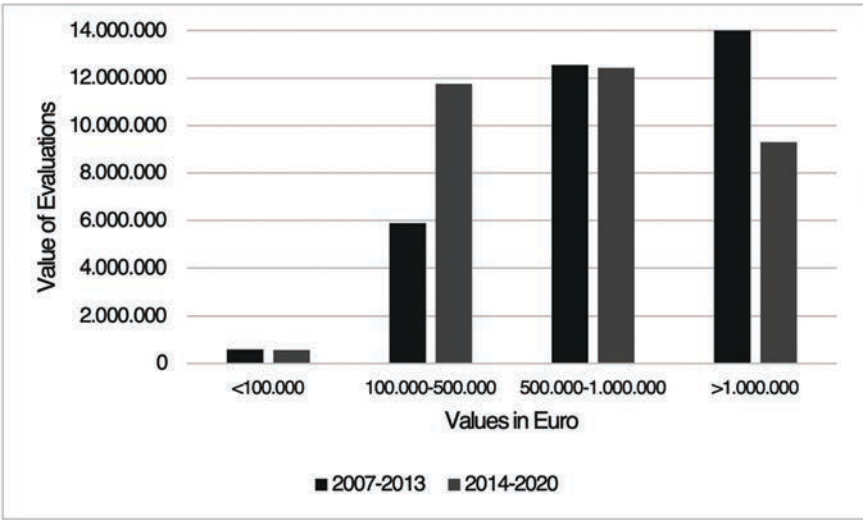


Figure 9.1 Value of evaluation contracts awarded by financial class and programming phase.
Source: Marra and Buscemi’s (2026) elaboration on evaluation tenders issued and Cohesion Policy Department data (2024).

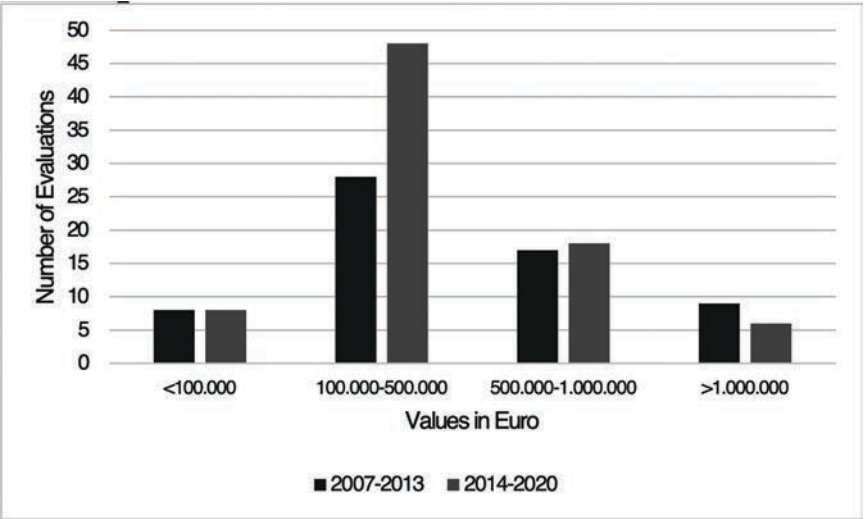


Figure 9.2 Number of evaluation assignments per cluster value and programming phase.
Source: Marra and Buscemi’s (2026) elaboration on evaluation tenders issued and Cohesion Policy Department data (2024).

Table 9.2 Number evaluation assignments and the value of evaluation contracts by financial classes

Cluster value (euros)	2007–2013 programming phase			2014–2020 programming phase		
	No. of contracts	Auction base	Rebate	No. of contract	Auction base	Rebate
N/A	15			20		
0–100	8	594,310	24	8	563,771	16
>100–200	16	1,980,200	30	23	3,290,497	27
>200–300	4	1,005,500	26	10	2,399,914	21
>300–400	6	2,060,600	32	8	2,968,109	18
>400–500	2	835,000	30	7	3,090,379	40
>500–600	3	1,565,000	32	6	3,123,575	36
>600–700	5	3,306,100	30	4	2,285,780	22
>700–800	3	2,184,750	32	3	2,285,940	43
>800–900	2	1,724,000	28	3	2,522,008	51
>900–1 mln	4	3,762,500	30	2	1,910,590	43
>1–1.5 mln	6	7,616,775	24	4	4,417,796	49
>1.5 mln	3	6,494,000	32	2	4,881,676	26

Source: Our elaboration on evaluation tenders issued and Cohesion Policy Department data (2007–2013/2014–2020).

values. This increase is likely to affect the types and quality of evaluation services provided, as well as the characteristics, size, and practices of the evaluation contractors.

However, as illustrated in Table 9.2, the average rebate applied to the auction base has considerably diminished the actual financial resources allocated to evaluation contracts. The prevalence of auction rebates suggests that these practices have become standard for both the agencies that accept them and the contractors that propose and implement them. Notably, rebates averaging nearly 50 percent of the auction base for higher-value contracts reveal a significant gap between committed and disbursed expenditure for evaluation services. The question is whether evaluation contracting out ensures effective selection procedures and the production of high-quality reports that support democratic discussions, as explored in the following sections.

Mapping out evaluation contracting at the regional and national levels

Figure 9.3 provides a more nuanced view of the contracts that shape the national and regional evaluation landscape, where expenditures appear relatively balanced across the country's three macro-areas, including regional governments and national ministries.

Upon closer examination of the national and regional distribution of evaluation expenditure depicted in Figure 9.4, Lombardy, Veneto, Apulia, and Tuscany regions demonstrate the highest expenditures for evaluation contracts, despite

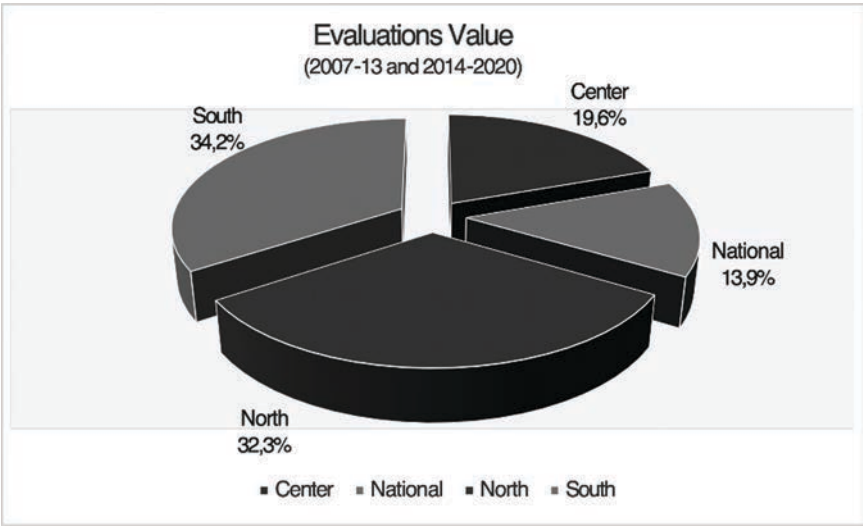


Figure 9.3 Distribution of evaluation expenditure per macro-areas in Italy.
Source: Marra and Buscemi’s (2026) elaboration on evaluation tenders issued and Cohesion Policy Department data (2024).

substantial discounts applied through auction mechanisms (as highlighted earlier). Insights gleaned from interviews indicate a growing administrative capacity in these regions to commission evaluation services and leverage evaluative information in democratic decision-making, programming, and more broadly in public life.

Indeed, Southern regional governments allocate fewer resources to evaluation relative to their overall ESI disbursements. Despite receiving approximately 75 percent of Cohesion Policy appropriations over the past two programming cycles, the value of evaluation contracts in the South accounts for only about 35 percent. Particularly, Campania and Sicily, which receive the largest Cohesion Policy allocations in Southern Italy, exhibit relatively low expenditures on market-based evaluation services. These regions account for nearly 23 percent of the total Structural and Investment Fund programs in both programming cycles, and the value of their evaluation contracts barely exceeds 11 percent of Italy’s total evaluation expenditure. While Campania and Sicily tend to favor in-house evaluation assignments over outsourced services, their performance has shown weaker administrative capacity in managing the disbursements of Cohesion Policy expenditures.

The data highlights a twofold democratic deficit: commissioners may evade results-based accountability for ESI Funds spending, while the limited production of evaluative information restricts citizens’ ability to access, interpret, and act on critical data—ultimately curbing their participation in public debate, as discussed later.

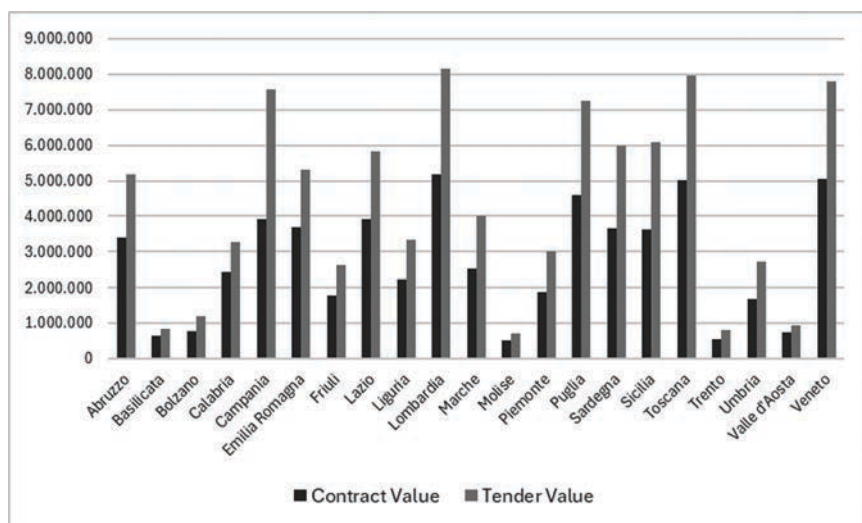


Figure 9.4 EU and national evaluation contract funds labeled by regions.

Source: Marra and Buscemi's (2026) elaboration on evaluation tenders issued and Cohesion Policy Department data (2024).

Types of evaluation services procured

Evaluation assignments include ex-post assessments, monitoring, midterm analyses, and technical assistance services, or a combination of these. While it is often difficult to discern the precise nature of these services from the financial data, the scope of evaluation assignments spans services that support program implementation and management, as well as ex-post impact assessments. The latter has become more predominant in the 2014–2020 programming phase, compared to the previous 2007–2013 phase, reflecting a growing trend towards retrospective evaluations both within public agencies and outsourced to the private sector (see also Marra, 2018, 2019, 2021; Buscemi & Cristiano, 2024). As the demand for outcomes and impact evaluations of ESI Funds programs continues to rise, there are important democratic implications. Evaluation methodologies must ensure both rigorous causal analysis and policy relevance to inform decision-making and public debate. While theory-based approaches are particularly suited to clarifying how observed changes align with program design and implementation, the commissioning process has not been thus far strategically oriented toward producing theory-based and policy-relevant studies that would not only enhance the practical utility of evaluations but also contribute to a more informed and inclusive democratic discourse.

Building the market

Data from official invitations to bid issued by public agencies for ESI Funds programs reveal that just over fifty percent of evaluation assignments are outsourced to the market. Figure 9.5 illustrates the key evaluation contractors active in Italy, showing the number of contracts awarded and their corresponding values, sourced from the evaluation tender database.

This dataset provides an estimate of the size and structure of Italy’s Cohesion Policy evaluation market, highlighting the dominance of specialized firms that have consistently participated in procurement and secured contracts over the past two programming periods. When considering both leading companies and consortia that include smaller evaluation firms and university departments, a total of 26 contractors have completed 147 evaluation assignments, totaling over 86 million euros. While the actual number of consulting firms engaged in evaluation services may be higher—since many companies act as partners in consortia or subcontractors—the firms represented in Figure 9.5 are not subsidiaries of

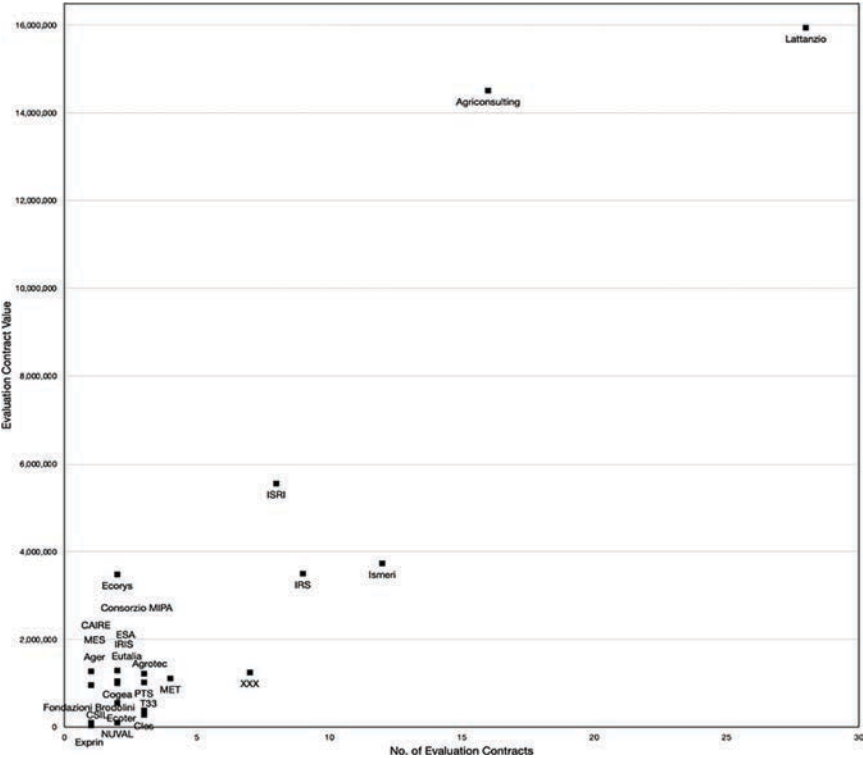


Figure 9.5 The configuration of the evaluation marketplace in Italy.

Source: Marra and Buscemi’s (2026) elaboration on evaluation tenders issued and Cohesion Policy Department data (2024).

large foreign consultancies. Instead, Italy's evaluation industry is characterized by domestically rooted businesses, reflecting a growing national capacity to address key policy-oriented issues.

However, Figure 9.5 also reveals a moderate concentration in the market for evaluation services, as indicated by a Herfindahl–Hirschman Index (HHI) score of 1,598—calculated based on the value and number of adjudicated contracts. A small number of major consultancies dominate the market, securing the majority of commissioned assignments at both the national and regional levels. In contrast, smaller firms and consortia tend to operate primarily at the regional level, often with limited access to contracts (Marra & Buscemi, 2026). This market configuration raises important democratic concerns, as the limited participation of a broader pool of contractors may constrain diversity in expertise and perspectives. A more inclusive and competitive procurement environment could enhance the overall quality, relevance, and legitimacy of evaluation activities.

Despite their deeper embeddedness in national and regional contexts—which could enhance the understanding of program dynamics and improve evaluation quality—many smaller firms remain underutilized, with lower diversity in evaluation perspectives and policy assessments. Interviews with evaluation contractors confirm this scenario, pointing to the emergence of a ‘quasi-market’ for evaluation services driven by public-sector outsourcing practices. In this environment, entry barriers are largely based on contractors’ economic and financial capacity, rather than their demonstrated quality of evaluation work. Public procurement requirements, such as financial collateral and turnover thresholds, serve as prerequisites for obtaining evaluation contracts. This financial criterion not only contributes to the expansion of evaluation assignments but also intensifies the market’s polarization between large consultancies and smaller, specialized firms. These trends have been exacerbated by the 2008–2009 global financial crisis, which deepened market fragmentation and polarization due to industry restructuring and public sector austerity measures.

Interviews suggest that the concentration and polarization within the evaluation industry are primarily driven by public funding, underscoring the government’s significant role in shaping the market for evaluation services at both the national and regional levels. Regarding the relationships between consulting firms and public commissioners, interviews reveal that, beyond financial guarantees, the ability of medium-sized consultancies to secure evaluation contracts is increasingly linked to their reputation. This reputation has become an asset for firms seeking to expand within the industry and gain access to a network of potential commissioners. As a result, leveraging political connections and influential experts is seen as crucial in the contract adjudication process, alongside evaluating bids based on pricing, deliverables, and field experience. Evaluation contractors interviewed highlight a somewhat passive stance among commissioners, who are influenced by market reputations but have yet to fully engage in coordinating evaluation services through collaboration, stakeholder participation, and interaction with policy users, as further discussed in the next section.

Approach and relationships involved in the evaluation of procurement contracts

Interviews highlight a critical aspect of evaluation contracts: the substantial financial investments required to procure evaluation services for national and regional operational programs. These contracts, typically ranging from 1 to 4 million euros, are initiated by regional or national agencies seeking comprehensive assessments of programs supported by the ESF or ERDF. While larger contracts exceed the ‘community threshold’ and thus require transparency and anti-organized crime safeguards, making the evaluator selection process more cumbersome, they also encompass a wide range of deliverables, including annual reports, communication strategy assessments, evaluations of program contributions to EU strategies, outcome evaluations, ex-post evaluations, and thematic reviews (Marra & Buscemi, 2026).

Despite the significant financial investment, the value of these contracts may not always suffice to comprehensively assess all measures within an operational program. Contractors may face the challenge of addressing evaluation questions while integrating multiple deliverables into a cohesive study. The collection and analysis of primary data, combined with qualitative assessments, may remain shallow, resulting in descriptive evaluations that merely catalog interventions across regional or national programs. As a result, large procurement contracts with numerous deliverables can raise serious concerns regarding the quality of evaluations, which may, in turn, affect the quality of democratic debate.

Procurement decisions also influence the management of evaluation services, with significant variation in administrative arrangements set by regional authorities coordinating ESI Funds Programs. Some agencies form steering groups with diverse internal personnel to address thematic, methodological, data collection, and analysis issues, as well as the dissemination of findings. Interviews suggest that these organizational arrangements are linked to the commissioners’ role and agency in fostering a constructive relationship with evaluators. When commissioners are proactive in engaging with evaluators and providing clear evaluation questions, the process benefits. However, when a single manager or unit is responsible for the entire evaluation process, the political governance orientation can undermine transparency and participation, ultimately affecting the quality and policy relevance of the outcomes as well as the democratic nature of evaluations.

Reforming commissioning in line with democratic principles

This chapter has examined how evaluation commissioning—including funding trends, procurement approaches, contracting mechanisms, and evaluation management—has shaped privately organized evaluation services in recent phases of regional Cohesion Policy and Common Agricultural Policy. Issues of transparency, quality, and market concentration are deeply influenced by how evaluators are hired, the expectations set for their work, the relationships between evaluators and commissioners, and the technical management frameworks in place. Addressing these challenges is crucial for reforming the evaluation commissioning process

according to the principles of democratic governance. Three key areas warrant attention.

Strengthening commissioners' agency and democratic accountability

Commissioners play a pivotal role in upholding democratic accountability in the allocation and evaluation of regional development policy—such as the European Cohesion Policy. Their agency lies in the strategic choices they make regarding the scope, quality, and policy relevance of evaluation activities—choices that directly affect program effectiveness, result-based accountability, and evidence-informed programming. This role becomes especially critical in Southern Italy, where a substantial share of European Structural and Investment (ESI) Funds is allocated to regional development.

In this context, commissioners who actively champion the expansion and enhancement of evaluation practices can significantly reinforce managerial accountability. By ensuring that evaluations generate credible, useful insights into the outcomes of Cohesion Policy and the Common Agricultural Policy, they empower both policymakers and citizens with knowledge that supports transparency, learning, and continuous improvement. Ultimately, commissioners who leverage evaluation as a tool for public value creation can help foster a more informed and constructive public debate on regional development policies.

Rethinking evaluator selection and the value of evaluation firms

Commissioners need to reconsider how evaluators are selected and how they perceive the value of evaluation firms, as these commissioning dimensions shape market dynamics with crucial democratic implications. First, commissioners should rethink about their choice of relying on a single contractor for gauging the ESI Funds national and regional programs. While relying on a single contractor may offer continuity and economies of scale, it also risks concentrating evaluation capacity, limiting methodological plurality, and reducing competition. Second, commissioners should reassess their preference for large firms which often stems from the complexity of evaluations requiring substantial expertise, human capital, and financial resources. This trend mirrors procurement decisions observed in the US evaluation system (Lemire et al., 2018; Peck, 2018).

As EU programs frequently demand diverse expertise and long-term engagement, commissioners could then opt for dividing contracts into specialized tasks to foster transparency, independence, and methodological plurality, exploring market segmentation strategies. This choice would require case-by-case assessment, as a single contractor managing evaluations across an entire programming period might provide a more integrated understanding of policy impacts and facilitate participatory processes that reinforce democratic accountability (Buscemi & Cristiano, 2024).

The perceived value of evaluation firms is another critical issue involving market-based evaluations with additional democratic implications. Critics argue that big

consultancies, although they can reduce transaction costs and generate profits, often prioritize reputation and perceived value over substantive contributions (Mazzucato & Collington, 2023). As for market-based evaluation services, political connections frequently influence pricing and contract awards in favor of large firms; the issue is whether evaluations should be outsourced or conducted in-house. While this decision largely depends on agencies' internal capacity, commissioners could strengthen evaluation culture—particularly in Italy's Southern regions—to develop a more interactive commissioner–evaluator relationship, enabling stakeholder engagement throughout the evaluation process and encouraging the use of evaluative findings.

Moreover, commissioners should critically assess the strategies firms use to market their services. While smaller firms often rely on networking, larger firms leverage significant rebates in auction-based contracting, as observed not only within the EU but also in the US evaluation system (Peck, 2018; Hwalek & Straub, 2018). Thus, setting a cap on auction rebates could encourage contractors to avoid excessive price reductions, ensuring fair competition and more resources for evaluation. In addition, creating opportunities for collaboration between large and small firms—leveraging their complementary strengths—could also encourage smaller contractors' participation in auctions.

Reassessing large-scale procurement and promoting collaborative learning

Large-scale procurement contracts that bundle multiple services pose significant challenges for evaluations, often limiting their ability to address core research questions, collect primary data, and analyze diverse information sources. To mitigate these constraints, commissioners should adopt a collaborative learning approach by:

- Defining evaluation questions that reflect local contexts and diverse stakeholder perspectives.
- Establishing transparent criteria for selecting evaluators and awarding contracts to ensure methodological rigor, quality, and policy relevance.
- Providing fair compensation that reflects the complexity and demands of evaluation tasks.
- Maintaining an ongoing dialogue with evaluators to strengthen long-term relationships and enhance learning.
- Actively engaging stakeholders throughout the evaluation process and ensuring that findings inform policy decisions.

These steps are essential to aligning commissioning practices with democratic governance principles in market-based evaluation services. A key consideration is the clear assignment of responsibilities, ensuring that evaluation contracts contribute to the generation of robust knowledge, capacity building, and sustained mutual learning through dynamic engagement with evaluators (Buscemi, 2024).

Evaluators should prioritize context-sensitive approaches that go beyond regulatory compliance and outcome measurement. Instead of focusing solely on indicators, they should explore how programs drive transformation in specific contexts. Recognizing the non-standard nature of interventions and the heterogeneity of beneficiaries is key to producing meaningful insights. Theory-based evaluation approaches can help bridge these gaps, ensuring relevance and facilitating the use of findings in democratic decision-making (Funnel & Rogers, 2011; Weiss, 1998). By embracing these reforms, evaluation commissioning can move toward a more transparent, accountable, and democratic process—one that ensures both the integrity and the utility and usability of evaluation findings in shaping policy, practice, and public life.

References

- Allen, Pauline. 2013. "An Economic Analysis of the Limits of Market-Based Reforms in the English NHS." *BMC Health Services Research* 13 (S1): 1–10.
- Bachtler, John, Laura Polverari, Ekaterina Domorenok, and Paolo Graziano. 2024. "Administrative Capacity and EU Cohesion Policy: Implementation Performance and Effectiveness." *Regional Studies*, 58 (4): 685–89. <https://doi.org/10.1080/00343404.2023.2276887>.
- Barca, Fabrizio. 2009. *An Agenda for a Reformed Cohesion Policy*. Independent report prepared at the request of Danuta Hübner, Commissioner for Regional Policy.
- Broer, Tineke, Roland Bal, and Martyn Pickersgill. 2017. "Problematisations of Complexity: On the Notion and Production of Diverse Complexities in Healthcare Interventions and Evaluations." *Science as Culture* 26 (2): 135–60.
- Buscemi, Virgilio, and Simona Cristiano. 2024. "La valutazione delle politiche comunitarie in Italia." *Rassegna Italiana di Valutazione* 85–86: 55–94. <https://doi.org/10.3280/RIV2023-085004>.
- Christie, Christina A., and Sebastian T. Lemire. 2019. "Why Evaluation Theory Should Be Used to Inform Evaluation Policy." *American Journal of Evaluation* 40 (4): 490–508.
- Cox, Jayne, and Pete Barbrook-Johnson. 2020. "How Does the Commissioning Process Hinder the Uptake of Complexity-Appropriate Evaluation?" *Evaluation*, 27(1), 32–56. <https://doi.org/10.1177/1356389020976157>.
- Davies, Philip, Stephen Morris, and Christopher Fox. 2018. "The Evaluation Market and Its Industry in England." *New Directions for Evaluation* 160: 29–43.
- European Commission. 2007. *Responding to Strategic Needs: Reinforcing the Use of Evaluation* (SEC(2007)213). Brussels: European Commission.
- Funnel, Sue, and Patricia Rogers. 2011. *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco, CA: John Wiley & Sons.
- Furubo, Jan-Eric, Ray C. Rist, and Rolf Sandahl, eds. 2002. *The Atlas of Evaluation*. New Brunswick, NJ: Transaction Publishers.
- Gates, Emily F. 2017. "Learning from Seasoned Evaluators: Implications of Systems Approaches for Evaluation Practice." *Evaluation* 23 (2): 152–71.
- Hwalek, Melanie A., and Victoria L. Straub. 2018. "The Small Sellers of Program Evaluation Services in the United States." *New Directions for Evaluation* 160: 125–43. <https://doi.org/10.1002/ev.20340>.

- Jacobs, Steve, Sandra Speer, and Jan-Eric Furubo. 2015. "The Institutionalization of Evaluation Matters: Updating the International Atlas of Evaluation Ten Years Later." *Evaluation* 21 (1): 6–31.
- Jordahl, Henrik. 2019. "Perspectives on Public Sector Outsourcing: Quasi-Markets and Prices." *CESifo Economic Studies* 65 (4): 343–48. <https://doi.org/10.1093/cesifo/ifz015>.
- Lahey, Robert, Catherine Elliott, and Sarah Heath. 2018. "The Evolving Market for Systematic Evaluation in Canada." *New Directions for Evaluation* 160: 45–62. <https://doi.org/10.1002/ev.20346>.
- Le Grand, Julian. 2007. *The Other Invisible Hand: Delivering Public Services through Choice and Competition*. Princeton, NJ: Princeton University Press.
- Leeuw, Frans L. 2009. "Evaluation: A Booming Business—But Is It Adding Value?" *Evaluation Journal of Australasia* 9 (1): 3–9.
- Leeuw, Frans L., Jacques Toulemonde, and André Brouwers. 1999. "Evaluation Activities in Europe: A Quick Scan of the Market in 1998." *Evaluation* 5 (4): 487–96. <https://doi.org/10.1177/135638999400830020>.
- Leeuw, Frans L., and Piet J. Rozendal. 1994. "Policy Evaluation and the Netherlands Government: Scope, Utilization and Organizational Learning." In *Can Governments Learn?*, edited by Frans L. Leeuw, Ray C. Rist, and Rolf C. Sonnichsen, 67–86. New Brunswick, NJ: Transaction Publishers.
- Lemire, Sebastian, Leslie A. Fierro, Alana R. Kinarsky, Emi Fujita-Conrads, and Christina A. Christie. 2018. "The U.S. Federal Evaluation Market." *New Directions for Evaluation* 160: 63–80. <https://doi.org/10.1002/ev.20344>.
- Marra, Mita. 2018. "The Ambiguities of Performance-Based Governance Reforms in Italy." *Evaluation and Program Planning* 69: 173–82.
- Marra, Mita. 2019. "Italian Evaluation Policy: Centralization and Judicially Enforced Accountability." In *The Evaluation Enterprise: A Critical View*, edited by Jan-Eric Furubo and Nicoletta Stame. London: Routledge.
- Marra, Mita. 2021. "A Behavioral Design to Reform Italy's Evaluation Policy." *American Journal of Evaluation* 42: 1–22.
- Marra, Mita, and Virgilio Buscemi. 2026. "A Quasi-Market for Evaluation Services: Government–Business Dynamics in Italy's EU Cohesion Policy." In *Evaluation*, online first, <https://doi.org/10.1177/13563890261446960>
- Martin, Steve, Sandra Nutley, James Downe, and Clive Grace. 2016. "Analysing Performance Assessment in Public Services." *Public Administration* 94: 129–45.
- Maynard, Rebecca A. 2000. "Whether a Sociologist, Economist, Psychologist or Simply a Skilled Evaluator." *Evaluation* 6 (4): 471–80.
- Mazzucato, Mariana, and Rosie Collington. 2023. *The Big Con: How the Consulting Industry Weakens Our Businesses, Infantilizes Our Governments, and Warps Our Economies*. New York: Penguin.
- McTaggart, Robin. 1991. "Principles for Participatory Action Research." *Adult Education Quarterly* 41 (3): 168–87. <https://doi.org/10.1177/0001848191041003003>.
- Nielsen, Steffen B., Sebastian Lemire, and Christina A. Christie. 2018. "The Evaluation Marketplace and Its Industry." *New Directions for Evaluation* 160: 13–28. <https://doi.org/10.1002/ev.20344>.
- Pattyn, Valérie, and Marleen Brans. 2015. "Organizational Analytical Capacity: Policy Evaluation in Belgium." *Policy and Society* 34 (3–4): 183–96.
- Peck, Laura R. 2018. "The Big Evaluation Enterprises in the United States." *New Directions for Evaluation* 160: 97–124. <https://doi.org/10.1002/ev.20341>.

- Perrin, Burt, and Tony Tyrrell (Eds.). 2001. *Changing Bureaucracies. Adapting to Uncertainty, and How Evaluation Can Help*. New York: Routledge.
- Polverari, Laura, John F. Bachtler, Martin Ferry, Carlos Mendez, and Jayne Ogilvie. 2020. *The Use of Technical Assistance for Administrative Capacity Building in the 2014–2020 Period*. Final report to the European Commission (DG Regio). Glasgow: European Policies Research Centre, University of Strathclyde.
- Rodríguez-Pose, Andrés. 2013. “Do Institutions Matter for Regional Development?” *Regional Studies* 47 (7): 1034–47.
- Rodríguez-Pose, Andrés, and Enrique Garcilazo. 2015. “Quality of Government and the Returns of Investment.” *Regional Studies* 49 (8): 1274–90.
- Rosenstein, Barbara. 2015. *Status of National Evaluation Policies*. Global Mapping Report, Implemented by Parliamentarians Forum on Development Evaluation in South Asia jointly with EvalPartners, 2nd ed., Ottawa, Canada.
- Schneider, Carmen H., Andrew J. Milat, and Gabriel Moore. 2016. “Barriers and Facilitators to Evaluation of Health Policies and Programs.” *Evaluation and Program Planning* 58: 208–15. <https://doi.org/10.1016/j.evalprogplan.2016.06.011>.
- Stockmann, Reinhard, Wolfgang Meyer, and Lena Taube, eds. 2020. *The Institutionalization of Evaluation in Europe*. Cham: Palgrave Macmillan.
- Trochim, William M. K. 2009. “Evaluation Policy and Evaluation Practice.” *New Directions for Evaluation* 123: 13–32.
- Weiss, Carol H. 1998. “Theory-Based Evaluation: Past, Present, and Future.” *New Directions for Evaluation* 76 (Winter): 41–55.

10 Evaluation Management Services at the MFA of Finland

Sari Laaksonen and Markus Palenberg

Introduction

This chapter analyzes the partial outsourcing of evaluation functions at Finland's Ministry for Foreign Affairs (MFA) from 2016 to 2024. It focuses on the effects of the Evaluation Management Services (EMS) framework that was introduced in 2016 on evaluation commissioning and governance. Based on our direct experiences as evaluators and EMS coordinators, we explore how EMS has shaped transparency, commissioner roles, and evaluation quality. We also examine its broader implications for democratic commissioning, as outlined in Chapter 1 of this book (Stame & Bustelo, 2027).

We have been part of MFA evaluation teams on several occasions since 2014, including evaluations under and outside of the EMS contracts. Our roles were evaluation team leader, deputy team leader, and team member. One of us also served as EMS Service Coordinator between January 2022 and June 2024. Apart from evaluation work for the MFA of Finland, we have worked for various other international clients in the development sector, experiencing different evaluation arrangements. The views expressed in this chapter are solely ours and do not necessarily reflect those of the MFA or EMS consortia.

After this introduction, the next two sections provide the necessary background on how the Development Evaluation Unit of Finland's MFA outsourced some evaluation functions (second section) and on what the rationales for this were (third section). In the fourth section, we then share our experiences with how EMS has impacted central evaluations at the MFA and to what degree the motivations for introducing EMS were fulfilled. In the final fifth section, we draw more general conclusions on how the outsourcing of evaluation functions may impact democratic evaluation practices.

Management of centralized evaluations at the MFA of Finland 2014–2024

The Development Evaluation Unit of the Ministry for Foreign Affairs (MFA) of Finland is responsible for evaluation guidelines and the development of the evaluation function.¹ It is an operationally independent unit in the sense of not reporting to any of the MFA's departments in charge of implementing Finland's development

DOI: 10.4324/9781042000968-12

This chapter has been made available under a CC-BY-NC-ND license.

policy and cooperation (Figure 10.1). Instead, the unit directly reports to the MFA's Under-Secretary of State responsible for development policy and development cooperation. Recently, the MFA has undergone an organizational reform, but because our experiences draw on the past, this does not affect the observations made in this book chapter.

In addition to its other duties,² the MFA's Development Evaluation Unit is responsible for managing "centralized" evaluations. These are comprehensive strategic and policy-level evaluations of Finland's development policy and development cooperation. Most centralized evaluations are targeted primarily at senior MFA management and the Finnish government and parliament, in addition to serving broader information and accountability purposes.

Under EMS, between three and five centralized evaluations were usually conducted per year on themes and issues of particular interest. Sometime before EMS was introduced, 10 or more smaller evaluations were implemented per year, but then the Development Evaluation Unit moved to fewer and larger evaluations to allow for comprehensive analysis of questions concerning overall Finnish development policy and cooperation. These larger evaluations required more refined methodology and expertise than earlier ones (MFA, 2019, p. 55). As a continuing principle, all centralized evaluations were commissioned to independent experts.

Until 2016, centralized evaluations were tendered and managed individually by the MFA's Development Evaluation Unit. In early 2016, the MFA launched a first call for tenders for EMS. In the summer of the same year, a consortium of two consulting companies – Particip GmbH and Indufor Oy – was contracted. The implementation of this first "EMS 1" framework contract continued until late 2020. Towards the end of this first period, Indufor Oy was replaced by another company, NIRAS Finland Oy, without interrupting the agreement. In 2020, a call for tender for a subsequent phase was launched, and the same consortium won that bid and was contracted for "EMS 2" (Box 10.1). At the time this chapter was written, the MFA launched a third framework, now called "Framework Agreement for Centralized Evaluations" (FACE) for the period 2024–28.³

Box 10.1 The EMS 2 consortium companies

The lead company of EMS 2 – Particip GmbH – has 85 staff and is headquartered in Freiburg (Germany).

The second consortium company – NIRAS Finland Oy – has 30 staff in Vantaa (Finland). NIRAS Finland Oy belongs to the NIRAS group with about 3,000 employees worldwide.⁴

Source: www.particip.de/en/companyoverview/companyfacts/ and www.niras.com/offices/finland/, www.niras.com/about-niras/, accessed March 2024.

Organisation of the Ministry for Foreign Affairs

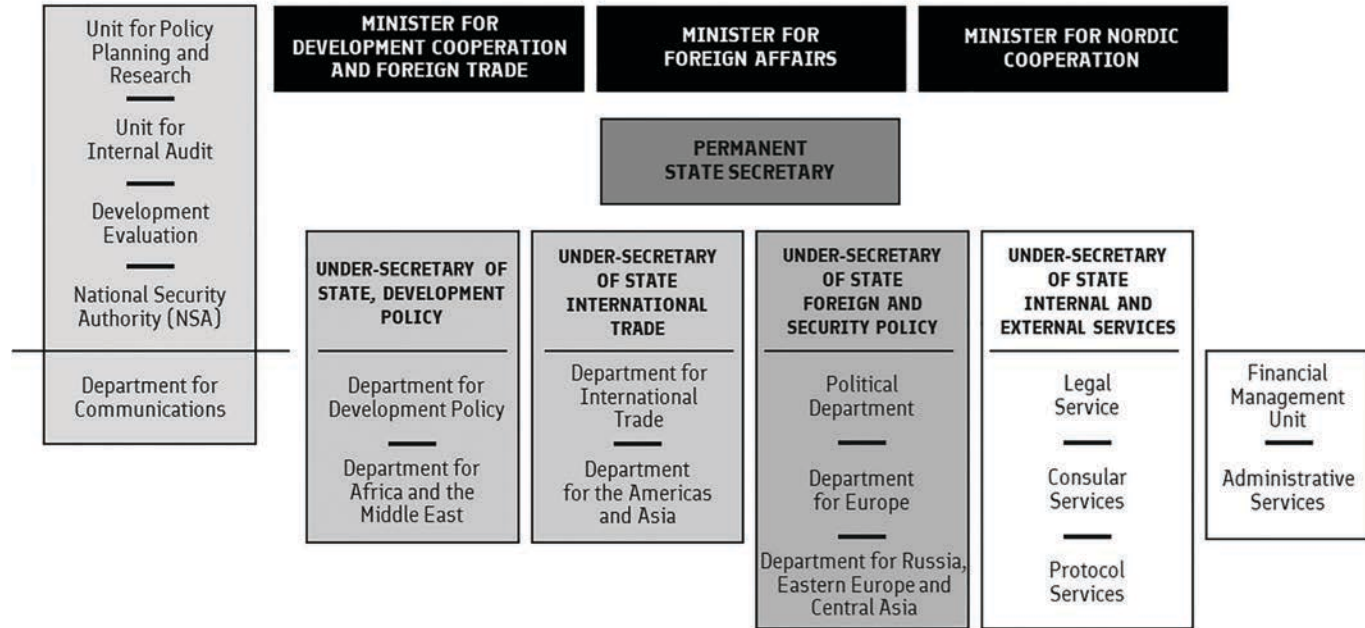


Figure 10.1 Organization of the MFA of Finland in 2023.

Source: <https://um.fi/the-organisation-of-the-ministry-for-foreign-affairs>, accessed March 2024.

Under the EMS framework contracts, the Development Evaluation Unit remained the commissioner of evaluations – following its evaluation plan – with overall decision-making authority. This means that the unit leads and steers the planning, design, and implementation of evaluations; the quality assurance of deliverables; dissemination; and the promotion of use. For each centralized evaluation, the Development Evaluation Unit designated one staff member as “Evaluation Manager” to oversee that evaluation.

EMS framework contracts implied several other governance and management arrangements. These are the Consortium Board, EMS Project Managers, the EMS Service Coordinator, and Evaluation Management Teams and are briefly explained below.

- The overall management of EMS contracts was overseen by a “Consortium Board”, the highest level oversight body within the EMS consortium, with members representing senior management from both consortium companies. The Consortium Board reported to and coordinated with the Development Evaluation Unit.
- Service provision for individual centralized evaluations was overseen by one of the two consortium companies. Between them, the two companies decided which evaluation would be led by which company. For this, each company appointed a staff member as “Project Manager”, with responsibility for leading and coordinating the administrative service provision from the consortium for that specific central evaluation.
- Another position was created with responsibilities both at the level of individual evaluations and the overall EMS contracts: the “EMS Service Coordinator”. As a backup to that position, a “Deputy Service Coordinator” was retained in case the Service Coordinator would not be able to fulfill his or her duties. EMS Service Coordinators were contracted independently from the EMS Consortium and reported to the Development Evaluation Unit. The Service Coordinator was responsible for supporting the MFA in the finalization of the evaluation ToR; suggesting evaluation team leaders and evaluation team members; advising evaluation team members in Finland’s development policy and cooperation, as well as with regard to the MFA’s organization and its partners; supporting the development of the evaluation approach and methodology; and quality-assuring the evaluation process and all evaluation outputs.
- To coordinate at the level of individual evaluations, “Evaluation Management Teams” were established. Members represented all parties involved: the Evaluation Manager from the MFA’s Development Evaluation Unit, the Service Coordinator, and, once appointed, the leader of the evaluation team.

Apart from these EMS-specific arrangements, centralized evaluations continued to be implemented by teams of independent experts, as before EMS was introduced.

Rationales for outsourcing Evaluation Management Services

In this section, we describe the reasons for introducing EMS at the MFA. We do this based on our own observations and understanding.

We consider the introduction of EMS to be driven by several interwoven factors:

- The quality of centralized evaluations,
- The capacity of the MFA's Development Evaluation Unit,
- Coordination and synergies between centralized evaluations,
- Learning from centralized evaluations, and
- Capacity development of emerging Finnish evaluators.

These motivations aligned with the MFA's broader approach to professionalizing evaluation commissioning while maintaining oversight. By shifting administrative and operational aspects of evaluation commissioning to the EMS consortium, MFA was able to focus on strategic oversight, albeit in a more indirect role than before. In what follows, the above factors are briefly described.

Quality of centralized evaluations. In our view, a central motivation for introducing EMS was the MFA's desire to ensure a constantly high quality of centralized evaluations. This concerned the methodological design of evaluations, their implementation, and the resulting end products. There were concerns that, because of staff rotations and non-expert positions, the Development Evaluation Unit lacked consistent in-house expertise to ensure high-quality centralized evaluations.

Capacity of the MFA's Development Evaluation Unit. Another driver was the limited staff capacity at the Development Evaluation Unit. For example, a recent funding decision notes in relation to the unit:

Due to the volume of concurrent evaluations and the limited human resources, the model based on commissioning requires external services in order to be efficient and to maintain high quality of evaluations deliverables.

(MFA databank, n.d./b)

Related expectations for EMS were to outsource some of the administrative tasks, such as searching for evaluation team candidates, contracting, travel arrangements, and formatting of evaluation outputs.

Coordination and synergies between centralized evaluations. EMS also aimed at a higher degree of standardization of centralized evaluations and increased coordination and synergies between them. This, so was the intention, should then improve the flexibility, effectiveness, and efficiency with which central evaluations were planned and implemented.

A recent evaluation of the utility of centralized evaluations at the MFA noted:

The EMS arrangement has multiple purposes [...]: it is expected to increase quality and coordination of the evaluations that the EVA-11 [the MFA's Development Evaluation Unit] outsources. It is also expected to increase flexibility, efficiency and effectiveness in planning and commissioning the evaluations.

(MFA, 2023)

Learning from centralized evaluations. Centralized evaluations are important knowledge instruments for the MFA (MFA, 2019), and, among other things, EMS was introduced to further support this. For example, a recent funding decision mentioned:

The use of EMS provides a platform for accumulating knowledge also within the commissioning organization.

(MFA databank, n.d./a)

Capacity development of emerging Finnish evaluators. In EMS, the MFA has systematized the inclusion of “emerging evaluators” into the teams of centralized evaluations. These are young Finnish professionals with experience of working in development cooperation and the desire to become evaluators. This reflects the intention of developing the capacity of emerging evaluators through their participation in centralized evaluations.

Our observations

In this section, we discuss our experiences and observations related to the previously discussed drivers and expectations associated with EMS. We do this along with some of the following key features of the EMS:

- 1 Outsourcing to a consortium,
- 2 EMS governance and management structure, especially the position of EMS Service Coordinator,
- 3 Standardization of management processes and requirements,
- 4 Contributions to knowledge management, and
- 5 On-the-job learning for emerging evaluators.

At the end of each section, we briefly review how these feature-specific experiences and observations relate to the three analytical aspects of democratic commissioning, namely i) the role and agency of commissioners, ii) approaches, methods, and relationships in the commissioning process, and iii) the systems in which commissioning takes place and that commissioners must navigate.

Outsourcing to a consortium

An obvious central feature of EMS is the fact that certain evaluation-related activities were outsourced to external companies rather than being delivered by the MFA. Based on our experiences, this had several effects that catered to some of the goals associated with EMS discussed in the previous section.

By contracting a consortium of established evaluation firms, MFA gained access to extensive professional networks, expanding the candidate pool for evaluation teams and enhancing evaluation quality.

However, while this structure enhanced technical quality, it also led to a concentration of evaluation commissioning within a somewhat more closed network. The structured selection process increased predictability but also reinforced a procurement model that, while not exclusionary, could limit the entry of new evaluators or firms.

A clearly positive effect is related to the MFA's human and financial resources for evaluation management. Staff resources at the MFA have been under pressure for decades. For example, a recent review of the use and utility of centralized evaluations noted that *"Half of the centralized evaluations made recommendations about the understaffing or competence development needs, and many others had findings or conclusions on them"* (MFA, 2023). Outsourcing certain evaluation management components relieves some staff pressure in the Development Evaluation Unit by transferring tasks to outside partners who are paid out of the MFA's development cooperation budget and not subject to staff headcount ceilings. For example, while the unit used to manage separate contracts for each centralized evaluation, in EMS, it managed a single contract for all centralized evaluations in a four-year period. We believe that some of the services provided under the past two EMS contracts would have been difficult for the MFA to provide from its own ranks. At the time this chapter was written, the unit had only four full-time staff.

Finally, in our view, potential conflicts of interest between the business interests of the consortium companies and the goals of the MFA were managed well. For example, evaluation teams were usually staffed with independent experts and not with employees of either company.

With relation to the broader subject of this book – democratic commissioning – we note that some activities previously implemented by the Development Evaluation Unit were now implemented by external companies. This impacted the role and agency of the Development Evaluation Unit through an increased distance between commissioners and evaluators. While previously contracting the evaluation team and overseeing the evaluation process, the unit now contracted and oversaw the external companies that did this. Similarly, EMS also impacted the commissioning process and the system in which commissioning took place, as explained further below.

EMS governance and management

A unique feature of EMS contracts was the introduction of the Service Coordinator, a coordinating position in-between the MFA's Development Evaluation Unit, the consortium, and the evaluation team.

In our experience, the Service Coordinator contributed to consistency between evaluation substance and evaluation management processes, for instance, by providing guidance on appropriate evaluation approaches to meet the MFA's evaluation purpose while remaining within available capacities and resources. Service Coordinators also contributed to consistency across centralized evaluations because the same person was closely involved in several consecutive and parallel evaluations over time.⁵

The position contributed to synergies between evaluations by informing one evaluation team of relevant findings from others. During evaluation design, Service Coordinators helped avoid duplication, particularly when previous or parallel evaluations involved the same MFA units, embassies, or stakeholders. By playing an intermediary role, Service Coordinators helped align evaluations with strategic priorities.

The Service Coordinators were senior experts in evaluation and Finland's development policy and cooperation. They were able to support all processes relating to evaluation and fact- and context-check for Finland's development efforts and guide and connect evaluation teams to managers and staff within and outside the MFA. While their role in Quality Assurance for evaluation outputs was significant, their value in ensuring process quality was oftentimes more important. They provided a fair deal of troubleshooting, problem-solving, and gap-bridging in managing expectations of stakeholders, facilitating mutual understanding between the MFA and evaluation teams, and providing hands-on assistance to evaluation teams and their end products.

We consider that the Evaluation Management Teams introduced with EMS have usually worked well. They represented the principal planning and coordination mechanism at the level of individual evaluations between the MFA's Development Evaluation Unit, the Service Coordinator, and the evaluation team. Having all parties involved at the table greatly facilitated communication, coordination and the establishment of mutual consensus on the way forward. In a sense, these teams were a way to ensure that some direct engagement between commissioners and evaluators continued to happen, despite the introduction of Service Coordinators as intermediaries. At the same time, the presence of the Service Coordinator in all evaluation processes facilitated creating trust between the actors involved and provided for a mechanism of expectation management facilitation between the MFA's Development Evaluation Unit and the evaluation teams.

With respect to democratic commissioning, we note that the introduction of an intermediary role altered the nature of the commissioner–evaluator relationship. As discussed in Chapter 1 of this book (Stame & Bustelo, 2027), democratic commissioning requires engagement between commissioners and evaluators. While the EMS model increased efficiency, the outsourcing, as well as the introduction of a Service Coordinator, added a layer between these actors. Yet, democratic commissioning also benefits from trust, management of expectations and support provided to the joint efforts of the commissioners and evaluators, and the added layer of the Service Coordinator did its part to add these values.

Standardized procedures

In 2023, the MFA reviewed the use and utility of its centralized evaluations (MFA, 2023). That evaluation found that the EMS process had contributed to managing evaluations in a standardized manner and with consistent quality.

Our experience suggests that while evaluation management standardization increased, it did not lead to bureaucratization. Standardized criteria for team selection, evaluation milestones, and reporting formats improved consistency while not restricting adaptability in evaluation design. The two-step contracting approach described below allowed for participatory finalization of the Terms of Reference (ToR), which contributed to evaluations being methodologically sound and strategically aligned.

One area in which standardized criteria were introduced under EMS was the selection of team leaders and different categories of team members. For each team member role (team leader, senior evaluator, evaluator, and emerging evaluator), minimum criteria were defined, and the consortium companies and the Service Coordinator made use of them when searching for candidates. Also, as part of the selection process, team leader candidates were interviewed by the Service Coordinator and asked to draft a brief methodological note.

Other areas with standardization under EMS included the EMS process flow with set milestones, as well as report formats, writing instructions, toolkits and vocabularies. We found these useful in managing and steering the evaluation, where all stakeholders involved could “sit in the process and enjoy the ride”. We also found these useful when leading evaluations and serving as team members because the form and boundaries of the work were set, and we could focus on substantive matters at hand.

Under EMS, evaluation contracts were split into two “Service Orders”. The first (“Service Order 1”) included a brief contract for the evaluation team leader that allowed for some preparatory activities as well as a limited number of working days for the Service Coordinator. This was followed by the second contract (“Service Order 2”) that covered all other evaluation activities.

We find that this approach has two advantages. First, it allows the team leader and Service Coordinator to help finalize the evaluation ToR. Usually, the MFA’s Development Evaluation Unit would provide a draft ToR and then ask for feedback and additional input, especially on the methods section, which was usually left open to be developed or adapted by the team leader and the Service Coordinator, in close alignment with the Development Evaluation Unit. We consider this a good practice because it ensures that the evaluation approach was realistic and in line with the skills and experiences of the evaluation team. In line with this, a recent MFA funding decision advocated that the “*service-based approach allows a more flexible use of evaluation expertise, also during the planning phases*” (MFA databank, n.d./a). In several instances, the interactions during this first phase had contributed to further clarifying the purpose and the scope of the evaluation and to developing fewer but more meaningful evaluation questions, which we also consider good practice. In our view, this participatory approach to finalizing ToRs reduced the

need for later course corrections during (or after) the inception phase. While we experienced ourselves that such later corrections were possible, having a ToR in place that was fully owned by those managing and implementing the evaluation was more efficient. It also had positive effects on managing expectations of evaluation stakeholders and increased transparency, as only the ToR (not the inception report) was later published as part of the evaluation report.

Second, the first Service Order also provided capacity and time for selecting the remaining evaluation team members, which we found very useful. We could, for example, suggest candidates from our own networks and influence the selection process. While the MFA's Development Evaluation Unit retained final decision-making authority on team selection, our recommendations were often followed. In our view, this approach increased the probability of having the "right" evaluators on board, not only in terms of skills and expertise but also in terms of effective and efficient teamwork.

A 2023 evaluation of the use and utility of centralized evaluations (MFA, 2023) suggested a need to better balance evaluation expertise with understanding and knowledge of the MFA, which the reviewers found missing in some of the very senior international consultants selected in the past. To this end, we considered it helpful that one consortium member was a Finnish company. Also, the supporting role of the (Finnish) Service Coordinator was helpful to ensure sufficient understanding of the MFA. Moreover, the standardized selection process for team leaders and team members allowed for a balanced approach in mixing experts with prior experience evaluating Finland's development policy and cooperation with new entrants and international, Finnish and local experts in the teams.

With respect to democratic commissioning, we find that the EMS-related standardization of the commissioning process has been useful overall. In our view, the model represented a good balance between providing additional structure without negatively impacting adaptability.

Knowledge management

Centralized evaluations are important knowledge instruments. In 2019, one evaluation was dedicated to knowledge management at the MFA (MFA, 2019). It found that centralized evaluations "*produce useful results information, and have successfully influenced decision-making at the MFA*". The same evaluation also found that MFA had not yet fully put to use the comprehensive body of knowledge centralized evaluations generated. It was recommended that the MFA invest in targeted syntheses of specific results information and knowledge from centralized evaluations to effectively inform learning and decision-making in development policy and cooperation at the MFA. The 2019 evaluation did, however, not assess the degree to which the EMS framework might have contributed to this status quo.

In 2023, a review of the utility of centralized evaluations (MFA, 2023) praised the EMS framework but also highlighted several challenges in knowledge management. A key issue identified was the limited transfer of ownership of evaluation results from the Development Evaluation Unit to operational MFA departments

responsible for policy implementation. The lack of a comprehensive, accessible knowledge repository further limited continuity across evaluations, forcing teams to repeatedly collect similar information. Addressing these gaps would improve the extent to which evaluations contribute to informed decision-making and democratic accountability in MFA policy processes.

Assessing EMS's impact on knowledge use remains challenging. However, while existing evaluations indicate that EMS improved process coordination, it did not resolve issues related to knowledge retention and accessibility. Strengthening systematic knowledge management mechanisms could enhance the long-term contribution of evaluations to MFA policy and decision-making.

One issue related to the management of information (rather than knowledge) was the lack of a comprehensive, ready-to-use database for standard information used in centralized evaluations before and during EMS. This has two complementary reasons. First, the MFA's project and archive management systems were and remained inadequate (MFA, 2019). Second, despite the Consortium's efforts, it was not possible to establish a comprehensive database as part of EMS. Because of this chapter's focus on EMS, we will discuss the second point further.

In the centralized evaluations in which we were involved, the identification and collection of documented information usually represented an activity that consumed significant amounts of time and capacity. This usually concerned a wide range of documents, from project plans and reports to policies and data on budgets, expenses, and human resources. In a first step, the Evaluation Manager usually tries to recover this information from the MFA's information management systems and from public sources. In most cases, this left important gaps, and it thus became necessary to ask MFA staff in involved departments and units to somehow retrieve any missing information. These activities required considerable effort by MFA staff and delayed the evaluation – or required the evaluation team to start their work with incomplete information.

Because evaluation contracts stipulated that evaluation-related information had to be deleted after an assignment was complete, we sometimes had to ask the same people for the same information we had already required and used in a previous evaluation. On the level of the consortium, this concerned information repositories of earlier centralized evaluations managed by one of the companies that were not accessible to the evaluation teams anymore, apart from some limited cases where the consortium negotiated access to the earlier evaluations' selected files to the experts. After repeatedly going through this procedure, we felt that much value could be added by investing in establishing and maintaining a comprehensive repository of standard information for future centralized evaluations.

With respect to democratic commissioning, we find that the EMS has not been able to improve knowledge management. The transfer of ownership of evaluation results from the Development Evaluation Unit to operational MFA departments remained weak, and there were no significant synergies in terms of access to the same information required by subsequent evaluations. We also note that these weaknesses are related to data management systems and policies at the MFA more generally, which could not be changed by the EMS but within which it operated.

Participation of emerging evaluators

Emerging evaluators were usually part of evaluation teams under EMS. A recent evaluation on the use and utility of centralized evaluations found that “The inclusion of Emerging Evaluator posts in the teams is most welcome: it gives Finnish young professionals a chance to build experience in these evaluations”. In our experience, the on-the-job learning of emerging evaluators has likely become more effective in EMS 2 because the same individuals could participate in different evaluations. This was part of a capacity development plan formulated as part of EMS.

Based on two surveys conducted by NIRAS Finland Oy in October 2022 and May 2023, all emerging evaluators felt that the level of tasks was adequate for learning. It was noted that the tasks needed to be defined clearly, be sufficiently challenging, and ideally include both an independent contribution to the evaluation (i.e., overseeing a specific part) and supporting more senior team members.

For us, one lesson learned is that the role of emerging evaluators needs to be carefully balanced between more or less independent and demanding tasks on a case-by-case basis. Clearly, emerging evaluators learn more in more demanding roles, provided they can fulfill them well. We had very positive experiences with motivated and capable emerging evaluators, but we also remember cases when intense coaching and unbudgeted additional work were required for achieving satisfactory evaluation end products of emerging evaluators.

In several cases, the emerging evaluators were also the Project Manager from one of the consortium companies. We consider this double role beneficial because it further supports learning. For example, it offered them the opportunity to participate in the planning of the evaluation from the very beginning.

Overall, the inclusion of emerging evaluators in EMS teams has provided a structured pathway for early-career professionals to enter the evaluation field. When we began writing this chapter, five emerging evaluators had successfully transitioned into the “Evaluator” category, with two already engaged in MFA evaluations in an independent capacity.

Conclusions

Between 2016 and 2024, the EMS framework introduced a structured, professionalized approach to managing centralized evaluations at the MFA of Finland. By outsourcing selected evaluation management functions, EMS allowed for greater consistency, access to evaluation expertise, and relief from internal staffing constraints. Standardized evaluation management processes, transparent qualification criteria, and a flexible two-step contracting procedure supported evaluation quality and strategic alignment.

However, the introduction of EMS also redefined the role of MFA commissioners, shifting them toward strategic oversight while introducing an intermediary structure that managed core aspects of evaluations. This increased distance between commissioners and evaluators impacted the agency of the MFA’s Development Evaluation Unit along the commissioning process, while it also served to provide

further trust, expectation management and sense of continuity in the commissioners and evaluators' relations.

The EMS-related standardization of the commissioning process has been useful overall, with a good balance between providing additional structure without negatively impacting adaptability.

The EMS has not been able to improve evaluation knowledge management, which remains a systemic issue at the MFA, with institutional systems and policies beyond the scope of the EMS.

The EMS contributed positively to evaluator capacity-building, notably through the emerging evaluators program, and realized process synergies across evaluations through the role of the Service Coordinator.

Looking ahead, the key question is how commissioning models like EMS can continue to evolve to strengthen democratic commissioning. If public agencies increasingly adopt structured outsourcing models, future iterations of EMS – or similar frameworks elsewhere – must ensure that commissioners' roles and agencies are further strengthened, that the commissioning process is efficient and effective, and that it is supported by the systems within which it takes place.

Notes

- 1 This section draws on the Development Evaluation Norm of the Ministry for Foreign Affairs of Finland (MFA, 2015).
- 2 As of March 2024, other responsibilities of the Development Evaluation Unit were evaluation guidance concerning development policy and development cooperation; evaluation capacity building and training; disseminating evaluation results; general development of the evaluation of development policy and development cooperation; and participation in international evaluation networks and joint evaluations (<https://um.fi/development-evaluation>, accessed March 2024).
- 3 The related tender document “Finland – Research and development consultancy services – Framework Agreement for Centralised Evaluations (FACE)” was published on the TED website of the European Union: <https://ted.europa.eu/en/notice/-/detail/233846-2024>, accessed June 2026.
- 4 The company information was drawn from the two firms' websites (www.particip.de/en/companyoverview/companyfacts/, www.niras.com/offices/finland/, www.niras.com/about-niras/, accessed March 2024).
- 5 Since EMS 1 began in 2016, the Service Coordinator only changed once from the beginning of 2022, when one of us took that job.

References

- Ministry for Foreign Affairs of Finland (MFA). 2015. *Development Evaluation Norm of the Ministry for Foreign Affairs (MFA)*. Norm 1/2015. Unofficial English translation from Finnish. Helsinki: Ministry for Foreign Affairs of Finland. Accessed May 2024. https://um.fi/documents/35732/0/Development+evaluation+norm_MFA_2015+%281%29.pdf/dfdf4e72-a5b6-8628-d70b-a32f257f0c2c?t=1624960477459.
- Ministry for Foreign Affairs of Finland (MFA). 2019. *Evaluation: How Do We Learn, Manage and Make Decisions in Finland's Development Policy and Cooperation*.

- Evaluation on Finland's Development Policy and Cooperation 2019/3. Helsinki: Ministry for Foreign Affairs of Finland. Accessed May 2024. https://um.fi/publications/-/asset_publisher/TVOLgBmLyZvu/content/evaluointi-tietojohdamisesta-miten-opimme-johdammme-ja-teemme-paatoksia-suomen-kehityspolitiikassa-ja-yhteistyossa/384998.
- Ministry for Foreign Affairs of Finland (MFA). 2023. *Review of the Use and Utility of Centralised Evaluations*. Evaluation on Finland's Development Policy and Cooperation 2023/2. Helsinki: Ministry for Foreign Affairs of Finland. Accessed May 2024. https://um.fi/publications/-/asset_publisher/TVOLgBmLyZvu/content/tarkastelu-keskitettyjen-evaluointien-kaytosta-ja-hyodyllisyydesta/384998.
- Ministry for Foreign Affairs of Finland Databank. n.d.-a. *Databank on Finland's Development Cooperation: Framework Arrangement for Development Evaluation Services*. Project FI-3-2020-2019000731, activities from 01/06/2020 to 31/12/2024. Accessed May 2024. <https://openaid.fi/en/project/FI-3-2020-2019000731>.
- Ministry for Foreign Affairs of Finland Databank. n.d.-b. *Databank on Finland's Development Cooperation: Support Service for Centralised Evaluations*. Project FI-3-2024-2024240234, activities from 01/01/2024 to 31/12/2024. Accessed May 2024. <https://openaid.fi/en/project/FI-3-2024-2024240234>.
- Stame, Nicoletta and María Bustelo. 2027. "Commissioning Matters: Considerations and Challenges for Evaluation Commissioners Who Intend to Promote Democracy." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.

11 The art of commissioning

Lessons from commissioner–evaluator collaboration¹

Ida Lindkvist and Rob Lloyd

Dispute about commissioner competence

During the spring of 2024, Ida Lindkvist, one of the authors of this chapter, was involved in a debate about evaluation in *Morgenbladet*,² one of Norway’s oldest newspapers specialising in politics, culture and science. The discussion was sparked by an op-ed written by Tore Wig, a professor of political science at the University of Oslo. Wig argued that Norwegian evaluations were of low quality and claimed that consultants were “charlatans” producing quasi-scientific reports that served their own interests. In Wig’s view, the only way forward was for researchers to take over and conduct randomised controlled trials of development assistance programmes.

In response, Ida Lindkvist and Tove Sagmo, both employed in the independent Evaluation Department³ in Norad and working as commissioners at the time, published a rebuttal, and several weeks of debate followed. All agreed low evaluation quality was a problem; instead, the crux of the disagreement was whether government agencies needed higher evaluation capacity, including competent evaluation commissioners, to ensure useful high-quality evaluations.

Wig’s claims drew on the Evaluation Department’s meta-evaluations (2017; 2020b; 2021)⁴ of decentralised evaluations, which consistently raised concerns about poor methodological quality, lack of ethical considerations, and over-ambitious scope. In contrast, the OECD’s (2019) peer review commended the centralised evaluation function as a leader in its field. Both types of evaluations were commissioned by Norad staff, but only the Evaluation Department had a separate mandate and staff with specialised evaluation competence, allowing it to decide independently what, when, and how to evaluate.

Lindkvist and Sagmo, who commissioned many of these meta-evaluations, argued that the antidote to Wig’s “charlatans” was to strengthen evaluation functions and ensure that the staff who conducted and commissioned evaluations had the right competencies, including knowledge-brokering skills, analytical and methodological skills, project management skills and a deep knowledge of the organisation in which evaluations would be used to ensure evaluation relevance.

In this chapter, we, the authors; Ida from the commissioning body; and Rob from the evaluation team argue that competent commissioning is a critical driver

of evaluation quality and use in government. We illustrate this through four evaluations commissioned by Norad’s Evaluation Department between 2014 and 2024, in which we were directly involved as commissioner representative or as evaluation team lead.

Drawing on our dual perspectives, we examine the role of commissioning from multiple angles and show how it is shaped by broader organisational evaluation capacity. The chapter is structured in three parts: first, a brief presentation of the evaluations; second, an exploration of the interplay between commissioner and evaluation team from the perspective of the commissioner; and finally, concluding reflections with lessons for other organisations who wish to strengthen the quality of commissioning.

The evaluation series

The evaluation series we will discuss began in 2014 with an evaluation titled *Can We Demonstrate the Difference Norwegian Aid Makes?* This first evaluation aimed to identify the systemic barriers preventing partners/grants, the primary source of results data in the Norwegian system, from effectively measuring results.⁵ The evaluation found that while existing systems and policies did not prevent capturing results, existing practices were not conducive to a results focus. The evaluation was part of the broader “When will we ever learn” debate initiated by the Center for Global Development (Savedoff et al. 2005).

The second evaluation, *Evaluation of the Norwegian Aid Administration’s Practice of Results-Based Management* (Evaluation Department 2018), shifted focus from individual grants and partner results, the central theme of the 2014 evaluation, to the broader system of results-based management (RBM) within the Norwegian aid administration. The main difference from its predecessor was that it focused on the use of quality results data in RBM, and management of portfolios in particular.

Responding to the 2018 evaluation’s recommendations to improve portfolio management, and the Ministry of Foreign Affairs’ commitment to act on this, the third evaluation, *Evaluation of the Norwegian Aid Administration’s Approach to Portfolio Management* (Evaluation Department 2020a), explored how portfolio management was being implemented and how it could be further strengthened.

The final evaluation discussed here, *Evaluation of Norad’s Use of Knowledge in Portfolio Management* (Evaluation Department 2024), assessed the use of knowledge/evidence/results data in portfolio management. It picked up from the 2020a evaluation, tracking recommendations and exploring in more depth whether and how knowledge was being used in the design and implementation of portfolios. This evaluation also adopted a developmental approach, collaborating with the administration in embedding knowledge practices within portfolio management through the evaluation process (2024, 7).

Viewed collectively, we argue that these four evaluations appear to have significantly influenced and shaped how the Norwegian aid administration approached results management.

The foundational 2014 evaluation sparked public debate⁶ about results measurement and the use of evidence in decision-making and set the stage for subsequent evaluations and reforms, and also appeared to influence the critique of the Norwegian aid administration that Wig later was a part of. The 2015–2016 budget proposal emphasised the need for strengthened quality of results information (see note 5).

The RBM evaluation from 2018 recommended clarifying RBM's dual purpose of reporting and learning, adopting a portfolio-based approach with clear objectives and programme logic and conducting evaluations at the portfolio level. The Norwegian aid administration's response included a commitment to improving portfolio management through pilot portfolios and gradual implementation of the recommended changes.

The 2020 evaluation of portfolio management informed Norad-initiated reforms to manage its grants effectively and efficiently at the level of the portfolio. Most notably, the portfolio management standards recommended in the evaluation and accompanying evaluation brief were adopted, along with the recruitment of portfolio coordinators.

The 2024 evaluation found encouraging signs of progress, including the development of theories of change and knowledge plans across portfolios. It marked a shift towards stronger leadership support for the use of knowledge and dedicated training and guidelines. However, challenges such as limited resources and gaps in monitoring, evaluation and learning (MEL) at the portfolio level persisted. Even before the evaluation was finalised, Norad began acting on key findings, integrating tools into guidance and revising the portfolio theories of change.

The interplay between the commissioner and the evaluation team

In Figure 11.1, we have summed up how the Evaluation Department as commissioner collaborated with the evaluation team during the evaluation series we have discussed. As can be seen from the flowchart, commissioners have an important role throughout the evaluation process, both before and after. For some evaluations, the process in Figure 11.1 was repeated multiple times as the Evaluation Department continued to follow up on the same issues over a ten-year period. For results-based management at the grant and portfolio levels, the evaluations probably contributed positively to improvements in management of grants. It is less certain that they improved the use of evaluations at the decentralised level as part of an RBM system. This is perhaps not surprising, as evidenced in the above-mentioned OECD peer review, the organisational capacity to do decentralised evaluations was limited.

To emphasise the importance of the commissioner's voice, 'we' in the sections below refers to the commissioner, while reflections from the evaluation team are reflected in separate boxes.

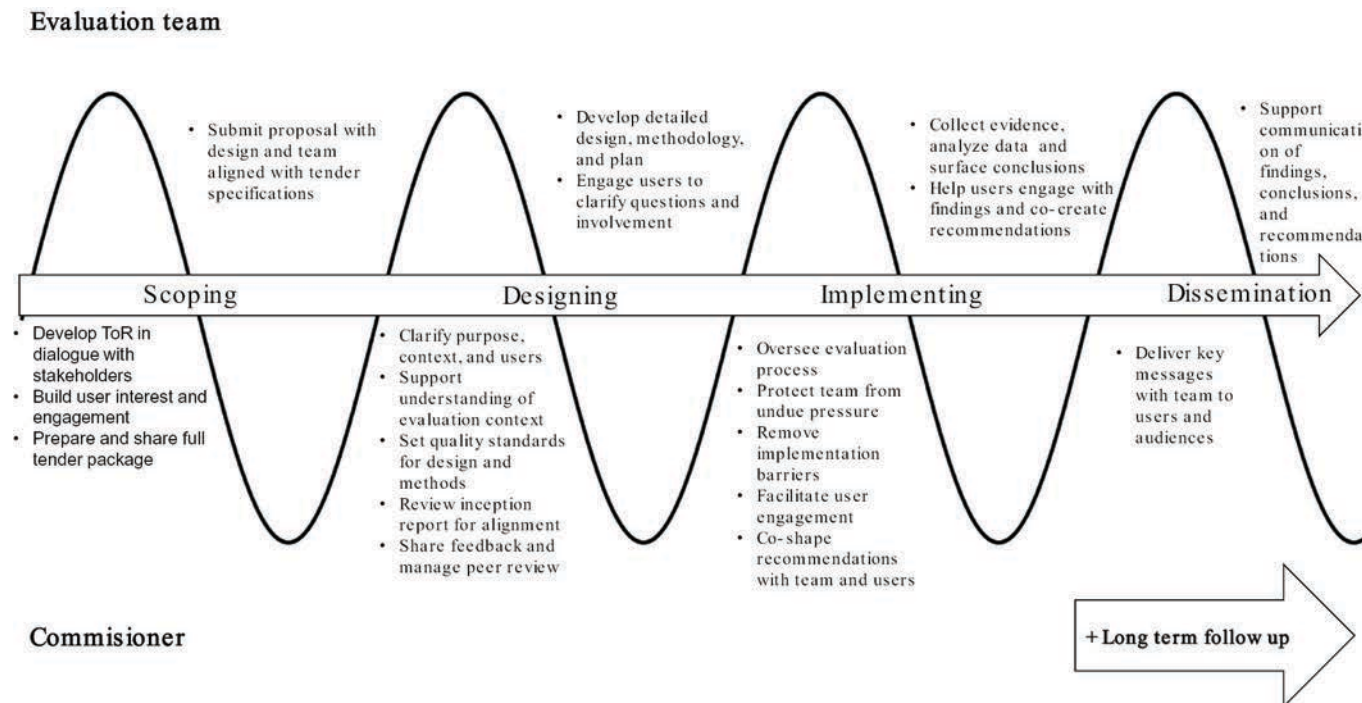


Figure 11.1 Collaboration between evaluation team and commissioner.

Scoping

Scoping evaluations involves developing the Terms of Reference (ToRs), building user interest and engagement and preparing a full tender package for procurement. Deciding what to evaluate, when and for what purpose is key.

The commissioner's role was not just to oversee evaluations but to actively diagnose challenges and shape the reform agenda, thereby demonstrating agency and leadership. This would not have been possible without the deep organisational knowledge we had developed over the course of the evaluation series. Acting as knowledge brokers, we identified and translated knowledge in ways that shaped the evaluations. Once the topic of results-based management was identified, we engaged stakeholders across the aid administration and among funded partners, such as NGOs, to capture how challenges were experienced in practice. To situate the Norwegian context within broader debates, we also consulted external experts—for example, John Mayne advised us on the results-based management evaluation during the development of the ToRs and beyond. Finally, we verified our approach against national standards by consulting the Norwegian Agency for Public and Financial Management (DFØ), whose guidelines served as a benchmark.

Each evaluation built on the last, with recommendations from one shaping the focus of the next. The first evaluation addressed the lack of documented results. The RBM evaluation tackled the limited use of results data in decision-making. The portfolio management evaluation focused on the absence of coherent structures to apply results in a strategic way. These topics were deliberately chosen and then socialised within the administration. Strong leadership was essential in both setting this agenda and bringing others along. The series of evaluations also benefited from retention of experienced staff in the commissioning function.⁷ Continuity in staff in the Evaluation Department facilitated the accumulation of institutional knowledge and established relationships that were important in maintaining the series' momentum. More specifically, some of the same staff developed ToRs, identified and communicated with key stakeholders throughout the above-mentioned evaluation processes.

Over the course of the evaluation series, we deliberately adapted the purpose of each evaluation to reflect the organisation's evolving context and needs, shifting from emphasising accountability to learning to maximise impact. The evaluation, *Can We Demonstrate the Difference that Norwegian Aid Makes?* (2014) was deliberately hard-hitting, built around hypotheses explaining why results data was not being gathered. The subsequent RBM evaluation continued this accountability focus. As openness to the results agenda increased, we recalibrated. The portfolio management evaluation leaned more towards learning: portfolios volunteered as case studies and benefited from support in developing theories of change and exploring how to operationalise portfolio management. By the time we conducted the evaluation of knowledge use in portfolio management, the approach had evolved into a close partnership with the Knowledge Department.

As commissioners, we also had a close relationship with Norad's procurement staff and worked extensively with procurement specialists to ensure an effective

procurement process. We also scoped the evaluations more tightly and limited the number of evaluation questions, as a wide range of questions lower the agency of the commissioner and can lead evaluators to cherry-pick evaluation questions according to their interests, rather than the organisation's needs.

Reflections from the evaluation team

In our experience, evaluation ToRs can often be overloaded with questions, frequently the result of compromises between multiple stakeholders. In contrast, Norad's ToRs stood out for their clarity and focus. They were tightly scoped, sometimes containing no more than three core questions, which gave the evaluation a clear sense of purpose from the outset. This focus sharpened our analysis and made it harder to hide behind a long list of loosely connected questions. It also raised the bar: with fewer questions, there was nowhere to hide; we had to answer each one clearly, robustly, and with confidence. This kind of discipline was refreshing and, ultimately, a key factor in the quality and impact of the evaluations.

Design

While the ToR set a clear vision and direction, including mentioning relevant methods, the tendering process left space for bidders to propose fit-for-purpose methodologies and innovative approaches. After identifying knowledge gaps, engaging users, and defining key questions, we discussed possible methods internally and, where needed, consulted external experts to validate options, sometimes highlighting this competence in the call for proposals. We also avoided being overly prescriptive, encouraging methodological development during the inception phase and explicitly welcoming revisions to the ToR if they improved alignment with purpose.

Reflections from the evaluation team

The ToRs for these evaluations struck the right balance. They were clear about the purpose and key questions, and while they provided initial direction on possible approaches, they didn't prescribe the methodology. This flexibility allowed us to design evaluations that were genuinely fit for purpose and to introduce new and innovative methods along the way.

One of the most impactful innovations was co-creating recommendations, which we first introduced in the RBM evaluation. This approach was carried forward into all subsequent evaluations and significantly improved

engagement with the “so what” of our findings. In the final evaluation, we took this further, working in close partnership with the Norad Knowledge Department to co-develop the recommendations. It felt like a real shift towards the organisation not just receiving findings but actively owning and using them.

Implementation

During evaluation implementation, the Evaluation Department played a dual role: continuously monitoring progress internally to identify and address emerging issues promptly and safeguarding the evaluation team’s independence by ensuring they could engage openly with stakeholders and ask all necessary questions.

A key milestone during the implementation phase was going through the inception report and assessing whether it responded to the ToRs and the tender proposal. This included assessing whether it adhered to OECD evaluation quality standards, the department’s guidelines and standards for the specific methods employed. We also shared the inception report with stakeholders and frequently external experts, either on the topic or method. We compiled all comments, including our own, in a table and shared it with the evaluation team and asked them to explain how they dealt with the comments in a revised inception report. We emphasised that it was up to the evaluation team to determine how to respond to stakeholder comments. The same process was repeated for the draft final report.

We held regular internal check-ins with the team to monitor how the evaluation was unfolding and to address blockages or concerns before they escalated and risked derailing the process. This proactive oversight helped maintain momentum and kept the evaluation on course. Equally important was our role in safeguarding the evaluation team’s ability to conduct candid inquiries. Through our relationships across the organisation, we received feedback and concerns both formally and informally. By addressing these issues or clarifying misunderstandings early, we prevented blockages and ensured the evaluation team retained access to people and information, crucial for maintaining the integrity and depth of the evaluation.

Reflections for the evaluation team

Regular check-ins with the Evaluation Department were crucial not only for coordinating the evaluations but also for discussing any emerging blockages or challenges. The Evaluation Department’s strong relationships across the organisation meant they had their finger on the pulse; if anyone had concerns or was unhappy with the evaluation process, they would often come to them first. These check-ins provided a safe space to surface and address such

issues early, helping to maintain trust and keep the evaluation on track. This ongoing dialogue between the evaluation team and the department was key to navigating challenges and preserving the independence and integrity of the evaluation.

In later evaluations, we introduced interim deliverables to stay closely aligned with the evaluation team’s thinking and progress. The evaluation of knowledge use in portfolio management (2024), which had a strong learning focus, required particularly close collaboration with the Knowledge Department. When decision points arose where early findings, such as assessments of portfolio theories of change and knowledge plans, could be most useful, we asked the evaluation team to share results at those moments when feasible.

We also participated in portfolio coordinators meetings. At these meetings we shared progress updates, and when new coordinators stepped in due to staff mobility, we held separate briefings to ensure continuity of understanding.

Reflections from the evaluation team

Introducing interim deliverables was a valuable adaptation that brought both discipline and flexibility to the evaluation process. These milestones encouraged us to work iteratively, cycling through data collection and analysis rather than waiting until the end—helping to surface insights more gradually and thoughtfully. They also served as important moments for dialogue with the Evaluation Department and other stakeholders, providing opportunities to test emerging interpretations and adjust course where needed. This ongoing engagement helped to build shared understanding and mitigate the risk of disconnects between the evaluation team’s evolving findings and stakeholder expectations. Importantly, the interim deliverables helped socialise preliminary insights across the organisation, fostering a collaborative atmosphere and reducing the potential for surprises at the final reporting stage. While not a perfect safeguard, they created a space for reflection and dialogue that ultimately strengthened the quality and usefulness of the evaluations.

Dissemination

For those evaluations that had a stronger accountability focus, we invested in public dissemination and stimulating debate on the evaluation findings both within the Norwegian Aid Administration and in the public. The evaluation director’s direct engagement with senior leadership helped us prevent findings from being dismissed. Bureaucracies often have a natural reflex to deflect or dilute hard messages; it took

strong leadership to hold the line, create space for difficult conversations, and push the system to respond rather than retreat. Staff would sometimes also write op-eds and, if relevant, share findings with the press.

In general, for most evaluations, we organised an evaluation launch seminar, inviting high-profile stakeholders, including policymakers and experts within the field. We prepared panellists in advance, ensuring they were familiar with the evaluation findings, and, when feasible, streamed the seminar online to maximise outreach.

Reflections from the evaluation team

From the outset, the Evaluation Department was proactive in communicating the findings of the evaluations, particularly those with a strong accountability focus. Their efforts often led to a level of external interest that is rare in our experience. Several reports generated coverage in national print and broadcast media, and public events were well-attended, frequently drawing senior MFA officials, including the Minister for International Development.

This visibility helped raise the profile of the evaluations and, we believe, contributed to shaping the broader discourse around evidence and accountability in Norwegian aid. While public attention doesn't guarantee internal change, it did create pressure and momentum that made it harder for the findings to be ignored. The willingness of the Evaluation Department to engage publicly and to take ownership of sometimes uncomfortable messages was a key factor in the evaluations' influence.

For evaluations with a strong learning focus, the department invested in targeted communication with strategic users. Preliminary findings were presented to different groups, such as the top management group and portfolio coordinators, tailored to their needs. Even when the results were critical of current practices, walking user groups through the findings and inviting discussion of their validity made the conclusions easier to accept.

Reflections from the evaluation team

Having the opportunity to work with the same organisation over nearly a decade is unusual and a real privilege. One of the biggest benefits was the ability to build relationships and, gradually, trust with both the Evaluation Department and key internal stakeholders. That trust wasn't automatic; it was earned over time as people saw how we operated: independent but constructive, focused on rigour but also on usefulness.

As trust grew, the nature of the engagement changed. Stakeholders became more open, more reflective, and more willing to lean into difficult findings. It allowed us to land messages that might otherwise have been met with defensiveness. Dissemination events, in particular, became more than just presentations; they were moments of real dialogue, where evidence could be interrogated and ideas for change could take root.

Just as important was the relationship we built with the Evaluation Department itself. Their trust in us gave us the space to innovate, to challenge, and to co-create. It meant we could work together as true partners, holding a mirror up to the organisation, while always keeping the focus on improvement. That partnership was the foundation for everything we achieved.

We used evaluation-specific learning briefs strategically to explore issues that could support the implementation of recommendations. For several evaluations, we commissioned, and sometimes co-authored, briefs that expanded on selected topics rather than serving as condensed versions of the full reports.

Reflections from the evaluation team

The Learning Briefs were a valuable way to explore specific issues raised in the evaluations in greater depth. Developed in collaboration with the Evaluation Department, we selected topics where we felt additional insight or practical guidance could add real value for the organisation.

For example, the brief “*10 Steps to Creating and Sustaining a Learning Culture*” built on a key recommendation from the RBM evaluation, namely the need to focus more intentionally on fostering a culture of learning. The brief translated this recommendation into actionable guidance that teams could use. Similarly, following the *Knowledge-Based Portfolio Management* evaluation, we developed a brief on how to approach MEL at the portfolio level, addressing a critical gap identified in that evaluation.

These briefs gave us an opportunity to complement the evaluations with practical tools and guidance, helping to bridge the gap between insight and implementation.

A way forward for commissioners

Stepping back from this series of evaluations, several factors stand out in explaining why it had the impact it did.

First, partnership mattered. The collaboration between the commissioner and the evaluation team was built on mutual trust and respect. Critical feedback, for

example, on the quality of an inception report, was received with curiosity and constructive debate rather than defensiveness. This dynamic strengthened the process instead of derailing it.

Second, the evaluations gained traction because both sides respected their distinct roles while working in ways that made those roles mutually reinforcing (as illustrated in Figure 11.1). Organisational change is often slow, but the evaluations had influence because of this balance.

Third, longevity counted. Continuity of staff on both sides meant that each evaluation could build on the last, without the need to start from scratch every time.

Finally, the relationship was possible only because the organisation had sufficient evaluation capacity. The Ministry of Foreign Affairs and later the Ministry of Climate and Environment gave the department freedom to commission evaluations, backed by stable funding and institutional support. The department had access to archives, a professional evaluation staff (at its peak 11 specialists, later 9), and deep experience from development programming, academia, and other evaluation agencies.

From this experience, we draw lessons for other commissioners:

Lesson 1: Exercise agency within the limits of independence. Commissioners should take responsibility for shaping ToRs and stay engaged throughout the process. But agency must be balanced with institutional independence, particularly where accountability is a strong purpose, to avoid undermining credibility (see for example Lindkvist and Kennedy-Chouane, 2024).

Lesson 2: Draw on external expertise. No commissioner can master all evaluation topics and methods. Independent advice can sharpen ToRs and strengthen dialogue with the evaluation team.

Lesson 3: Support collaboration. Mechanisms such as interim deliverables and framework agreements create space for productive interaction and help build trust over time.

Lesson 4: Share responsibility for communication. Effective communication is not just the evaluation team's job. Commissioners play a vital role in designing and sustaining long-term communication strategies within the organisation.

Lesson 5: Build on previous evaluations. To drive organisational change, commissioners must be strategic and patient. Linking new evaluations to earlier findings reinforces learning, strengthens continuity, and increases the chance that change will take root.

Returning to where we began, the debate with Wig highlights a blind spot not only among researchers but perhaps also among evaluators: commissioners matter. They are the crucial bridge between an evaluation team and the organisation that must act on the findings. Commissioners should not be viewed as passive administrators of contracts but as active agents who shape the success of evaluations and their uptake. Recognising commissioning as a discipline in its own right, and investing in the capacity of commissioners, is essential if evaluations are to achieve their full potential.

Notes

- 1 The views, thoughts, and opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of their current or previous affiliated institutions.
- 2 Interested readers are referred to *Morgenbladet* to read the entire debate (Lindkvist and Sagmo 2024c; 2024a; 2024b; Wig 2024c; 2024a; 2024b). Others also engaged in the debate (Eggen 2024; Høydal 2024; Reinertsen 2024).
- 3 From 1978 till 2004, it was organisationally situated in the Ministry of Foreign Affairs, while from 2004 till 2024, it was organisationally situated in Norad but reported to the MFA and, in later years, also to the Ministry of Climate and the Environment.
- 4 Wig only referred to ‘the Norad report’ that documented low quality; however, this report must have been the meta-evaluations, as no other reports systematically assessed evaluation quality.
- 5 Evaluation Department (2014), p. 5.
- 6 Individuals from both the evaluation team and the commissioning body participated in this debate. See, for example Lindkvist and Dixon (2014).
- 7 Staff involved had a combination of analytical skills and practical experience with development assistance, as they included staff with both research and evaluation backgrounds and with practical experience from development assistance.

References

- Eggen, Øyvind. 2024. ‘Bistandsevaluering Krever Forskere Som Forsøker å Forstå: Tore Wig Uttrykker Ikke Faglig Uenighet, Men Uvitenhet’. *Morgenbladet* (Oslo), May 10. www.morgenbladet.no/ideer/bistandsevaluering-krever-forskere-som-forsoker-a-forsta/9409317.
- Evaluation Department. 2014. *Can We Demonstrate the Difference That Norwegian Aid Makes?* Oslo: Norad.
- Evaluation Department. 2017. *The Quality of Reviews and Decentralised Evaluations in Norwegian Development Cooperation*. Evaluation No. 1/17. Oslo: Norad.
- Evaluation Department. 2018. *Evaluation of the Norwegian Aid Administration's Practice of Results-Based Management*. Oslo: Norad. www.norad.no/en/publications/2018/evaluation-of-the-norwegian-aid-administrations-practice-of-results-based-management/.
- Evaluation Department. 2020a. *Evaluation of the Norwegian Aid Administration's Approach to Portfolio Management*. Oslo: Norad. www.norad.no/publikasjoner/2020/evaluation-of-the-norwegian-aid-administrations-approach-to-portfolio-management/.
- Evaluation Department. 2020b. *Quality Assessment of Decentralised Evaluations in Norwegian Development Cooperation (2018–2019)*. Evaluation No. 6/20. Oslo: Norad.
- Evaluation Department. 2021. *Quality Assessment of Decentralised Evaluations in Norwegian Development Cooperation (2019–2020)*. Evaluation No. 3/21. Oslo: Norad.
- Evaluation Department. 2024. *Evaluation of Norad's Use of Knowledge in Portfolio Management*. Oslo: Norad. www.norad.no/publikasjoner/2024/evaluation-of-norads-use-of-knowledge-in-portfolio-management/.
- Høydal, Øyunn Syrstad. 2024. ‘Wig Og Norad Roper Fra Hver Sin Tue Det Er Viktigere at Framtidens Byråkrater Og Forskere Er i Stand Til å Tenke Kritisk Rundt Evaluering Enn at de Drilles i å Gjennomføre Eller Bestille Evalueringer På «den Rette Måten»’. *Morgenbladet* (Oslo), May 24. www.morgenbladet.no/ideer/Wig-og-norad-roper-fra-hver-sin-tue/8914874.

- Lindkvist, Ida, and Megan Kennedy-Chouane. 2024. *Independence: When, Why and How It Matters for Evaluation Credibility*. EVAL Reflection Note Series No. 2. Oslo: Evaluation Department, Norad.
- Lindkvist, Ida, and Vibecke Dixon. 2014. 'To "Feel Good", or to "Do Good"? Why We Need Institutional Changes to Ensure a Results Focus in Norwegian Development Assistance'. *Journal of Development Effectiveness* 6 (4): 350–60.
- Lindkvist, Ida, and Tove Sagmo. 2024a. 'Bistand: Vi Trenger Forskere Som Kan Evaluere Evaluering Kan Ikke Reduseres Til Én Metode, like Lite Som Forskning Kan Det.' *Morgenbladet* (Oslo), May 10. www.morgenbladet.no/ideer/bistand-vi-trenger-forskere-som-kan-evaluere/9409860?signedInSuccess=true&_cb=070e3035-a2aa-4598-9ff0-5912d107b243&purchasedSuccess=true.
- Lindkvist, Ida, and Tove Sagmo. 2024b. 'Evalueringsslaget Bør Brukes Der Tore Wig Vil Droppe All Kunnskap Fra Evalueringsslaget, Mener vi at Man Bør Bruke Denne Til å Øke Kvaliteten På Evalueringer.' *Morgenbladet* (Oslo), May 24.
- Lindkvist, Ida, and Tove Sagmo. 2024c. *Medisinen Mot Wigs Sjarlataner Evaluering Er et Eget Fagfelt, Men Det Synes Ikke Særlig Kjent Blant Norske Forskere*. *Morgenbladet* (Oslo), April 26. www.morgenbladet.no/ideer/medisinen-mot-wigs-sjarlataner/9408394.
- OECD/DAC. 2019. *Peer Reviews: Norway*. OECD Development Co-Operation Peer Reviews. Paris: OECD Publishing. <https://doi.org/10.1787/75084277-en>.
- Reinertsen, Hilde. 2024. 'Politisk Vingeklipping Av Bistandsevalueringens Integritet: Bistandsevalueringens Største Utfordring i Dag Er Ikke Manglende Vitenskapelighet, Men Politisk Arroganse Og Organisatorisk Vingeklipping.' *Morgenbladet* (Oslo), May 31. www.morgenbladet.no/ideer/politisk-vingeklipping-av-bistandsevalueringens-integritet/9413082.
- Savedoff, William D., Ruth Levine, and Nancy Birdsall. 2005. 'When Will We Ever Learn? Recommendations to Improve Social Development through Enhanced Impact Evaluation'. Washington, DC: Center for Global Development.
- Wig, Tore. 2024a. 'Bistandsevaluering Krever Ikke et Eget Fag En Byråkratdreven Tilnærming Står i Veien for Forskningsbasert Utvikling.' *Morgenbladet* (Oslo), May 3. www.morgenbladet.no/ideer/tore-wig-bistandsevaluering-kraver-ikke-et-eget-fag/9409581.
- Wig, Tore. 2024b. 'Evaluering Uten Forskningsforankring Manglende Bruk Av Randomiserte Studier Er Bare et Symptom På at «evalueringsslaget» Henger Etter'. *Morgenbladet* (Oslo), May 16. www.morgenbladet.no/ideer/tore-wig-evaluering-uten-forskningsforankring/9410670.
- Wig, Tore. 2024c. 'Sjarlatanene Med Dress, Slips Og Powerpoint-Presentasjoner Den Moderne Sjarlatan Selger Analyser Til Staten, Skriver Tore Wig.' Kommentar. *Morgenbladet* (Oslo), April 12. www.morgenbladet.no/ideer/sjarlatanene-med-dress-slips-og-powerpoint-presentasjoner/8871334.

Part III

Educating the commissioner, through a dialogue between commissioners and evaluators



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

12 The good commissioner

Competencies for enhancing democracy in commissioning processes

*María Bustelo, Ramón Crespo
and Juan Andrés Ligeró*

Introduction

The commissioning process for evaluations is an essential and intricate part of the evaluation function. It serves as a bridge between the conception of policies and interventions and the assessment of their results and value. This process involves various stakeholders, each playing a crucial role at different stages. The success of an evaluation and its contribution to democratic processes heavily depend on how well the commissioning process is managed.

In evaluation theory and practice, the complex and diverse contexts in which evaluations are conducted are well-recognized and addressed by evaluation theory and practitioners. However, much of the literature and guidance focuses on the role of evaluators. Evaluators are often seen as the central figures who must ensure the success of the evaluation through their expertise and interaction with commissioners and policymakers. But as evaluators with both research and practical experience, we know that no matter how well an evaluation is designed and performed, if the commissioning process goes wrong or gets difficult, the evaluation will never manage to be successful.

As a result, many existing ethical codes, guides, and competency frameworks are geared primarily toward evaluators, outlining how they should conduct evaluations and engage with other stakeholders. While these resources acknowledge that evaluators are not working in isolation, they often fall short in addressing the competencies needed by other roles. With some exceptions (UKES Guidelines for Good Practice in Evaluation), there is practically no guidance for commissioners, who are part of the evaluation community and play an important role in the evaluation process. Commissioners play a pivotal role in conceiving, facilitating, and guiding the evaluation process and use of its results. This gap found in codes and competency frameworks is significant because commissioners are not merely facilitators; they are also change agents who have “agency” in evaluation and can influence the entire evaluation process, its results, and use.

To address this gap, we aimed to identify a comprehensive framework of competencies, which considers not only evaluators but also commissioners and users

of evaluations. The need for a well-defined set of competencies for commissioners and users is underscored by the increasing complexity of the environments in which evaluations are conducted. These environments are characterized by diverse stakeholders, evolving policy landscapes, and the demand for evaluations that can drive meaningful change. Therefore, it is essential to expand the focus beyond evaluators and consider the broader spectrum of roles involved in the evaluation process. In this chapter, we are going to focus specifically on the commissioner's role and its specificities in comparison with the evaluator's role. With this chapter, we aim to open a discussion on several questions. Which competencies do commissioners of evaluation need to make the best of a commissioning process? What is their potential active role? What do they need to know, and in which skills should they be trained to be "good" commissioners? What does it mean to be a "good" commissioner? Are there specific competences, distinctive from those needed for evaluators, that should be potentially developed? What do commissioners need to be better equipped for the evaluation commissioning process? This chapter aims to contribute to this book in its attempt to identify and critically analyze the evaluation function and the relations that are established among the different stakeholders participating in evaluation commissioning processes. We aim at this endeavor guided by two principles we think can better enhance democratic commissioning processes. First, by moving toward a more collaborative, productive, and meaningful relationship between commissioning bodies and evaluators, and second, by identifying a more active role for commissioners with "agency" during the commissioning process.

What we present in this chapter is part of the analysis we initiated as a participatory reflection process in 2023 with alumni from the Master's in Evaluation of Programmes and Public Policies at the Complutense University of Madrid (UCM). With more than two decades of history, this program has trained a substantial number of professionals who now occupy various roles in the field of evaluation. The master's is a professionally oriented program, meaning that some of the participants each year act as either evaluators or evaluation commissioners/managers while they are being trained in the master's program, something that enriches classroom activities and debates. Several alumni have ended, after the master's program, playing an important role in developing the evaluation function in their organizations; others are contributing as policy makers/users to the evaluation function in their respective sectors. Their experiences provide a rich empirical base for identifying the necessary competencies for evaluators, commissioners, and users. The process started earlier in 2021 when the Spanish National Institute of Public Administration commissioned us for a study to identify competencies for establishing "evaluation capacity itineraries" for public servants and employees. Although this work was never fully implemented, it served as a base for the later developed Competence Framework in Public Policy Evaluation¹ and its division into different roles of evaluators, commissioners, and users, although the term that has been used is "evaluation managers" and not commissioners, and indeed, the framework tends to relate to a more managerial, rather than a political or strategic, role of commissioners.

By drawing on established frameworks such as the EES competencies framework and the UK Evaluation Society's Framework of Evaluation Capabilities, and through engaging with alumni and tutors from the master's program, we seek to develop a nuanced understanding of the competencies required for different roles in the commissioning process. This chapter explores the competencies for the commissioner's role in detail, aiming to contribute to the evaluation literature and practice by highlighting the importance of a comprehensive approach to competency development in evaluation.

The chapter presents first some theoretical notes on the concept of agency and the study of evaluation competencies. The second and third sections explain the methodology used and the analytical framework applied. In the fourth one, we present the results of our empirical research, ending with a concluding section with lessons learnt and future research.

Theorizing on the concept of commissioners' "agency" and evaluation competences

Thomas Schwandt explains in his article on "Competencies for Evaluation as a Civic Science" the connection of the identification of competencies with the purpose of evaluation and how this is central in evaluation systems:

The identification of competencies for professional evaluation work depends on a clear understanding of the purpose of that professional work (...). Learning from evaluation for making practical improvements and strategic adjustments importantly figures in this idea of what evaluation is for. Ideally, this purpose should be the central focus in robust, institutionalized national evaluation systems.

(Schwandt, 2023: 24)

Therefore, before entering the concept of agency, we would like to set up the scene in which commissioners exert that agency. On the one hand, there are evaluation systems developed in many countries and organizations where the evaluation commissioning processes take place. These evaluation systems that most authors claimed from the 1990s as structures to be in place for institutionalizing evaluation offer numerous benefits for the evaluation endeavor. Among them, they provide a framework for the systematic review and improvement of programs and policies, allowing for an integration of evaluation planning with organizational strategy. They also enhance evaluation culture and evaluation capacity building within public organizations, as well as strengthen evaluation quality. However, they can also generate unintended side effects. Frequently, external requirements impose excessive constraints on the systems, becoming an administrative burden, which might compromise their independence, adaptability, and democratic relevance. This phenomenon, known as "evaluation capture", occurs when a proliferation of regulatory frameworks and mandatory evaluation requirements gives rise to

inflexible structures that prioritize compliance over substantive inquiry (Raimondo and Leeuw, 2022). This might even lead to evaluation systems losing their ultimate mission.

On the other hand, we have a pledge of stakeholders – evaluators, commissioners, and evaluation users – executing their roles within those systems or structures but, at the same time, being constrained by the administrative burden that somehow seems to be parallel to the institutionalization of evaluation systems. This has been further analyzed, developing the concept of evaluators’ resilience for navigating those systems (Leeuw and Pleger, 2023). Therefore, the effectiveness of these evaluation systems depends not only on organizational design but also on the capacity and engagement of evaluators, commissioners, and decision-makers, being crucial the dynamic interaction between institutional arrangements and individual agency (Bustelo and Jacob, 2025). Summing up, evaluation systems are shaped by contextual factors and institutional arrangements but very much also by the behavior and contributions of key individuals, among them, not only evaluators but certainly commissioners, to whom the evaluation literature has not paid much attention so far.

“Agency” is an old category for political analysis that has been especially used and developed by feminist theory. Agency refers to human action, that is, the capacity or ability of an actor to act consciously (Kantola and Lombardo, 2017; Hay, 2002; Hinterberger, 2013). It is “commonly associated with autonomy, free will, and choice (...), and related to power in the sense that individual and collective capacity to act is both constrained and enabled in different contextual circumstances” (Kantola and Lombardo, 2017: 71). Apart from power, agency is also related to another political category: structure, which “refers to contextual factors (e.g., social practices, political institutions) that show some regularity over time and escape the immediate control of actors” (Kantola and Lombardo, 2017: 71).

What feminist theories have highlighted is the ability of human action to respond, resist, negotiate, and transform structures (Hinterberger, 2013:7–8). Informed by social constructivism, feminist political analysis proposes the idea that agency and structures shape each other, being mutually constitutive; that is, social and institutional structures bind agents, who also have the capacity through action for transforming the structural context in which they are immersed (Kantola and Lombardo, 2017). This also goes in line with the concept of “civic agency” Schwandt uses in his 2023 article, as “the capacity to navigate, negotiate, and transform the world around us, and it has a collective and not simply an individual dimension” (Schwandt, 2023: 25) as can be used “to collaborate with others across differences to address common challenges” (Boyte, 2008; quoted in Schwandt, 2023: 25).

All of these theories are very useful to apply to the concept of commissioners’ agency, which can be defined as the ability to act consciously and navigate a concrete evaluation system, shaped by contextual factors and organizational arrangements, but with the intention to influence – and ultimately transform – that evaluation system in which the commissioning processes take place, through

action and strategy, negotiation with stakeholders – including evaluators and policymakers – and resistance – as well as resilience – to over bureaucratization or democratic mission loss. That is, commissioners are not only evaluation managers, but they also have a crucial role in enhancing significant, useful, and democratic commissioning processes.

Methodology

To identify the necessary competencies for evaluators, commissioners, and users in the commissioning process, we employed a participatory and multi-method approach. Participants included alumni, some of whom have played different roles in commissioning processes and APROEVAL² members.

Participatory reflection process

Our primary method involved a participatory reflection process with alumni from the UCM master's program. This program has been running since 2002 and has trained over 500 professionals, many of whom now serve as evaluators, commissioners, or managers. By leveraging their diverse experiences and roles, we aimed at gathering a comprehensive understanding of the competencies required for effective evaluation commissioning. We first ran a workshop in June 2023 with a group of 15 alumni involved in the master's program (either as teachers or tutors), in which we used the ***persona technique*** for developing the different roles (evaluators, commissioners, and evaluation users). This method involves creating detailed and realistic profiles of different stakeholder roles based on collective input. Through group discussions, participants helped identify the various roles involved in the commissioning process, including evaluators, commissioners, and users. This technique allowed us to explore the specific needs, challenges, and competencies associated with each role. In a second phase, we implemented a larger workshop with approximately 40 participants in the Annual Seminar on Experiences of Evaluation in November 2023.

Documentary analysis

An analysis of existing competency frameworks, specifically the EES competencies framework and the UK Evaluation Society's framework of evaluation capabilities, was conducted. These frameworks provided a foundational structure for our study, allowing us to compare the competencies identified through our participatory methods with established standards in the field.

Data synthesis

The data collected were synthesized to identify common themes and patterns. We divided the competencies into theoretical-conceptual and practical-interpersonal

categories, reflecting both the knowledge and skills needed to design and conduct evaluations, as well as the attributes necessary for effective communication, negotiation, and ethical practice.

Subsequently, we engaged in an iterative review process with participants through email and participatory workshops. Preliminary findings were shared with alumni and tutors from the master's program for feedback and validation. This iterative process ensured that the identified competencies were relevant, comprehensive, and accurately reflected the evaluation community.

By combining participatory methods with existing frameworks and an iterative validation process, our methodology provides a robust basis for identifying and developing competencies for evaluators, commissioners, and users. This approach not only highlights the critical skills and knowledge required but also underscores the importance of continuous learning and adaptation in the field of evaluation.

Analytical framework and categories of competencies

Drawing on the participatory research process described above, the competency categories presented below were cross-referenced with the different phases of an evaluation process (see Annex: Competencies Table). This structure allowed us to explore how each type of competency is required at different moments of the evaluation lifecycle. Rather than listing the competencies in isolation, this approach highlights their contextual relevance and dynamic interaction depending on the stage of the process. It also oriented our fieldwork to reflect on when, why, and how each skill is most needed.

The categories of competencies we used were:

- A Understanding the discipline: ability to appreciate the unique role of evaluation in society, familiarity with evaluation theories, schools, concepts, and approaches, and recognition of its links with social sciences.
- B Quality standards: ability to identify relevant evaluation questions, constructively engage with stakeholders, understand the value of diverse evaluation approaches, and adapt designs and methods to the context.
- C Technical and methodological potential and methodological skills: understanding the formative and summative potential of various methods, techniques, and instruments, along with ensuring their validity and reliability.
- D Management capacity: ability to respond to the needs and concerns of various stakeholders, comprehend the theory of change underlying the evaluand, anticipate opportunities and challenges throughout the process, and prudently manage resources to achieve evaluation objectives.
- E Interpersonal skills: ability to clearly communicate objectives and results, promote teamwork, assertive negotiation, and positive conflict resolution. This also includes cultural and gender sensitivity.

- F Ethics and transformative vision: ability to uphold ethical standards and democratic values in evaluations, promote the use of evaluations, maintain independence in data interpretation and judgment, contribute to the professional evaluation community, and pursue continuous professional development.

These were crossed with the phases of an evaluation process, which were the following:

- a Commissioning and motivation for evaluation: Initial framing and conception of the evaluation. It includes understanding why the evaluation is needed, who will use its results, and for what purpose. It involves identifying the right approach, defining evaluability criteria, and shaping the scope of the task.
- b Definition of the evaluand and context analysis: Clarifying what is being evaluated and understanding its environment, it focuses on analyzing the policy, program, or service under evaluation. This includes reviewing documentation, identifying the theory of change, and mapping the institutional, political, and social context.
- c Methodological choice: selection of approaches, strategies, and designs that best respond to the evaluation questions and context. It includes the ability to compare options, assess their feasibility, and justify methodological choices (commissioners and evaluators need to be aware of the political and ethical implications of these decisions and ensure that the selected methods allow for credible and useful results).
- d Operationalization: defining indicators and rubrics, selecting data collection tools, and ensuring the methodological plan is coherent and realistic. It requires translating the evaluation framework into clear and practical steps while maintaining quality standards. Attention to gender, equity, and sustainability dimensions is also central in this phase.
- e Fieldwork: implementation of data collection in the field, either through interviews, surveys, workshops, or other techniques. It includes dealing with real-life situations, unexpected events, diverse actors, etc., and requires flexibility, communication skills, and a strong ethical approach to protect participants and ensure data quality.
- f Data analysis: organizing, processing, and interpreting the data. It requires methodological rigor, reflexivity about one's own assumptions, and integrating different sources and perspectives (ensuring triangulation and maintaining transparency for credible analysis).
- g Interpretation and judgment: giving meaning to the results, identifying patterns, and linking findings to the original evaluation questions. This phase also includes making value judgments, being aware of limitations, and facilitating deliberation among stakeholders.
- h Influence and communication: how the evaluation results are communicated, shared, and used. It includes developing dissemination strategies, writing clear reports, managing resistance, and supporting dialogue with decision-makers.

Results

The participatory reflection process yielded significant insights into the competencies necessary for both evaluators and commissioners involved in the commissioning process. We identified key competencies that are crucial across different roles, as well as the nuances specific to each role. The competencies framework³ is in the Annex to this chapter. Here, we elaborate on the changes proposed during the second workshop in November 2023, in which the attribution of each competency to each role was discussed. In the analysis, we will highlight the specificities of the commissioner's role. Curiously enough, after the second participatory workshop, the competencies assigned to commissioners tended to be amplified.

In general, commissioners' competencies are present in every category and every phase, although they are slightly more concentrated in the categories of "enables the evaluation", "management skills", and "maintains an ethical and transformative vision". Regarding the phases, commissioners tend not to be so present in methodological choice, operationalization, and data analysis. On the contrary, their role seems to be more active at the beginning of the evaluation process in "commissioning and motivation for evaluation" and at the end in "influence and communication".

Understanding evaluation theory and practice: There are common competencies with evaluators, such as familiarity with evaluation theories and approaches, knowledge of global trends and emerging skills, and recognizing the distinctive role that evaluation plays in society, and both roles should be able to scope evaluations plausibly. As something specific to their role, commissioners are also expected to contribute to the construction of the evaluation task by helping set priorities and by building alliances with other actors. They should be able to redirect or reshape the task when it is not well defined and translate it into a structured proposal.

Technical and methodological competencies: both evaluators and commissioners must have the skills to design and implement evaluations effectively, ensuring methodological rigor and validity of results, and the ability to analyze qualitative and quantitative data and draw valid conclusions to ensure that evaluation findings are reliable and actionable. They also need to understand the implications of different evaluation theories and methods and be empowered to discuss and choose the most legitimate approaches for each context.

In addition to the competencies already listed, commissioners should also be able to promote and understand participatory approaches, recognizing the contexts in which they are most appropriate. Commissioners are also expected to exercise flexibility in methodological planning, adapting to unforeseen changes in the evaluation context. Taking this as a starting point, commissioners should be capable of planning realistic evaluation timelines and adjusting them when needed. They also need the ability to assess the technical quality of proposed methodologies and evaluation approaches (which includes understanding the complexity of evaluation

designs and their practical feasibility so they can support informed decisions about methodology).

Another relevant aspect is their capacity to assess the quality of available data and to give credibility to non-traditional or innovative methods, if they are suitable for the evaluation's purpose. Finally, commissioners should understand that no method is inherently superior – its appropriateness always depends on the project type and evaluation objectives.

Management competencies: Both commissioners and evaluators require strong project management skills to handle budgeting, scheduling, and resource allocation effectively. Engaging stakeholders constructively is also critical for both roles. This competency includes communication, negotiation, and conflict resolution skills. Both roles must understand and align evaluations with strategic policy goals, ensuring that evaluations support broader organizational objectives.

As for commissioner-specific competencies, they should be able to formulate high-quality terms of reference for evaluations, including clear and focused evaluation questions. They should also understand and manage the political implications of evaluations, ensuring that evaluations are not only methodologically sound but also politically relevant and feasible.

In addition to managing political implications and writing good terms of reference, commissioners must recognize and respond to stakeholders' concerns and needs. They also should be able to support the creation of governance spaces, like steering groups, that promote participation and commitment. This is clearly connected to the commissioner's role in ensuring that evaluation findings are used. They should be able to develop communication plans, adapt language for different audiences, and identify possible resistance to change. They are also expected to help create the conditions for applying recommendations and to feel responsible for promoting their use.

Finally, beyond project management and stakeholder engagement, commissioners play a crucial role in facilitating access to relevant data and internal documentation. This involves understanding data protection regulations and ensuring that evaluators can consult necessary sources, including administrative records and digital systems. Commissioners should also be able to identify and clearly express the information needs that the evaluation should address, in line with its goals and the institutional context.

Interpersonal competencies: both evaluators and commissioners need to communicate evaluation goals, processes, and findings effectively. This competency ensures that complex information is accessible to diverse audiences. Promoting teamwork and collaboration among evaluation teams and stakeholders is essential for both roles. This fosters a positive working environment and constructive dialogue. In this regard, commissioners are expected to demonstrate facilitation skills (this means guiding stakeholder groups, leading constructive discussions, and fostering collaboration). Finally, both roles should demonstrate sensitivity

to cultural and gender issues, ensuring that evaluations are inclusive and equitable.

The fieldwork we carried out also emphasizes the importance of commissioners adapting to the communication styles and cultural backgrounds of the stakeholders involved in the evaluation. This means being able to identify the appropriate interlocutors and knowing how to interact with them effectively, with sensitivity to identity, diversity, and communication capacities (these support inclusive and respectful engagement throughout the evaluation process).

Ethical and transformative vision: a commitment to ethical practices in evaluation, including independence, confidentiality, and respect for participants, is crucial for both evaluators and commissioners. Also, promoting the use of evaluations by encouraging the application of findings to inform decision-making and drive improvements. This ensures that evaluations lead to meaningful changes and enhancements. Both roles should engage in continuous professional development, participating in professional networks and seeking opportunities to enhance their competencies.

Although not always with deep technical knowledge, commissioners are expected to promote attention to cross-cutting issues such as gender equality, human rights, and planetary health. This implies recognizing that methodological choices are not neutral and that they reflect broader ethical and ideological orientations.

Commissioners, therefore, should be aware of how these perspectives shape the evaluation and take responsibility for articulating their relevance clearly with stakeholders from the beginning. In other words, commissioners should not only uphold ethical principles but also show sensitivity to global challenges (e.g., alignment with the Sustainable Development Goals and awareness of how evaluation can promote sustainability and equality).

Additionally, during fieldwork, commissioners should ensure participant protection and confidentiality and show respect for local customs, beliefs, and social roles. And in data analysis and interpretation, they need to integrate gender perspectives, understand validity criteria, and support meaningful interpretation of results.

Conclusions

The above categories were very useful for guiding the process, but as can be seen easily, frequently they overlap and are connected to one another in multiple ways. In fact, after the second workshop, we thought that maybe a new category of “Promoting the use of evaluation” could be formulated for referring specifically to the commissioners’ role related to understanding the general and the organizational context in which the evaluation takes place, the identification of policy users’ information needs, and the continuous engagement and communication with stakeholders. This was a result after having worked on competencies, also from the perspective of the commissioners. There was a clear tendency to extend our scope

toward the competencies that give more “agency” to the commissioners, leaving more space for their influence in democratic commissioning processes: strategic, oriented to the participation of stakeholders and the use of evaluation, and to the creation of public value, as well as good coordination and joint work with evaluators.

Indeed, in democratic contexts, the evaluation process must guarantee the decision-making spheres of the different actors involved in the evaluation, and this responsibility falls mainly on the commissioners who, as actors with the greatest ability to influence the evaluation framework, can largely determine the commission and its performance.

However, this increased agency does not mean that commissioners, nor their institutions, are the sole “owners” of evaluation processes and evaluators are just external actors working for them and under their control. Exerting their power as commissioners should not imply taking it from evaluators and other stakeholders but balancing that power in a horizontal and collaborative relationship with them. From a democratic perspective, a recognition of both evaluators and other stakeholders as people with their own agency and important roles to play, and with whom they need to interact and work with, is necessary beyond the institutions they represent and the policy users to serve. That is, commissioners must exercise their agency, but not at the expense of the agency of evaluators and other stakeholders. Unfortunately, as an indicator that very frequently this is not the case, some professionals consulted reported experiences in which commissioners have denied that the evaluator’s name and recognition appear in the evaluation report.

During the interviews we conducted to identify competencies, some experiences were also reported where commissioners interfered outside the established channels in methodological decisions regarding which aspects of the program should not be evaluated, the weight to be given to some interpretations over others, or the modification or deletion of certain judgments in the final reports. The mechanisms for influencing are varied, ranging from subtle but untimely suggestions to more explicit mechanisms of pressure, such as express coercion or threats of various kinds, such as communicating professional misconduct on social media or not making the final payments for the evaluation.

Considering these bad practices reported, and given the “extension” of commissioners’ competencies once we started to take seriously identifying competencies from the commissioner’s perspective, we can raise questions around how much about evaluation a commissioner should know? Is it necessary to complete an education on evaluation for commissioners? Are there good commissioners – we are sure they are – who are not evaluation specialists? Are the best commissioners the ones who have been evaluators before? After more than two decades of educating evaluators (and some commissioners, too), we advocate for an evaluation education for commissioners. The fact of having both roles in the classroom is an extra opportunity for enriching perspectives and debates. More than having a specific education for commissioners, separated from evaluators, we advocate for having a joint education, when possible, reinforced by a continuous evaluation capacity-building

for both evaluators and commissioners – a joint education that should better reflect the specific competencies for commissioners and a commissioner’s perspective and that highlights the democratic nature of evaluation processes. Now that democracy is at stake, the role of commissioners for assuring democratic commissioning processes that contribute to public value is more important than ever.

Notes

- 1 This “Competency Framework for Public Policy Evaluation” was elaborated in 2023 in Spain by the Institute for Public Policy Evaluation and the National Institute of Public Administration. It identifies the competencies that civil servants with duties or functions related to the evaluation of public policy should demonstrate, based on their professional profiles. See https://digital.gob.es/content/dam/portal-mtdfp/funcion-publica/evaluacion-de-politicas-publicas/COMPETENCIAS/Marco_de_competencias_en_evaluacion.pdf (in Spanish).
- 2 APROVEAL (Asociación de Profesionales por la Evaluación) is a Spanish and Portuguese (Iberian) association of evaluation professionals.
- 3 The competencies framework elaborated is organized following the different phases of the evaluation on the vertical axis and the categories of competencies on the horizontal one.

References

- Boyte, Harry C. 2008. “Civic Driven Change and Developmental Democracy.” In *Civic Driven Change: Citizens’ Imagination in Action*, edited by Alan Fowler and Kees Biekart, 17–29. The Hague: Institute of Social Studies.
- Bustelo, María, and Steve Jacob. 2025. “From Studies to Systems: Ray Rist’s Influence on Evaluation Systems—Insights from the International Research Group on Policy and Program Evaluation (INTEVAL).” *Journal of MultiDisciplinary Evaluation* 21(50), 123–129. <https://doi.org/10.56645/jmde.v21i50.1175>
- Hay, Colin. 2002. *Political Analysis*. Basingstoke: Palgrave Macmillan.
- Hinterberger, Amy. 2013. “Agency.” In *Gender: The Key Concepts*, edited by Mary Evans and Carolyn Williams, 7–13. London: Routledge.
- Kantola, Johanna, and Emanuela Lombardo. 2017. *Gender and Political Analysis*. Basingstoke: Palgrave Macmillan.
- Leeuw, Frans L., and Lisa Pleger. 2023. “Evaluation Capture, Evaluator Resilience, and the Need for Competencies of Evaluators.” *Journal of MultiDisciplinary Evaluation* 19 (46): 45–53. <https://doi.org/10.56645/jmde.v19i46.863>
- Raimondo, Estelle, and Frans L. Leeuw. 2022. “Evaluation Systems and Bureaucratic Capture: Locked.” In *Changing Bureaucracies: Adapting to Uncertainty, and How Evaluation Can Help*, edited by Burt Perrin and Tony Tyrell, 143–160. New York: Routledge.
- Schwandt, Thomas A. 2023. “Competencies for Evaluation as a Civic Science.” *Journal of MultiDisciplinary Evaluation* 19 (46): 23–30. <https://doi.org/10.56645/jmde.v19i46.859>

- EVALUATORS' COMPETENCIES
- COMMISSIONERS' COMPETENCIES

ANNEX: COMPETENCIES' FRAMEWORK PROPOSED AND ELABORATED BY THE UCM MASTER ON EVALUATION

EVALUATORS' COMPETENCIES		Related to the adequate understanding of the history, approaches, models, and theories of evaluation and their implications for governance, design, purposes, practices, uses, and methods of evaluation. These competencies allow for the design, conduct, and supervision of evaluations effectively and for making informed decisions about the most appropriate methodologies and approaches for each situation.			Related to what actually happens in the field. These include communication skills and interpersonal attributes involved in the design and management of the evaluation process. Listening and negotiation skills, the willingness to adapt to the context, independent thinking, as well as strong ethical principles, are essential attributes for evaluation.		
COMMISSIONERS' COMPETENCIES							
PHASES		Understands the discipline of evaluation	Enables the evaluation	Manages technical and methodological aspects	Has management skills	Has interpersonal skills	Maintains an ethical and transformative vision
COMMISSIONING AND MOTIVATION FOR EVALUATION	Understands the distinctive role that evaluation plays in society ● ●	Is capable of identifying the purposes of the evaluation, considering who will use the findings and what types of decisions will be made based on the recommendations ● ●	Has the ability to convert the task into a structured proposal ● ●	Identifies and responds to the needs and concerns of different stakeholders ● ●	Can communicate the needs of others and build appropriate alliances for the evaluation ● ●	Defends ethical standards and democratic values in the conceptualization of the evaluation ● ●	
	Is familiar with the theories, schools, concepts, and approaches of evaluation ● ●	Is capable of establishing evaluability criteria and scoping the task accordingly	Recognizes the importance of adequately appropriating the task (grasping it) ●	Has the capacity to formulate quality terms of reference for all types of evaluations, including clear and well-focused evaluation questions ●	Is capable of assisting in establishing priorities during the construction of the task ● ●	Shows cultural and gender sensitivity ● ●	
	Has knowledge of global trends in evaluation that affect their work as well as emerging techniques and skills that advance the profesión ● ●	Has the ability to redirect the task when it is not well-defined ● ●		Understands and manages the political implications of the evaluation ● ●	Is competent in dialogue and participatory leadership ●	Shows sensitivity to planetary sustainability and knows how to align the task with the SDGs ● ●	
	Knows how to plausibly scope the evaluation task considering available resources ● ●	Has the ability to redirect the task when it is not well-defined ● ●				Participates in professional networks and provides spaces for self-development to improve competencies ●	
DEFINITION OF THE EVALUAND AND CONTEXT ANALYSIS	Understands the links between the discipline and social sciences ●	Relates constructively with all stakeholders ● ●	Has solid knowledge of various types of theories of change and logical models ●	Knows how to negotiate evaluation priorities considering available resources ● ●	Knows how to consolidate a steering group with stakeholders to facilitate consultation and governance spaces that ensure their commitment to the evaluation ● ●	Is capable of identifying potential challenges to the independence and impartiality of the evaluation and knows how to navigate them under pressure ●	
	Understands the links between the discipline and other fields of thought (feminism, interculturality, environment, etc.) ●	Identifies relevant evaluation questions and demonstrates the ability to prioritize them ● ●	Is capable of developing theories of change and logical models to facilitate evaluative thinking among stakeholders ●	Has sufficient ability to assess the technical skills of evaluators tasked with conducting an evaluation ● ●		Shows sensitivity to various axes of inequality (gender, disability, age, sexual orientation, etc.) and knows how to help neutralize them through evaluation ● ●	
	Knows how to search for, process, and legitimately use information related to the task (bibliographic references, previous evaluations, etc.) ●	Knows what internal and external context information to share to benefit the evaluation ●	Understands the theory of change underlying the evaluand (policy, service, or program) ● ●				
		Knows their rights in the evaluation and knows when to speak up ●					

● EVALUATORS' COMPETENCIES ● COMMISSIONERS' COMPETENCIES	COMPETENCIES 2					
	Related to the adequate understanding of the history, approaches, models, and theories of evaluation and their implications for governance, design, purposes, practices, uses, and methods of evaluation. These competencies allow for the design, conduct, and supervision of evaluations effectively and for making informed decisions about the most appropriate methodologies and approaches for each situation.			Related to what actually happens in the field. These include communication skills and interpersonal attributes involved in the design and management of the evaluation process. Listening and negotiation skills, the willingness to adapt to the context, independent thinking, as well as strong ethical principles, are essential attributes for evaluation.		
PHASES	Understands the discipline of evaluation	Enables the evaluation	Manages technical and methodological aspects	Has management skills	Has interpersonal skills	Maintains an ethical and transformative vision
METHODOLOGICAL CHOICE	Understands the political implications of different evaluation theories and their use ● Is sufficiently empowered to discuss the legitimacy of various approaches and methods ● ●	Understands the summative and formative potential of methods (experimental, quasi-experimental, qualitative, participatory, and mixed) ● ●	Knows how to design evaluations with different levels of complexity, based on best existing practices and innovative approaches ● ● Has sufficient ability to assess the technical quality of the proposed evaluation approach and methodology ● ●	Has the ability to determine the scope of an evaluation and articulate the mechanisms to implement it ● ● Knows how to plan a realistic work schedule for evaluations with varying levels of complexity (and adjust the work plan as necessary) ● ●	Is capable of objectively evaluating the pros and cons of different methodological choices ●	Understands participatory approaches and the contexts in which they are most appropriate ●
OPERATIONALIZATION	Knows how to operationalize different informational needs (vertical and horizontal work) ●	Is capable of validating or giving credit to unorthodox (or innovative) methods to explore what is considered relevant for the evaluation ● Is capable of discerning the quality of available data ●	Identifies the most appropriate data collection instruments for each task ● Has sufficient ability to assess the technical quality of evaluation metrics and instruments ● Designs appropriate inquiry instruments to answer evaluation questions (construct validity) ● ●	Manages resources prudently to achieve the evaluation objectives ● ●	Demonstrates people and relationship management skills to ensure effective work and meeting deadlines ● ●	Possesses solid knowledge on gender equality, human rights, and planetary health, and is capable of ensuring that the evaluation design, questions, and analysis incorporate this knowledge in any context ● ●

- EVALUATORS' COMPETENCIES
- COMMISSIONERS' COMPETENCIES

COMPETENCIES 3

Related to the adequate understanding of the history, approaches, models, and theories of evaluation and their implications for governance, design, purposes, practices, uses, and methods of evaluation. These competencies allow for the design, conduct, and supervision of evaluations effectively and for making informed decisions about the most appropriate methodologies and approaches for each situation.

Related to what actually happens in the field. These include communication skills and interpersonal attributes involved in the design and management of the evaluation process. Listening and negotiation skills, the willingness to adapt to the context, independent thinking, as well as strong ethical principles, are essential attributes for evaluation.

PHASES	Understands the discipline of evaluation	Enables the evaluation	Manages technical and methodological aspects	Has management skills	Has interpersonal skills	Maintains an ethical and transformative vision
FIELDWORK	Understands theoretical approaches that specifically indicate how to include gender perspective (PDG) in fieldwork ● ●	Demonstrates a collaborative attitude and openness to teamwork ● ●	Gathers rigorous and relevant evidence, knowing how to analyze and interpret the context ● Knows how to involve stakeholders in fieldwork, using techniques that foster open and sincere dialogue ● Knows how to scale fieldwork according to available resources and the implementation context ● ●	Has the ability to assess the implicit risks in fieldwork and to design and implement contingency plans when necessary ● ●	Works collaboratively to address challenging dynamics and/or deal with controversial issues, ensuring diverse perspectives are integrated into the evaluation ● ● Knows how to ask relevant and non-coercive questions and actively listens to responses ● Demonstrates an empathetic disposition ● Shows flexibility in the face of unforeseen events and the ability to react and resolve conflicts ● ● Has the ability for critical observation ● Can reflect on how their own identity influences the interpretation process ●	Is capable of ensuring the protection and confidentiality of evaluation participants ● ● Takes into account that the evaluation does not endanger the individual or the organization ● ● Understands and knows how to respect local customs, religious beliefs and practices, and ethnic and gender roles ● ●
DATA ANALYSIS	Has the ability to critically reflect on their own beliefs, assumptions, and theoretical preconceptions (and is predisposed to inter-paradigmatic dialogue) ●		Conducts solid data analyses, whether qualitative or quantitative ● Knows how to operationalize methodological triangulation at different levels ●			Consistently integrates a gender perspective in the analysis to address planned objectives and facilitate the communication and use of the results in this regard ● ●
INTERPRETATION	Understands the limits of interpretation and its relevance for proper judgment ●	Actively participates in the interpretation and avoids non-constructive opposition ●	Has the ability to integrate and interpret results ● Understands that validity and reliability criteria differ according to the method. ● ●		Negotiates assertively and manages conflicts positively ● ● Has the skills to extract and communicate lessons learned from difficult situations ● ●	Maintains independence in the interpretation and judgment of data ● Holds a situated perspective and considers the cultural, social, political, and economic environment in which the intervention takes place ●

● EVALUATORS' COMPETENCIES ● COMMISSIONERS' COMPETENCIES		COMPETENCIES 4				
PHASES		Related to the adequate understanding of the history, approaches, models, and theories of evaluation and their implications for governance, design, purposes, practices, uses, and methods of evaluation. These competencies allow for the design, conduct, and supervision of evaluations effectively and for making informed decisions about the most appropriate methodologies and approaches for each situation.				
		Understands the discipline of evaluation	Enables the evaluation	Manages technical and methodological aspects	Has management skills	Has interpersonal skills
INFLUENCE AND COMMUNICATION		Knows how to present oneself as an expert and trustworthy person, and how to describe and promote their work to demonstrate its value and relevance ●	Writes clearly and concisely, and can communicate technical evaluation concepts in a way that non-experts can easily understand ●	Is capable of developing a communication plan that shares the knowledge acquired through the evaluation to ensure the appropriate dissemination of results ● ●	Anticipates the implications that results may have for the process and acts accordingly ●	Communicates objectives and results clearly and is capable of adapting communication to different audiences ● ●
		Understands the theoretical uses described for evaluation as well as their political implications ● ●	Knows and knows how to use different communication media ● ●	Has sufficient ability to assess the technical quality of evaluation reports ● ● Is capable of supporting the comprehensive dissemination and use of evaluation results to inform and improve programming ● ● Knows how to design and manage a system to track the use of evaluation recommendations (influence) ● ●	Knows how to leverage networks and associations with various stakeholder groups to scale results and the use of evaluation ● ● Feels empowered to suggest commitments to the use of evaluation to commissioning agencies ●	Is effective in communicating with senior management and stakeholders ● ● Knows how to identify possible resistance to change among stakeholders and how to neutralize it ● ● Contributes to the professional evaluation community and seeks continuous professional development ● Considers their part of the responsibility in creating conditions for the recommendations to be applied ● ● Has facilitation skills and is capable of guiding stakeholder groups, leading discussions, fostering collaboration, and managing group dynamics ● ●

13 Training for critical evaluative thinking

*Per Øyvind Bastøe, Øyunn Syrstad Høydal and
Knut Fossetøl*

Introduction

This chapter discusses an initiative to address deficiencies in commissioning and using evaluations in public administration in Norway through a training course. The purpose of the course, held twice in 2024, was to build evaluation capacity by teaching effective, useful, and professional evaluation practices. The aim was to improve the outcomes of evaluation activities in the organizations the participants belong to (Preskill and Boyle, 2008; Wandersman, 2014).

Building evaluation capacity has also been discussed in previous books in this series. For instance, it was one of the conclusions in the book edited by Boyle and Lemaire (1999) that unique institutional characteristics in each country make generalization difficult. In line with this conclusion, our purpose was to build an evaluation capacity custom-made for the Norwegian context.

The course was based on a recognized discrepancy between public administration assigned to solve persistent problems requiring extensive knowledge, collaboration, and cross-disciplinary dialogue and a government governed by indicators and parameters that are overly simplistic and narrow to encompass this method of operation. The assumption was that the solution to address this challenge was not merely to enhance the knowledge of commissioning, implementing, and using evaluations. There was also a need to cultivate dialogical skills and critical thinking at both individual and organizational levels. The strategy was to encourage critical, evidence-based, and independent approaches to evaluation by promoting broad reflections related to evaluative thinking. Evaluative thinking refers to an attitude characterized by curiosity, belief in the value of evidence, and the ability to reflect and ask questions as a basis for action (Baker and Bruner, 2012). This way, building evaluation capacity holds the potential to promote democratic values within public administration.

The need for strategic anchoring, systematization, and planning of evaluations has been highlighted as an important improvement area to strengthen the use of evaluations in central government agencies (Haugnæss and Gram, 2017). Such considerations are shared by the OECD, referring to governments being confronted with complex policy challenges, declining levels of trust, and resource constraints. To establish robust institutions and practices that promote the use of public

DOI: 10.4324/9781042000968-16

This chapter has been made available under a CC-BY-NC-ND license.

policy evaluations, OECD has launched recommendations for policy evaluations structured around three mutually reinforcing pillars; (i) Institutionalize evaluation, (ii) Promote the quality of evaluations, and (iii) Conduct evaluations that impact decision-making.

Despite being a requirement since 1997, the evaluation practices in Norway are remarkably variable across agencies and ministries (DFØ, 2016; Høydal, 2019). Arguably, there is a fragmented understanding of what evaluation is and how it can be implemented and used in a way that produces knowledge leading to better policy and improved use of knowledge in governance. Furthermore, there is a lack of knowledge about how civil servants commission and utilize evaluations and other forms of evidence, even though a key task for this group is to provide the knowledge required for policy formulation (Høydal, 2020). Efforts over the years have been made to strengthen public evaluation practice in Norway through recommendations and guidelines issued by the Ministry of Finance (MoF) and the Directorate of Public and Financial Management (DFØ). So far, these efforts have had limited effects.

The training course was developed and implemented through collaboration between an international evaluation expert, an experienced evaluator/researcher, a researcher in the field of evaluation, and representatives of DFØ.¹ Together, we represented a wide range of evaluation expertise and insights in some of the central perspectives present in the national evaluation context.

The Norwegian context

The global trend of evaluation takes place within national knowledge regimes characterized by their respective institutional and cultural features (Campbell and Pedersen, 2014). For instance, Duffy (2017) claims UK policy evaluation to be close to economic traditions like auditing and monitoring, while the Norwegian evaluation tradition is heavily influenced by the social sciences, both theoretically and methodologically (Dahler-Larsen, 2013). The influence from social sciences could be explained by the establishment of several institutes of applied social science to produce knowledge for policymaking in the post-war period. Institutes that later came to play a central role in evaluations of public sector initiatives. An important institutional base for Norwegian policy development and implementation is the close collaboration between researchers, administration, politicians, and interest organizations. Different research institutes and research competencies are closely linked with different functions in public administration. The various sectors maintain research institutions and frequently use experts for investigation. The close connection between research and evaluation could be the reason why Norwegian bureaucrats generally believe evaluations are of good quality and why most are positive toward evaluation more generally (Høydal, 2019).

The anchoring in the social sciences may also explain the lack of academic training specific to evaluation and why a distinct evaluation profession has not emerged. Researchers from universities and research institutes have traditionally performed most public evaluations and are still, together with a growing number

of consultants, significant contributors. Consequently, the general awareness of evaluation is low, and terms like “applied research”, “reports”, and “evaluations” are, to some extent, overlapping. The international evaluation vocabulary has also had limited use and influence. However, it is reasonable to assume that consultants representing international companies are more familiar with the international evaluation standards and concepts than evaluators based in Norwegian research institutions. Winckler Andersen (2014) argues that interests and strategies may also be prioritized differently between commissioners, consultants, and researchers.

All parts of the Norwegian central administration are formally required to provide information on their effectiveness and efficiency by conducting evaluations. The evaluations are meant to be integrated into plans and strategies along with other forms of knowledge production (DFØ, 2016; Tornes, 2012). Hence, evaluations are in practice, on the one hand, linked to research and are seen as central in providing a knowledge base for political decisions, in line with the ideal of evidence-based policymaking. On the other hand, the government administration, in its guidelines, predominantly sees evaluation as one of several tools in performance management. Performance management is integrated with financial management and planning and follows a set structure with budget, letters of allocations from ministries to agencies, planning, and reporting both within each agency, between agencies and ministries, and from sector ministries to the Ministry of Finance.

How evaluations are initiated varies between ministries and agencies and between different sectors. Evaluations commissioned by ministries tend to be more political and ad hoc than the evaluation activity initiated by agencies (DFØ, 2016). The agencies play a central role as providers of evidence for the central administration. The relationship between agencies and ministries balances the professional autonomy of the agencies against the political agenda of the ministries and the political regime. While some agencies initiate most of their own evaluations, others mainly follow instructions from their ministries. There are also differences in the selection of programs and other activities to be evaluated. Some follow a structured approach of requiring evaluations of all new programs, while others initiate evaluations only where there is clear political interest (Høydal, 2019).

There are significant variations in the evaluation frequency between different sectors. While sectors like defense, transport, and communication rarely commission external evaluations, sectors like health, education, and social policy do this frequently (Tornes, 2012). The variation could be explained by differences in complying with the regulative text as well as sector-specific characteristics like the size and number of public interventions. According to Bogenschneider and Corbett (2010), the field of social policy is, for instance, particularly vulnerable to debate and disagreement, both because it is characterized by value-laden phenomena such as family, education, and health, and because it contains a high proportion of so-called wicked problems. This could be a reason why social policy is among the most frequent evaluation initiators.

No agencies or ministries have a separate evaluation unit, although the big agencies have departments dedicated to knowledge management with personnel specialized in commissioning evaluations.² The role of this personnel is to identify

and select the best evaluators as well as ensure the quality of the evaluations. The commissioning process is usually a collaboration between experts in research and development units or technical departments and procurement units. The general perception is that the competence of those involved in the commissioning processes is variable, and there is also a certain fear of making formal mistakes, which leads to a certain degree of rigidity in the processes.

Even though most public evaluations are in the threshold that requires international tender (DFØ, 2016) and thus are strictly regulated, there are ways for the public sector to influence the choice of evaluators. In interviews, civil servants have revealed how they influence the choice, for instance, by asking for project leaders or team members with specific academic qualifications for large evaluations. However, they may choose consultants without such qualifications for smaller evaluation projects, where effectiveness is more important than analytic competence. Similarly, they can formulate the tender announcement specifically regarding which aspects to include in the evaluation and thus influence who is eligible (Høydal, 2020).

The experts are not *avant-garde*

The outset for the collaboration in the organizing of the training course was an emerging intention within DFØ to establish a closer relationship with researchers in academic institutions and involve them in tasks such as training. We all shared an interest in evaluation and an understanding that central administration needs better knowledge about evaluation. At the same time, we realized that we had different perceptions of how these elements are connected and how they ideally should be interconnected. We considered ourselves less as experts and more as facilitators for dialogue to increase mutual understanding with the course participants.

DFØ was primarily concerned about the seemingly widespread confusion concerning the links between evaluation and other tools used for performance management in the central administration. Some of the confusion was caused by evaluations being administered by units focusing on knowledge creation, while performance management was administered by central administrative units. Several other aspects were also brought to the table, such as:

- The overall lack of practical skills concerning how to commission evaluations and how to use evaluations.
- The need for better insight concerning methodological possibilities and limitations.
- The lack of overarching knowledge about evaluation *per se*.

It turned out that planning the course contained many of the same challenges we thought the course would address. We agreed that there was a need to view performance management and evaluation in context, but we did not have a unanimous understanding or even a common answer to how to do this. The representatives from DFØ were to some degree bound to policy regulations in their communication, and

it took a while before the other members of the group understood the directorate's perception of the problem related to evaluation and performance management. In the end, the representatives from DFØ acknowledged that one of their main concerns was that evaluations rarely serve as a significant supplement to the quantitative and indicator-based process of performance management due to the silo mentality in the bureaucracy. One of the problems is the lack of collaboration between research units and performance management units. A consequence of this separation is that management parameters are detached from knowledge creation.

As the training evolved, new issues emerged, for instance, related to the use of international concepts. The international evaluation expert and DFØ were influenced by traditions and concepts from the international evaluation community, like the OECD standards and criteria. While the evaluation researcher and the evaluator, both sociologists, came from a Norwegian research background, they were reluctant to adopt these concepts. From their perspective, the strength of the Norwegian evaluation tradition is its influence from social sciences and a belief that this is a better guarantee for quality than advice from the OECD. Instead of trying to develop a common language and a shared conceptual platform, we chose to accept the different views and to be open about this plurality in the course and to use our multiple perspectives as an asset.

Based on the experience from DFØ, we also identified a potential conflict between participants' request for quick fixes, that is, the high interest in practical guidelines describing why and how to evaluate, and our shared ambition to promote broader reflections through evaluative thinking. The latter, for instance, involves reflections on evaluation in relation to other forms of information or knowledge produced for the sector as well as evaluation as a broader societal phenomenon (Høydal, 2019). It also implies insights into the potential power embedded in evaluations and evaluation processes as well as the limits of this knowledge production (Høydal, 2021). Knowledge, we believed, was crucial for understanding the role of evaluation in relation to other tools and processes used in the public sector.

The course takes form

After defining common ground, we started to develop a theory of change for the course. Weiss (1995) defines a theory of change as a theory of how and why an initiative works. In this case, the outset was the hypothesis that the training would improve the evaluation capacity in the participants' respective organizations. However, developing the details of the theory was more complicated. One issue was related to who the potential participants would be. Who would be our target group, and who would like to come? Since the course involved a participant fee, the number of participants from each organization would be limited. From DFØ's former experience, civil servants with a particular interest in evaluation would always turn up, and this group typically belongs to development and research units.

Based on DFØ's concern about the confusion regarding the relationship between evaluation and other tools used for performance management, we decided to primarily market the course as an opportunity to learn more about evaluation and

commissioning of evaluation in relation to performance management. Consequently, we hoped to reach bureaucrats occupied with evaluation as learning, as well as control and management. Based on our focus on capacity building, we stressed that the participating organizations would benefit more from the course if they sent two or more participants.

In the end, we established a simple theory of change that could work for a mixed group of participants:

Problem:

- Limited use of evaluations in public performance management.
- Limited or incomplete understanding of what evaluation is.
- Limited practical knowledge related to commissioning and use of evaluations.

Activity:

- Four-day training course.

Output:

- Completed course.

Outcome

- Increased knowledge of and interest in initiating and using evaluations in one's own agency/ministry.
- Increased understanding of the possibilities and limitations of evaluations in performance management.
- Increased knowledge of how to plan and commission evaluations.

Impact

- Improved evaluation capacity of key personnel in ministries and agencies. The enhanced capacity will then lead to increased institutionalization, improved quality, and better commissioning and use of evaluations, and thus to increased transparency and accountability, and ultimately to more effective resource use.

Assumptions

- Relevant lectures, knowledge sharing, and group discussions.
- Active interest and involvement of participants.

To reflect the theory of change, the course was organized into four modules. Module one was aimed at giving the participants a broad introduction to evaluation as a method and societal phenomenon (Høydal, 2019) and the challenges of evaluating in the public sector. In this session, the participants were introduced to evaluation from all four perspectives: the international evaluation expert, the experienced evaluator, the researcher in the field of evaluation, and DFØ.

The next two modules were focused on the commissioning of evaluations and the data gathering/reporting phase. In these modules, the participants were also

provided with practical guidelines, for instance, on how to develop a program theory/theory of change and how to design an evaluation. In module four, the focus was back on a more general level when the topic was the use of evaluations in the public sector. Here, we focused on the different forms of use of evaluation results, the communication of results, and the use of evaluations in relation to performance tools. All modules were structured with group activities to engage the participants and facilitate the application of their newly acquired knowledge.

Lessons and reflections

To do the right thing

People who work in public administration want to do things right. They want to master their daily tasks and the daily operation of the organization in a good way. They want to follow recommendations and standards and get clear and useful answers. At the same time, the situations and tasks they are to handle are always complex and consist of various instances with different expectations that are set for them and that cannot be met simultaneously (benefit realization/utility, insight work, and compliance/reliability, to mention a few). It requires experience and insight to make clear priorities in the process of commissioning evaluations, and it takes courage to balance political and professional interests.

In addition, it requires a willingness from managers and commissioners to dialogue, to ask naive questions, and to try and fail, which are often very demanding to handle, while at the same time showing that they are responsible, have control, and maintain daily operations. One of the most important functions of the course was to open up for such an experience-based dialogue. This was to be done, not least, by the four course leaders participating in the entire course all the time, representing different roles and viewpoints and demonstrating disagreement and uncertainty in their own presentations and plenary discussions.

The administration is not one

The government administration is large and complex, and the participants in the course were in different situations. From the large welfare sectors with extensive external deliveries related to income security and services related to work or children, and where much of the service production takes place in municipalities that the state cannot directly control, to small enterprises which wanted to streamline their own governance and organization. The former have large research and development units with employees experienced in ordering external research and evaluations, as well as large administrative units primarily geared at performance management tools. The participants from such organizations understood the need for interaction between knowledge production and performance management, while smaller organizations sought assistance to think better about how knowledge could support efficiency and reduce resistance to change in their own employee professions. It was demanding to develop a course to meet such diverse needs.

What is evaluation?

Our ambition was to give room for evaluative thinking by creating a setting open for curiosity, critique, and collective reflections. In line with this premise, we sought to establish openness for the many different forms, functions, and uses of knowledge and information in public administration and less rigidity related to such important aspects. We stressed the meaning of context in relation to evaluation and how the understanding of evaluation has varied through the past decades (Dahler-Larsen, 2011). However, we also presented the official evaluation definition provided by DFØ and guidelines for public evaluations. While some enjoyed the broadening of their evaluation horizon, others became frustrated, and a fundamental question repeated by participants throughout the course was *what is evaluation really?* This uncertainty was reiterated by some in the evaluation of the course. On the one hand, this persistent confusion could be due to evaluation being a phenomenon that is difficult to grasp. On the other hand, it could be a negative consequence of our attempt to broaden the participants' understanding of evaluation. It's possible that the participants' prior knowledge influenced the extent to which they benefited from our open approach and that participants who already had some evaluation experiences were best equipped to make use of the more holistic perspective presented.

How should I evaluate?

In line with DFØ's prior experience and the feedback from the pilot, many participants were very interested in "step by step guidance". Our ambition was to teach the participants that even though guidelines could be useful, every evaluation has its own individual characteristics and challenges that must be handled accordingly. Hence, we sought to teach the participants about the different tasks and challenges during the different phases of an evaluation process, without presenting quick-fix recipes. However, we did inform about different aspects to focus on during the different stages of an evaluation process, and we informed about more practically oriented online resources. The participants were also given a presentation on procurement regulations by an expert from DFØ. Our strategy to avoid "step by step guidance" partly seems to have failed. In the evaluation of the courses, participants still asked for such information and "more practical knowledge".

Do we need to build an evaluation system?

We believe that having some sort of evaluation system in place in organizations will improve the implementation of the evaluations (Martinaitis, Christenko, and Krauciuniene 2019). A system helps clarify what the evaluation activities include (roles, responsibilities, etc.) and how evaluations fit in relation to other functions in learning and performance management. At the same time, we wanted to acknowledge the limitations of establishing a rational system and pay sufficient attention

to norms, routines, and beliefs. Consequently, we talked about the risks that a too-rigid system may increase the focus on metrics, risk-aversion, conformism, and a compliance mindset. Again, we stressed the need to encourage behaviors that promote autonomy, flexibility, and critical reflection (Dahler-Larsen and Raimondo 2024). At least some of the participants expressed appreciation of this approach. Once again, we believe the more experienced participants to be the most positive in this respect.

Did we manage?

The aim of the course was to build the evaluation capacity of the participants and the agencies and ministries they represented through evaluative thinking. In our theory of change, we hoped for an outcome related to

- Increased knowledge of and interest in initiating and using evaluations in one's own organization.
- Increased understanding of the possibilities and limitations of evaluations in performance management.
- Increased knowledge of how to plan and commission evaluations.

Despite the critique previously mentioned, the course evaluations all the participants gave at the end were generally positive. However, we gained a more accurate impression of the actual outcomes during follow-up discussions with participants two months after the course. Seven people participated in a follow-up meeting, representing five different organizations with varying levels of experience in evaluation, ranging from very limited to advanced. Despite this variation, all participants had utilized the course content in various ways. The participant with no prior experience had reflected on “evaluation and how to evaluate” specific cases or aspects in their work. Another participant had conducted a thorough evaluation of a seminar in their leadership group, and the results had been constructively debated. Two participants from another organization believed the course to be very useful. However, despite support from their closest leader, they were frustrated due to the top leaders' lack of interest in implementing evaluation in management. This led to a meaningful discussion where relevant experiences and strategies on how to overcome such organizational hindrances were shared. The most experienced participant, who had previously been part of a successful project where evaluation became integrated into management, aimed to replicate this success in her new role but had not yet succeeded. From her perspective, the course had validated her ideas, reinforced her perspective, and strengthened her drive in her current project.

These short glimpses of individual experience gave us the impression that the course had indeed led to an increased interest in and understanding of evaluation. It has also led to practical use on different levels, as well as reflections and discussions related to knowledge use in administration and, more specifically, related to evaluations in performance management.

Conclusions

Some lessons may be drawn from this effort to build evaluation capacity in the Norwegian central government administration. One lesson relates to the chosen approach to focus on evaluative thinking rather than on evaluation tools and techniques. A second relates to the attempts to refer to international evaluation standards and criteria in the Norwegian context. The third lesson is about building evaluation capacity to promote democratic values.

When it comes to the first lesson, we wanted, as described in the introduction, to stimulate critical, evidence-based, and independent reflections by promoting broad reflections related to evaluative thinking. This was also reflected in the organization of the course, where we presented ourselves as facilitators, representing different perspectives, rather than a unified group of experts. When some participants kept asking for guidelines or asking about the core of the evaluation phenomenon, it could, on the one hand, be seen as a sign of our strategy being a failure. On the other hand, it could be seen as an expression of learning or a reflection process starting from a very basic outset. We believe the fact that participants kept asking at least reflects the success of our ambition to make room for dialogue and openness.

The second lesson is related to the relationship between the international and the Norwegian approaches to evaluation, reflecting the gap between international organizations' ideas about evaluation and national values and interests (Schwandt, 2009). The course itself gradually made us more reflective and articulate around this fact, and this led to a growing acknowledgment of the value of both paradigms – in combination in a Norwegian context – a point of view we tried to bring to our participants.

A third lesson is related to the overall theme of this book; how evaluation contributes or potentially contributes to promoting democratic values and practices. Evaluation is currently closely associated with the results-based management tradition of measuring and reporting on the achievements of intended objectives of an economic transfer, a grant, or a public sector reform initiative. We aimed to expand the understanding of evaluation beyond merely being another tool for measurement. On the one hand, we emphasized the understanding of not only the possibilities but also the limitations of evaluation throughout the course. This included discussing topics like evaluation and power, as well as the relationship between evaluators and the commissioners of evaluations. On the other hand, we emphasized evaluation as useful for promoting reflection and critical thinking. We tried to widen the participants' perspective on the use of evaluation. To understand that, evaluation should also be understood as a means to uncover abuse of power, misuse of funds, and negative consequences for vulnerable user groups. We think that at least some of the participants started to understand this potential for evaluation.

By spending time together – participants and organizers, representing different organizations, perspectives, and levels of knowledge and experience related to evaluation – we believe the course contributed to an open dialogue, a dialogue we believe to be essential for democratic values and practices.

The overall success of the training course will rely not only on how well we managed to conduct the training, but also on several factors within the participants' own organizations and, not least, what the central actors in the government administration (e.g., the Ministry of Finance and the DFØ) will do in the future to institutionalize evaluation and evaluation use. In the near future, there may be a need for clearer priorities in the Norwegian public sector, and evaluation can then play an important role. However, whether evaluations are used wisely will depend on the bureaucracy's knowledge and understanding of evaluation.

Notes

- 1 DFØ administers the regulations on financial management in government institutions and provides advice and guidance on governance, socio-economic analysis, administration, and management.
- 2 The Agency for Development Cooperation has been an exception until recently by hosting an independent evaluation unit responsible for evaluating all aspects of Norwegian international evaluation cooperation. This unit is now outsourced to a smaller public organization in another part of the country.

References

- Baker, Anita, and Beth Bruner. 2012. *Integrating Evaluative Capacity into Organizational Practice*. Cambridge, MA: The Bruner Foundation. www.evaluativethinking.org/docs/Integ_Eval_Capacity_Final.pdf.
- Bastøe, Per Øyvind. 2020. *Fem forutsetninger for gode evalueringer (Five Prerequisites for Good Evaluations)*. Oslo: Gyldendal.
- Bogensneider, Karen, and Thomas Corbett. 2010. "Family Policy: Becoming a Field of Inquiry and Subfield of Social Policy." *Journal of Marriage and Family* 72(3): 783–803.
- Boyle, Richard, and Donald Lemaire (ed). 1999. *Building Effective Evaluation Capacity*. New Brunswick and London: Transaction Publishers.
- Campbell, John, and Ove Pedersen. 2014. *The National Origins of Policy Ideas: Knowledge Regimes in the United States, France, Germany, and Denmark*. Princeton: Princeton University Press.
- Dahler-Larsen, Peter. 2011. *The Evaluation Society*. Stanford, CA: Stanford University Press.
- Dahler-Larsen, Peter. 2013. Evaluering i Norge set udefra. In *Evaluering. Tradisjoner, Praksis, Mangfold (Evaluation. Traditions, practice and diversity)*, edited by A. Halvorsen, E. Lier-Madsen and N. Jentoft. Bergen: Fagbokforlaget.
- Dahler-Larsen, Peter, and Estelle Raimondo. 2024. "The Sceptical Turn in Evaluation and What to Do with It." *Evaluation*, 30(1): 69–81. <https://journals.sagepub.com/home/evi>.
- DFØ – Direktoratet for økonomistyring. 2016. *Bruk av evaluering i statlig styring (Use of Evaluation in Public Sector Management)*. Oslo: DFØ.
- Duffy, Deirdre Niamh. 2017. *Evaluation and Governing in the 21st Century: Disciplinary Measures, Transformative Possibilities*. Hampshire: Palgrave Macmillan.
- Finansdepartementet. 2003. *Reglement for økonomistyring i staten. Bestemmelser om økonomistyring i staten*. Fastsatt 12. desember 2003 med endringer, senest 5. november 2015.
- Finansdepartementet. 2005. *Veileder til gjennomføring av evalueringer (Guide to Evaluation)*.

- Haugnæss, Gitte, and Ida Gram. 2017. "Brukes evalueringer i statlig styring? (Are Evaluations Being Used in Public Sector Management)." *Stat & Styring*, nr. 1. Universitetsforlaget.
- Høydal, Øyunn Syrstad. 2019. *Evaluering – symbol og metode (Evaluation – Symbol and Methods)*. OsloMet Avhandling nr. 9. Oslo: Oslo Metropolitan University.
- Høydal, Øyunn Syrstad. 2020. "Worlds Apart or Coproduction? Investigating the Relationship between Knowledge Users and Producers." *Evaluation* 26(4): 404–421.
- Høydal, Øyunn Syrstad. 2021. "Evaluation, Framing and Power." *Scandinavian Journal of Public Administration*, 25 (3–4).
- Martinaitis, Žilvinas, Aleksandr Christenko, and Lina Kraučūnienė. 2019. "Evaluation Systems: How Do They Frame, Generate, and Use Evidence?" *Evaluation* 25: 46–61.
- OECD. 2020. *Improving Governance with Policy Evaluation: Lessons from Country Experiences*. OECD Public Governance Reviews. Paris: OECD Publishing. <https://doi.org/10.1787/89b1577d-en>.
- OECD. 2022. *Recommendation of the Council on Public Policy Evaluation*. OECD/LEGAL/0478.
- Patton, Michael Quinn. 2006. "Evaluation for the Way We Work." *Nonprofit Quarterly* 13(1): 28–33.
- Preskill, Hallie, and Shanelle Boyle. 2008. "A Multidisciplinary Model of Evaluation Capacity Building." *American Journal of Evaluation* 29(4): 443–459.
- Schwandt, Thomas A. 2009. "Globalizing Influences on the Western Evaluation Imaginary." In *The SAGE International Handbook of Educational Evaluation*, edited by Jack Smith and Philip Riley. Thousand Oaks, CA: Sage.
- Stockdill, Stacey Hueftle, Michael Baizerman, and Donald W. Compton. 2002. "Toward a Definition of the ECB Process: A Conversation with the ECB Literature." *New Directions for Evaluation* 93: 1–25.
- Tornes, Kristin. 2012. "Evalueringstradisjoner." In *Evaluering i teori og praksis (Evaluation in Theory and Practice)*, edited by K. Tornes, 71–97. Bergen: Fagbokforlaget.
- Vedung, Evert. 2009. *Utvärdering i politik och förvaltning*. 3rd ed. Lund: Studentlitteratur.
- Vedung, Evert. 2010. "Four Waves of Evaluation Diffusion." *Evaluation* 16(3): 263–277.
- Vedung, Evert. 2017. *Six-Times-Two Uses of Evaluation*. Uppsala: Evert Vedung's Personal Archive.
- Wade, Jay, and Leanne Kallemeyn. 2020. "Evaluation Capacity Building (ECB) Interventions and the Development of Sustainable Evaluation Practice: An Exploratory Study." *Evaluation and Program Planning* 79: 101777.
- Wandersman, Abraham. 2014. "Getting to Outcomes: An Evaluation Capacity Building Example of Rationale, Science, and Practice." *American Journal of Evaluation* 35(1): 100–106. <https://doi.org/10.1177/1098214013500705>.
- Weiss, Carol H. 1995. "Nothing as Practical as Good Theory". In *New Approaches to Evaluating Community Initiatives*, edited by James P. Connell, Ann Kubitsch, Lisbeth B. Schorr, and Carol C. Weiss. New York: Aspen Institute.
- Winckler Andersen, Ole. 2014. "Some Thoughts on Development Evaluation Processes." *IDS Bulletin* 45(6): 77–84.

14 Academy of Evaluation

A capacity-building initiative for
commissioning authorities

*Karol Olejniczak, Marlena Stępień
and Piotr Strzęboszewski*

In memory of Marek Kozak, a friend and colleague dearly missed.

Introduction

Evaluation units (EUN) are key commissioning authorities in the context of evaluation related to the European Union's Cohesion Policy (CP). They shape, oversee and coordinate the agendas of evaluative activities for specific policies, programs or agencies. EUN, acting as commissioning authorities, initiate evaluations and supervise the execution of individual studies, often run part of the analyses, and feed study results back into the policy-decision process. EUNs are relatively common; they operate in the context of the EU Cohesion Policy (Kupiec et al., 2023), US federal programs (see: US Congress, 2018), international agencies (Raimondo, 2018) and various policy-specific domains (Smith, 1984). However, the pivotal role of EUN as commissioning authorities that co-shape merit and the use of evaluation is overlooked in the current literature (see Chapter 1: Stame & Bustelo, 2027).

This chapter focuses on the case study of the Academy of Evaluation (AE) – an educational effort to build capacity among commissioning authorities in Poland. AE was a one-year postgraduate free-of-charge university program dedicated exclusively to civil servants (co-financed by European Union technical assistance). The Polish National Evaluation Unit (NEU) and the University of Warsaw implemented eight program editions from 2008 to 2017, reaching over 250 staff from over 50 Polish national and regional evaluation units and civil servants not directly related to evaluation activities. The case of the AE is an interesting example of a systemic effort to build competencies among commissioning authorities. It offers insights into evolving ideas among Polish policy practitioners and academics on what capacities are needed for effective commissioning of evaluation and what methods are suitable for teaching personnel of commissioning authorities. The implementation of the AE ended in 2017 because most of the direct addressees in the Polish CP administration had been trained, and there were organizational changes within the institution that coordinated the implementation of this initiative.

DOI: 10.4324/9781042000968-17

This chapter has been made available under a CC-BY-NC-ND license.

The chapter is based on the following group of sources: (1) a review of the literature on evaluation and public administration in Poland; (2) the detailed documentation from eight editions of AE (the content of tenders and offers for AE, final results of each edition, regular post-classroom surveys with participants, supervisors' observations from classrooms, notes from review meetings, etc.); (3) two evaluations executed during and at the end of AE implementation (Muszyńska et al., 2011; Śliwowski, 2018); and (4) our recollections of the events. Our case study is not an objective analysis since two of us were directly involved in running the AE.¹ Rather, it is the contemplation of "reflective practitioners" on lessons learned from the journey. We attempted to mitigate our subjectivity in two ways. First, we brought the third author to the conversation – Marlena Stępień, a young expert who was not involved in AE but is currently operating in the Polish evaluation context. Second, we implemented a structured process to reflect on the main strengths, weaknesses and factors influencing the academy.

We hope this chapter contributes to the ongoing discussion on evaluation capacity building among commissioning authorities for co-producing quality evaluations and utilizing results effectively.

The context and origins of Academy of Evaluation

Polish system of evaluation

Before the 1990s, Poland lacked public sector evaluation. Modern administration began after Poland's democratic transition (Olejniczak, 2013). External donors introduced evaluation practices, though studies remained scarce (Żuber & Bienias, 2008). Poland's 2004 EU accession required the implementation of strategic planning and evaluation practices. The EU provided funding and assistance to enhance public administration capacity. The EU has driven Poland's evaluation culture development, aligning with European CP financial perspectives (see Table 14.1). The Polish evaluation system has three key features: predominant external evaluation through contractors, domestic consulting firms as evaluators and gradual influence of EU standards on national policies.

The beginning of AE and its formal arrangements

The AE initiative arose from three enabling factors. First, a favorable environment existed for government-academia cooperation on policy and administration capacity building. EU regulations and funding created an opportunity, emphasizing evaluation in CP implementation (Council of the European Union, 2006, articles 47–49) and advising capacity-building efforts. Thus, the activities of the Academy could be co-financed by the European Union through the technical assistance programs.

The second element was a proven framework for long-term training, created with academia. The Academy of Regional Development, a joint initiative by the Ministry of Regional Development and EUROREG – University of Warsaw, started

Table 14.1 Summary of the evolution of Polish evaluation system

<i>Period</i>	<i>System architecture</i>	<i>Scope of operations</i>	<i>Key challenges and milestones</i>	<i>Number of evaluations</i>
2004–2006	Formation of the system: creation of EUN, development of procedures 7 units	Limited in scale but foundational development of the system	-NEU designated to coordinate all development strategies and programs in 2006 -Focus on setting up the structure -basic mix of qual-quant methods used in studies	84 (60 in 2006)
2007–2013	Expansion: regional units established, further decentralization 57 units	Rapid growth in the number and robustness of evaluations.	-Counterfactual methods introduced and popularised -NEU competencies expanded -Strong focus on results-oriented policies and impact evaluations	883 (158 in 2010)
2014–2020	System stabilized, though some loss of momentum in evaluation culture 33 units	Fewer evaluations conducted, with some EUN closed.	-Decline in annual number of evaluations -Units decreased from 57 to 33.	518
2021–2027	The system remains stable. Some of the positive practices have stopped	The system remains decentralized with 33 EUN still operating.	-Ongoing implementation, though decline in annual evaluations noted -More co-design approaches used in study developments	240

in 2005 as postgraduate studies on regional policy for public administration. This model served as a blueprint for establishing a new evaluation academy.

The third factor was the growing demand for evaluation education, which focused on the commissioning authorities' perspective. In 2008, during the planning of AE's first edition, the evaluation system was in the early stages of growth. New commissioning authorities were established to conduct evaluation studies, staffed by young individuals with academic training but lacking practical experience in evaluating public policies. The AE aimed at addressing this gap.

The first edition of the AE was conducted through a free-hand procurement procedure, allowing discussions among public administration employees, experts and academics on developing an innovative study program. The procurement aimed to organize workshops on evaluating socio-economic development programs titled "Academy of Evaluation of Socio-Economic Development Programs" for central and regional institution employees. The workshops totaled 170 hours over 12 two-day sessions in Warsaw, with 50 expected participants.

The tender's objective was to provide current evaluation theory and practice, emphasizing evaluation's importance for improving public interventions, particularly for EU CP goals. The training aimed to teach participants to use evaluation as a learning and management tool for regional development. This tender first identified core competencies needed for efficient evaluation implementation. Annual tenders for single editions provided both a risk of discontinuity and an opportunity for continuous program adaptation.

Academy's vision, participants and merit

The Theory of Change for AE

The terms of reference for the consecutive editions of the AE were translated into three main objectives that the program pursued over the years:

- 1 embedding evaluation within regional development and management by introducing contemporary theories and public sector management standards;
- 2 fostering a multidisciplinary discussion of various evaluation research concepts, methodologies and approaches based on both domestic and international experiences; and
- 3 enhancing the practice of evaluation research functions in the public sector by studying both domestic and international cases.

Table 14.2 provides a generic Theory of Change (ToC) for the AE based on all editions' documentation. The first ToC version emerged after the 2nd edition of the AE. The ToC aimed to articulate AE's vision and role in the evaluation system and plan changes in new editions for coherence and continuity. During the first classes, it also served as a canvas for participants to map their expectations about knowledge gaps, training needs and realistic anticipation of AE's career impact.

Specific recruitment strategies, learning methods and teaching content selection followed those assumptions. We discuss them in more detail in the following sections.

The participants of the Academy

The primary information source on AE participants and learning outcomes is the 2018 study with summaries of participants' surveys conducted during each AE edition and follow-up on alums (Śliwowski, 2018). Surveys conducted during each edition were obligatory for all participants.

AE started with a target group of potential commissioning authorities of EU CP evaluation in Polish central and regional public administration – personnel of EUN, monitoring and strategic planning departments for EU programs. Over the years, the AE has evolved, attracting participants with high managerial roles in national programs and people with extensive experience coordinating evaluation projects. That led to a more specialized focus. Despite this trend, participants' profiles remained diverse, with varying experience levels in the evaluation field present in each edition.

Two hundred fifty-six individuals participated in the eight editions of the AE. Half were ministry employees, predominantly from the Ministry of Regional Development (44 people), and numerous participants were from the Prime Minister's Office (28). One-fifth of the participants worked at the Marshal's Office or its units. Thus, most participants came from the central administration and regional levels of self-government.

Sixty-seven percent of all participants were women, mirroring the trend in Polish public administration.² Most participants were from Warsaw (83%), likely because the classes were held there. The average participant age was 32, indicating a young demographic, with two-thirds under 40, the typical peak professional age. Thus, the AE knowledge and skills could benefit individual careers and institutional operations.

Surveys from the fourth to the eighth editions revealed that 52% had no experience in commissioning or conducting studies, 32% had commissioned studies, 10% had conducted studies, and only 16% had done both. These figures highlight the Polish evaluation system's focus on external evaluations over internal ones. Moreover, 83% of participants said the AE was their sole educational experience in evaluation, and 49% had no prior educational exposure to the subject.

Participant motivations were equally varied, as surveys from the fourth to eighth editions showed. About one-third attended to acquire new skills for daily work related to CP evaluation. Others aimed to enhance skills in project and strategy evaluation areas such as control, auditing, monitoring, reporting and regulatory impact assessment or to develop competencies relevant to public sector roles. Thus, the AE attracted civil public sector personnel who worked as commissioning authorities for evaluations, but also wider socio-economic analyses used in government decision-making. This diversity fostered a unique participant group interested in evaluation, underlying the program's broad applicability.

Table 14.2 The Theory of Change for AE

<i>IF... we organize good learning process for right participants...</i>	<i>THEN... we will trigger positive cognitive and behavioral changes among participants...</i>	<i>AND THAT... would lead to better evaluation practices in our public institutions.</i>
Right participants Participants already have some experience in evaluation. Participants have intrinsic motivation for participation. Participants are embedded in a supportive organizational environment.	Cognitive change Increase in individual knowledge and skills on the latest evaluation expertise. Strengthen the number of evaluation experts in the Polish public administration system.	Micro-organizational change AE alumni share their expertise with the direct team in which they work. The unit where AE alumni work improves their way of working, and their approach to E&M tasks.
Good learning process The merit is of high quality and relevant to participants' work. The teaching is based on effective educational strategies and methods for adult learners.	Behavioral change Alumni start using the gained knowledge and skills in their daily work Alumni gain confidence in participating in the evaluation-related activities of their organizations and the Polish public sector. Alumni continue to develop networks among like-minded practitioners of evaluation.	Structural change The improvement of practices at the level of departments and institutions. The improvement of practices in various sectoral and functional areas (e.g., policy themes, fields of action). AE alumni maintain contact with other AE graduates.

Source: Evaluation report of editions 1 and 2, mind maps from AE participants.

Curriculum for commissioning authorities

Over eight years, the AE’s curriculum evolved. Initially, it emphasized tools and skills for drafting evaluation tenders for specific programs. Later editions expanded to cover various applications and approaches for sectoral evaluations. The last editions adopted a holistic approach to orchestrating evaluation and other evidence-based activities in government. The curriculum recognized that EUN act as change agents and facilitators of organizational learning processes, eventually adopting the term “knowledge brokers” (Olejniczak et al., 2016).

Definition of evaluation units as knowledge brokers

Knowledge brokers are intermediaries that steer flow of research evidence between knowledge producers (researchers, experts, evaluators) and knowledge users (decision-makers, public managers). They are instrumental for

connecting the demand for evidence-based knowledge with its supply in the ecosystem of government decision-making.

Source: Based on: Dobbins et al. (2009), Frost et al. (2012) and Olejniczak et al. (2016).

The AE's curriculum was an ongoing attempt to address the practical question for educating commissioning authorities: **What essential knowledge and skills are needed to be an effective knowledge broker in public administration?**

Insights from literature reviews, discussions with practitioners, debates from major evaluation associations brought us to the conclusion that knowledge brokers in the context of Polish public administration must: (1) understand government decision-making dynamics, (2) formulate practice-oriented research questions (including evaluative questions), (3) comprehend the credibility of social research design and methods, and (4) disseminate knowledge by building an evidence base for policy discussions, networking, communicating findings. Table 14.3 summarizes the knowledge broker curriculum.

Methods of teaching adults professionals

The teaching team included leading national and international experts in evaluation and public policies from academia, research institutions, public administration and international organizations. International experts introduced emerging ideas, sectoral experts shared national practices, and EUROREG researchers facilitated workshops and theory-to-practice sessions. This approach allowed integrating diverse conceptual and methodological innovations into Polish context in a coherent way.

The Theory of Change for the AE assumed that teaching, in order to be effective, needs to use educational methods for adult learners. That is why the didactic approach was based on four pillars from adult learning literature. The first was **backward design**, where educators initially identified the course's core goals – what students should achieve and its significance. Course activities were then built around these core elements and connected to the final assignment (Lang, 2016; Tokuhama-Espinosa, 2021, 42–60). AE program centered on knowledge brokers and core competencies (see Table 14.3). Additionally, participants' final assignment – a draft of a commissioned evaluation – was integrated and initiated in Week 1 of the course.

The second pillar, **Elaboration Theory**, states that learning content should be structured from general to complex. This approach uses zoom lenses: starting with an overview to understand how parts relate, then focusing on each part in detail, and finally zooming out to show connections again. This gist-to-details method significantly enhances retention (Kirschner and Hendrick, 2020, Chapter 16; Medina, 2021, 59). At AE, we began with the overarching concept of evidence-informed policies and evaluation in modern public policies, and then delved into

the processes of designing, conducting and utilizing evaluations. To manage complexity, AE participants used mind maps for note-taking, designing and discussing their commissioned evaluation ideas.

The third pillar of the AE was **the Experiential Learning Model (ELM)**, grounded in trial-and-error learning. Learners engage in a cyclical process of experiencing, reflecting, thinking and acting, allowing them to adjust their mental models based on practical experiences, leading to continuous relearning (Kolb, 2013). In AE practice, this translated into extensive workshops and practical exercises following lectures to apply theoretical concepts to real problems. Applied tasks constituted 70% of teaching time. ELM was implemented through interactive methods, including case studies relevant to participants' contexts, debates, discussions and simulations. We even developed a serious board game to simulate typical activities of EUN for regional development programs (Olejniczak, 2017).

The final component of our didactic strategy facilitated three interaction types during courses: student-student (SS), student-teacher (ST) and student-content (SC). These interactions correlate with enhanced achievement and student satisfaction (Bernard et al., 2009). On AE, participants consistently worked in small teams (SS), received supervision and engaged in discussions with instructors (ST) and accessed readings and practical materials before, during and after classes (SC).

The effects and impact of the Academy

The impact on participants

The AE evolved over its various editions, leading to changes in graduate surveys and learning outcomes. While the profiles and competencies of graduates varied, the AE's impact can be illustrated with two core areas: understanding intervention logic and acquiring reliable knowledge.

The first area, understanding intervention logic, was focused on organizing public management concepts through mind maps, which participants rated highly across all editions. However, despite positive evaluations, participants found it difficult to transfer this knowledge to their daily work, highlighting the challenge of integrating new tools into public administration.

The second area focused on practical skills, such as formulating evaluation and research questions and coordinating studies. Workshops on Public Procurement Law and research methodology were particularly well-received, although attempts to increase sophistication (e.g., introducing advanced methods) faced challenges due to time constraints and language barriers.

Career impact data is limited, but we have some insights on how the knowledge gained is passed on to the professional responsibilities of AE graduates. The follow-up study of alumnus (Śliwowski, 2018) noted that while many graduates took on operational and management roles, 6% transitioned to the private sector. Most graduates (56%) are involved in identifying socio-economic challenges and evaluating policy implementation, with some also working on project design and evaluation.

Table 14.3 AE – generic curriculum (synthesis of eight editions)

<i>MODULE 1: Understanding the context and logic of policy interventions (50 academic hours)*</i>	<i>MODULE 2: Identifying knowledge needs (40 academic hours)*</i>	<i>MODULE 3: Obtaining credible knowledge (70 academic hours)*</i>	<i>MODULE 4: Sharing knowledge with users (30 academic hours)*</i>	<i>Exam session (15 academic hours)*</i>
Learning goals: (1) Understand the broader context in which brokers operate. This refers to the decision-making dynamic, organizational, and regulatory environment of evidence-informed public policies. (2) Comprehend the object of evaluation activities – public interventions. This involves unpacking the Theories of Change that underlie public policies, programs, projects, and regulations.	Learning goals: (3) Identify and articulate the knowledge needs of different knowledge users. The spectrum of users includes senior decision makers, public managers, and policy stakeholders. (4) Develop skills in formulating clear research questions, and prioritizing knowledge needs (need to know vs. nice to know).	Learning goals: (5) Grasp the spectrum of research designs and methods for acquiring credible knowledge for public policy purposes. (6) Learn the logic of evaluative assessment, ethical and effective conduct of evaluation contracts.	Learning goals: (7) Explore the spectrum of activities and processes brokers can proactively plan to increase the use of evidence in public policy. (8) Explore effective ways to share evidence with various knowledge users in the policy process. That means choosing appropriate forms and channels of communication and combining the results of various studies into evidence streams.	Objectives of the exam: (a) Test the factual knowledge of the individual participants (b) Test how they apply knowledge and skills in real-life scenarios, working as a team.

(Continued)

Table 14.3 (Continued)

<i>MODULE 1: Understanding the context and logic of policy interventions (50 academic hours)*</i>	<i>MODULE 2: Identifying knowledge needs (40 academic hours)*</i>	<i>MODULE 3: Obtaining credible knowledge (70 academic hours)*</i>	<i>MODULE 4: Sharing knowledge with users (30 academic hours)*</i>	<i>Exam session (15 academic hours)*</i>
Courses in Module 1:	Courses in Module 2:	Courses in Module 3:	Courses in Module 4:	Content of the exam:
(i) Introduction to Evidence-Informed Public Policies	(i) Evaluation and Monitoring cultures and systems in OECD countries	(i) Research designs (RCTs, quasi-experiments, macroeconomic modeling, descriptive studies, case studies, Theory-Driven Evaluations, participatory approaches, simulations)	(i) Types of evidence use and effective strategies for dissemination of results	(i) Knowledge test (multiple choice questions)
(ii) New developments in selected sectoral policies	(ii) Knowledge management in the Polish system of CP	(ii) Research methods and tools – quantitative, qualitative, mixed-methods	(ii) Effective data visualization	(ii) Capstone teamwork project: designing Terms of Reference for evaluation contract (different topics available)
(iii) Theories of Change for public interventions	(iii) Asking research questions, defining concepts and indicators	(iii) Management of evaluation contracts–Commissioning Authority’s and contractors’ perspectives.	(iii) Argument-building and storytelling	
(iv) Knowledge Brokers simulation game			(iv) Writing in plain Polish	
(v) Idea mapping and MindManager				

* 1 academic hour = 45 min.

Graduates frequently performed tasks like searching for evidence, preparing reports and reviewing research assumptions. However, only some are involved in primary research or database management. Regarding key skills, data searching and visualization were identified as critical.

Regarding communication and networking, most graduates collaborate within or between authorities. Only 24% work with target groups, and only some present their research findings to external stakeholders, political leaders or the media.

The study identified three core areas for future development: general public administration skills, analytical capabilities and communication. Notably, research coordination skills were deemed more important than analytical abilities, reflecting the role of public administration in managing rather than conducting evaluations. Communication skills were also undervalued.

The study of alums raised key questions about the future of the similar projects: whom it should educate and who its target audience should be. Options included even more specialized courses for EUN and policy analysts, more generic courses for civil servants or a hybrid model. However, the closure of the AE's program left that dilemma undecided.

The role of the Academy in the system

The AE significantly influenced the maturation of evaluation commissioners and the entire evaluation system in Poland. This long-term program, coupled with developing an evaluation culture under the CP, has substantially enhanced commissioning authorities' methodological knowledge and analytical skills, challenging evaluators (contractors) to use more ambitious methods. One notable achievement is the incorporation of qualitative criteria in bid assessments for selecting evaluation contractors.

The competencies and work of graduates who stayed within the public system influenced the system's overall performance. This impact was visible in three ways (Bartosiewicz-Niziolek et al., 2020; NEU observations from the network dynamics over the years). First, the quality of research plans and terms of reference improved, along with contractor collaboration, through clear expectations and goal-oriented pursuits. Second, the quality and depth of discussions improved at professional meetings, such as Evaluation Steering Groups and proceedings of M&E committees. Third, the cooperation and experience sharing of EUN within the CP system grew in vertical (central-regional level and contractors) and horizontal terms (direct exchange between regional units). Thus, although planned follow-up networking initiatives for graduates did not materialize, the memory of the AE remains strong among participating EUN employees. The knowledge imparted years ago is now well-established and developed.

The AE did not change the structure of evaluation commissioning practices in Poland, which means external dominance over internal evaluations. There are growing arguments for implementing internal or hybrid evaluations (related to, among other things, a shrinking market of contractors), but attempts to implement them are still limited mainly due to the small size of the EUN, which is burdened

with many other tasks. With this in mind, the AE has contributed to improving the quality of the activities of EUN as commissioning authorities.

The final element of the evaluation process – utilizing and promoting study results – remains problematic and undervalued. Issues persist in effectively communicating evaluation results to decision-makers and the public, which is crucial for participatory evaluation. Evaluation reports lack media appeal, and meetings typically involve only public administration representatives, emphasizing the necessity for evaluators to possess communication skills, including networking, result presentation, research promotion and participatory techniques. These abilities contribute to the profile of a modern knowledge broker. While enhancing analytical and methodological skills is vital, the previously identified needs during the AE's implementation are now more pertinent than ever.

Reflections and lessons learned

The primary lesson from years of the AE initiative is that capacity-building activities for commissioning bodies should be viewed as an ongoing, evolutionary process influenced by feedback, stakeholder reflection, environmental changes and varying learner groups.

We found that a comprehensive academic program offers a valuable form of systemic capacity-building for commissioning authorities and broader public administration personnel, serving as a viable alternative to short-term training and summer schools. This approach provides greater stability and depth, making the academic diploma an appealing milestone in the career development of civil servants despite requiring a greater time investment.

Regarding the program's content, there was a clear evolution from the technical skills of "contracting evaluations" to a more holistic perspective of "commissioning knowledge insights". The idea of commissioning authorities as knowledge brokers eventually emerged from adapting the AE curriculum. Knowledge brokering remains a useful framework for addressing the needs and competencies of staff commissioning evaluations. However, it may require updates to address modern evaluation challenges, such as new strategies for addressing the erosion of evidence-based discussions among decision-makers (Marra et al., 2024), new methods for co-designing evaluations with target groups, and the effective and ethical use of Generative Artificial Intelligence (Nielsen et al., 2024).

When reflecting on what makes these types of initiatives possible, we can point out the combination of endogenous and exogenous factors. First, Polish decision-makers and management were open to disseminating evaluation knowledge and developing the system. This openness led to the second aspect – the real collaboration with the academic institution, where dialogue between practitioners and academics shaped the program. The third factor was European funds, which provided essential and sustainable financial support, covering the costs of the academy. This combination of factors enabled the creation of a unique project within public administration, where pioneering ideas are often met with resistance.

Looking more broadly at how this type of initiative contributes to promoting democratic values in evaluation work, we identify four key aspects. Firstly, this initiative has fostered an understanding among commissioning authorities about the critical role of evaluation in supporting democratic decision-making. Public policies need evaluations to establish their merit and worth, and EUNs facilitate evidence-based decision-making with evidence, thereby strengthening democratic processes.

Secondly, the AE empowered commissioning authorities by equipping them with the knowledge and skills to engage meaningfully with experts and researchers conducting evaluations. This empowerment enhanced their ability to shape discussions on policy effectiveness and enabled them to guide external experts, foster constructive critiques and demand higher-quality contracts and outcomes.

Thirdly, at the methodological level, the AE introduced a range of approaches and techniques that reflect democratic principles. These methods enabled more inclusive practices, including co-designing research processes with marginalized groups and stakeholders, thereby ensuring that diverse voices were heard and considered in the evaluation of public policies.

Lastly, the AE contributed to the development of a community of practice, bringing together individuals and institutions dedicated to advancing evaluation and, more broadly, evidence-informed decision-making. This community serves as a critical space for knowledge exchange, peer learning and collaboration, further enhancing the democratic ethos in evaluation work.

The AE also signals the limits of capacity building. Even the best-educated commissioning authorities, working with evaluators, need an audience for evidence – civil society, journalists and decision-makers interested in evidence-informed debates. Thus, when discussing ambitious schemes of evaluation practices promoting democratic ways, we should be aware that the evaluation craft is embedded in a broader socio-institutional context that shapes and often limits our efforts.

Funding information

The authors acknowledge that the activities of the Evaluation Academy were co-financed by the European Union, mainly through the technical assistance program under the 2007–2013 and 2014–2020 financial perspectives.

Acknowledgments

We extend our gratitude to those who made the AE possible. We thank prof. Grzegorz Gorzelak and Piotr Żuber, whose visionary leadership initiated this educational endeavor. Our deep appreciation goes to all participants across eight editions – their dedication and professional commitment were the foundation of this initiative's impact.

We are grateful to the Polish and international lecturers who shared their expertise, bringing together diverse perspectives from academia, practice and

international organizations. Lastly, we would like to thank our colleagues from the National Evaluation Unit and EUROREG, whose administrative work ensured the Academy's smooth operation.

The AE was truly a collective achievement, and we remain honored to have been part of this remarkable journey in building evaluation capacity in Poland.

Notes

- 1 Karol Olejniczak was, together with Marek Kozak, co-supervisor of the program at the University of Warsaw. Piotr Strzęboszewski was the supervisor of the program from the side of the Polish Ministry and National Evaluation Unit.
- 2 According to 2023 data, nearly 63% of public sector employees are women (Statistical Yearbook of Labour 2023, Statistics Poland).

References

- Bartosiewicz-Niziołek, Monika, Beata Ciężka, Weronika Felcis, Katarzyna Hermann-Pawłowska, Seweryn Krupnik, and Piotr Strzęboszewski. 2020. "Poland." In *The Institutionalization of Evaluation in Europe*, edited by Reinhard Stockmann, Wolfgang Meyer, and Lena Taube, 405–33. Cham, Switzerland: Palgrave Macmillan.
- Bernard, Robert, Philip Abrami, Eugene Borokhovski, Anne Wade, Rana Tamim, Michael Surkes, and Edward Bethel. 2009. "A Meta-Analysis of Three Types of Interaction Treatments in Distance Education." *Review of Educational Research* 79 (3): 1243–89.
- Dobbins, Maureen, Paula Robeson, Donna Ciliska, Steve Hanna, Roy Cameron, Linda O'Mara, Kara DeCorby, and S. Shawna Mercer. 2009. "A Description of a Knowledge Broker Role Implemented as Part of a Randomized Controlled Trial Evaluating Three Knowledge Translation Strategies." *Implementation Science*, 4(1).
- Frost, Helen, Rosemary Geddes, Sally Haw, Caroline A. Jackson, Ruth Jepson, John D. Mooney, and John W. Frank. 2012. "Experiences of Knowledge Brokering for Evidence-Informed Public Health Policy and Practice: Three Years of the Scottish Collaboration for Public Health Research and Policy." *Evidence & Policy*, 8(3): 347–59.
- Kirschner, Paul, and Carl Hendrick. 2020. *How Learning Happens. Seminal Works in Educational Psychology and What They Mean in Practice*. London, New York, NY: Routledge.
- Kolb, David A. 2013. "Experiential Learning Theory and Learning Styles." In *Encyclopedia of Management Theory*, edited by Eric Kessler, 277–80. Thousand Oaks: SAGE Publications.
- Kupiec, Tomasz, Dominika Wojtowicz, and Karol Olejniczak. 2023. "Structures and Functions of Complex Evaluation Systems: Comparison of Six Central and Eastern European Countries." *International Review of Administrative Sciences* 89 (1): 202–20.
- Lang, James M. 2016. *Small Teaching: Everyday Lessons from the Science of Learning*. San Francisco: Jossey-Bass.
- Marra, Mita, Karol Olejniczak, and Arne Paulson, eds. 2024. *Evaluation in the Post-Truth World*. New York, NY: Routledge.
- Medina, John. 2021. *Brain Rules for Work. The Science of Thinking Smarter in the Office and At Home*. Seattle: Pear Press.

- Muszyńska, Anna, Barbara Piwońska, Monika Stefanowicz, and Paweł Śliwowski. 2011. *Badanie Ewaluacyjne Akademii Ewaluacji Programów Rozwoju Społeczno-Ekonomicznego*. Warszawa: Ministerstwo Rozwoju Regionalnego.
- Nielsen, Steffen Bohni, Francesco Mazzeo Rinaldi, and Gustav Jakob Petersson, eds. 2024. *Artificial Intelligence and Evaluation: Emerging Technologies and Their Implications for Evaluation*. New York: Routledge.
- Olejniczak, Karol. 2013. "Mechanisms Shaping Evaluation System – A Case Study of Poland 1999–2010." *Europe-Asia Studies* 65 (8): 1642–66.
- Olejniczak, Karol. 2017. "The Game of Knowledge Brokering: A New Method for Increasing Evaluation Use." *American Journal of Evaluation* 38 (4): 554–76.
- Olejniczak, Karol, Estelle Raimondo, and Tomasz Kupiec. 2016. "Evaluation Units as Knowledge Brokers: Testing and Calibrating an Innovative Framework." *Evaluation* 22 (2): 168–89.
- Raimondo, Estelle. 2018. "The Power and Dysfunctions of Evaluation Systems in International Organizations." *Evaluation* 24 (1): 26–41.
- Śliwowski, Paweł. 2018. *Ewaluatorzy Analitycy Brokerzy Wiedzy. Akademia Ewaluacji a Kształcenie Kadr Dla Polityk Publicznych Opartych Na Dowodach*. Warszawa: Ministerstwo Inwestycji i Rozwoju.
- Stame, Nicoletta, and María Bustelo. 2027. "Commissioning Matters: Considerations and Challenges for Evaluation Commissioners Who Intend to Promote Democracy." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Smith, Nick L. 1984. "Evaluation Units in State Departments of Education: A Five-Year Update." *Evaluation News* 5 (3): 37–44.
- Tokuhami-Espinosa, Tracey. 2021. *Bringing the Neuroscience of Learning to Online Teaching. An Educator's Handbook*. New York, NY: Teachers College Press.
- U.S. Congress (2018). P.L. 115-435, Foundations for Evidence-Based Policymaking Act of 2018. (approved: January 14, 2019). H.R. 4174. 115th Congress, 1st Session.
- Żuber, Piotr, and Stanisław Bienias. 2008. "System Ewaluacji W Polsce – Dotychczasowe Doświadczenia I Wyzwania Na Przyszłość." In *Teoria I Praktyka Ewaluacji Interwencji Publicznych. Podręcznik Akademicki*, edited by Karol Olejniczak, Marek Kozak, and Bartosz Ledzion, 84–98. Warszawa: Wydawnictwa Akademickie i Profesjonalne.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Part IV

Conclusions



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

15 Democratic commissioning

Lessons from theory and practice

María Bustelo and Nicoletta Stame

In memory of Thomas A. Schwandt, for his inspiration about the “democratic purpose of evaluation”, values and valuing, Ethos and professionalism. And for the good, inspiring and meaningful moments shared with him throughout the EES Biennial Conferences.

The project for this book began with a discomfort with mainstream commissioning and its predominantly bureaucratic characteristics, such as evaluation capture, think-do and ritualism (Raimondo & Leeuw, 2021). These characteristics had daunting effects that were contrary to the avowed democratic mission of evaluation (Schwandt, 2023). We began investigating the links between commissioning, the demand side of evaluation, and democratic values. Could democratic commissioning exist? What would it be like? We could only answer the first question positively if we could provide evidence for the second question and offer suggestions for more encouraging practices.

Why the uneasiness? First, we had to isolate the commissioning process from institutionalized evaluation systems. How did mainstream commissioning become “bureaucratized” and lose part of its mission? Over the past few decades, much of the evaluation practice has become institutionalized within evaluation systems that establish evaluation as a standard practice within public organizations. Commissioning processes usually take place within these systems and are influenced by them. These systems are intended to promote environments that are favorable to evaluation and useful by making evaluation more integral to decision-making. They also aim to promote transparency by communicating findings and supporting accountability. Additionally, these systems are meant to safeguard the independence and integrity of evaluation (Bustelo & Jacob, 2025; Leeuw & Furubo, 2008).

However, some authors have pointed out “a number of pernicious facets that can accompany the institutionalization of evaluation within bureaucracies” (Raimondo & Leeuw, 2021, p. 143). One problem is that evaluation systems can become an administrative burden, leading to overly rigid structures in which compliance takes precedence over meaningful inquiry (Bustelo & Jacob, 2025). Others argue that evaluation systems “can limit actors’ agency and impose top-down pressure on how and when evaluative activities should be conducted” (Benerdal & Larsson,

2021, p. 64). Another problem is the risk of overstandardization. Institutionalizing evaluation practices should not mean conducting all evaluations in the same way. Uniformity often prevents adaptability across policy contexts and limits the capacity of evaluations to produce nuanced, context-specific insights (Andersen, 2021). Other authors have noted that the shortcomings of these evaluation systems are particularly problematic for democratic approaches to evaluation because they often leave the responsibility of ensuring stakeholder participation and influence to the individual evaluator (Benerdal & Larsson, 2021).

We might conclude that any democratic intention within an evaluation system should consider the role of other participating actors, particularly commissioners, who are important guarantors, enablers and drivers of democratic commissioning processes. Therefore, this volume focuses not on evaluation systems “per se” but on the commissioning process and, specifically, on the often-overlooked role of commissioners and how they can exert their full agency within those systems.

Second, it was necessary to consider the role of commissioners in their relationship with evaluators. While the supply side of evaluation could claim a long tradition of evaluation approaches leaning toward democracy, the topic of commissioning – not to mention democratic commissioning – had just begun to attract serious research attention (Brousselle, 2022; Cox & Barbrook-Johnson, 2021).

When analyzing what commissioning implies within the entire evaluation process (deciding what to evaluate, terms of reference, choosing an evaluator, contracting, management and use), we had to focus on its processual nature. At each step, responsible choices must be made in collaboration with other actors.

To address the deficiencies, and make commissioning compatible with democratic evaluation, we needed to explore to what extent the system in which evaluations are asked could be open to allow democratic practices. We also needed to examine the role attributed to commissioning in each system and the strategies commissioners might undertake to promote democracy.

This is what the studies presented in the previous chapters have done. Then, we could venture theorizing about democratic commissioning. To take this new step in our inquiry, we realized that we could benefit from Barry MacDonald’s pioneering work and its further utilization.

- MacDonald (1977) defined democratic evaluation vis-à-vis other forms of evaluation (bureaucratic and autocratic): could we draw a similar distinction for the commissioning policies of evaluation systems?
- Schwandt (2018) inquired into how evaluation could be considered a form of democratic professionalism (Dzur): could we draw a similar conclusion for a managerial task (Catsambas & Davidson, 2024) like commissioning?

Our experience in writing this book enables us to answer these questions positively. The rest of this chapter will explain why we believe this to be true.

We will proceed as follows: first, we will review the analytical frameworks used to distinguish democratic and non-democratic practices in evaluation and apply them to

the actors involved: evaluators and commissioners. In the second part, based on what we have learned from the previous chapters, we will offer more explicit guidance on what this means for commissioners' competence and behavior in practice.

Part I – Defining democratic commissioning

Democratic evaluation as a policy

Barry MacDonald was the first to theorize what it means for an evaluation – and not just the political regime in which it operates – to be democratic. His pioneering views have inspired broader developments and will also inspire our work. Our proposed approach follows MacDonald's **typology of evaluation studies** based on the supply side, which is determined by whom the **evaluator** is responsible for.

- 1 **Bureaucratic** evaluation occurs when the evaluator accepts the values of those in office and acts as a *management consultant* trying to satisfy their client, who owns the final report. The techniques used must be credible to policymakers, who are mainly concerned with “utility” and “efficiency”. Its key justificatory concept is the “reality of power”.
- 2 **Autocratic**¹ evaluation occurs when the evaluator offers external validation of policy in exchange for compliance with his recommendations. The evaluator is aware of the bureaucracy's constitutional and moral obligations, but his techniques need validation from the academic research community. The key concepts are “principle” and “objectivity”. The key justifying concept is “the responsibility of office”.
- 3 **Democratic** evaluation occurs when the evaluator collects definitions of and reactions to the program, recognizes value pluralism, and acts as a broker of information exchanges between different groups of informed citizens. The report must be accessible to non-specialist audiences and should not contain recommendations. The key concepts are “confidentiality”, “negotiation” and “accessibility”. The key justifying concept is “the right to know” (MacDonald, 1977).

As we have seen in Chapter 1 (Stame & Bustelo, 2027), MacDonald's definition of democratic evaluation has provided the basis for further elaborations on democratic evaluation. Moreover, it resonates with more recent elaborations on democracy, recalled by our authors to locate democratic commissioning. In this way, Rogers & Montague (based on Hanberger, 2006) distinguish between democracy: i. To the people (elitist); ii. For the people (participatory); iii. With the people (discursive). Similarly, Ling & Millard refer to Fishkin (2009)'s deliberative democracy.

While MacDonald's typology is based on the supply side of the evaluation, it can be utilized also when we look at the counterpart, that of **the demand side**, that is, commissioning.

In Chapter 3, Rogers & Montague (2027) have identified three basic situations where evaluations are demanded, which we can consider as forms of **evaluation policy**:

1 **undemocratic** evaluation policy when evaluation is:

- absent: in dictatorial regimes, or in specific sectors (like fossil fuel subsidies), there is no demand for evaluation.
- token: reviews undertaken by external agencies are done for compliance purposes only, with limited public access to the findings or scrutiny of the process.

2 technocratic evaluation policy² that focuses on providing information to elected officials so they can make transparent, evidence-informed decisions.

3 **democratic** evaluation policy that explicitly supports active dialogic and deliberative engagement with communities or supports their self-determination.

Our chapters on commissioning experiences (second section of this volume) provide examples of situations characterized by these different kinds of evaluation policies.

Some chapters analyze evaluation policies of the second type, their different aims and results: Chapter 10's analysis (Laaksonen & Palenberg, 2027) of the Evaluation Management Services of the Finnish Ministry of Foreign Affairs (MFA) shows how it created structured processes that made the provision of evidence more transparent. The MFA policy for the creation of the Evaluation Management Services (EMS) was motivated by the concern for methodological quality and the need for coordination of centralized evaluations, that are considered crucial for learning and policy planning. The authors, however, observe that, notwithstanding the well-oiled machine of the EMS, the whole process of knowledge management was wanting; it was not possible to establish a comprehensive database as part of EMS, which would have helped avoid repeating requests for information from previous centralized evaluations. In its turn, the Polish policy of the Learning Agendas analyzed in Chapter 8 (Olejniczac, Kupiec & Park, 2027) focuses on identifying evaluation demands that can elicit useful and novel information. They well embody an evaluation policy of the second type.

Other chapters analyze what an evaluation policy of the third type looks like: Chapter 5 (Koleros, Liston & McGhee, 2027) refers to the movement of the Equitable Evaluation Initiative promoted by a group of philanthropies, that is premised on the collaboration between commissioners and evaluators, where both parties are actively involved in defining the evaluation's goals, methods and outcomes. Such a policy is based on a shared set of internal practices rather than a series of external actions performed by evaluators alone.

Democratic commissioning as a way of interpreting a management role

In all those situations, it is not only a matter of evaluation policy, which depends on the political authority, but also on the actions to implement it. These actions include the choices made by people charged with specific commissioning tasks, such as selecting the evaluand, writing the terms of reference (ToR), contracting, managing and utilizing. These people may be functionally appointed, as well as operators who perform these tasks alongside other duties (Catsambas & Davidson,

2024). It is precisely in the implementation of the commissioning evaluation policy when the commissioners gain (or not) agency.

Commissioners play a crucial role in interpreting the evaluation policy mandate (Lipsky, 1980). Regardless of the intentions of the evaluation policy, implementing that policy within an evaluation system is very different and depends greatly on the commissioners. Their interpretation and implementation of any evaluation policy – certainly of the second or third types listed above – is key to democratic commissioning. This is why, throughout this book, we have been interested in how commissioners who intend to contribute to democracy through evaluation might exploit their situation. As Catsambas (2027) suggests in Chapter 6, similarly oriented commissioners can develop appropriate strategies regardless of their location. This is an encouraging inspiration for our work.

To conceive of a democratic way to perform this managerial role, we can proceed analogously to how we approached evaluation in general. In the former, we compared variations in the demand and supply sides of evaluation (as proposed by MacDonald). Here, we can compare variations in the managerial role of commissioning (demand) with those in the professional role of evaluation (supply). Following Schwandt's analysis of the evaluator's professional role, we can distinguish between technocratic and democratic managerial styles of commissioning.

Inspired by Dzur (2008) on democratic professionalism, Schwandt (2018) inquired into "the ethical and civic dimension of what it means to be a professional evaluator", particularly a democratic one. "What is the public good that the aspiring profession of evaluation is to serve?"

Dzur was interested in the link between professionalism, the realm of sociology and democracy, the realm of political science, and how democracy might inform the characteristics that traditionally define a profession. These characteristics include (a) training in specific knowledge, (b) self-regulation, such as codes of conduct and certification and (c) social responsibility, or an orientation toward community interests instead of individual self-interest. This involves protecting critical social interests from being "colonized by market forces" (Dzur, 2008, pp. 44, 47).

Dzur's concept of "democratic professionalism" is based on "commitment to knowledge as well as co-direction of professional services" with laypeople, and public collaboration (2008, p. 130). It draws on Dewey's conception of the democratic professional as the one "who worked *with* rather than *for* the public, who *facilitated* public understanding and practical abilities rather than *led* the public" (Dzur, p. 5).³

Elaborating on Dzur's argument, Schwandt (2018) identifies three types of **professional evaluators**:

- **Social trusteeship**, "the belief that evaluators best serve their clients and the public interest by producing valid, empirical evaluative knowledge, independent of any political agenda or special appeals" (p. 233)
- **Technocratic professionalism**, "apparent in evidence-based approaches to decision making based in rigorously developed causal claims stemming from analyzing randomized experiments" (p. 234).

- **Democratic** professionalism, “participatory evaluation that are genuinely collaborative, share authority and responsibility for decisions, and are grounded in an ethical (versus pragmatic) justification (...) it results from deliberative democratic theory and a conception of democracy as a way of life (not just a set of political arrangements)” (p. 234).

Can this distinction have a parallel in the opposite role of commissioner? While professionals, and evaluators for that matter, work in the civil society sphere (that Dzur calls “middle democracy”), commissioners belong to the public sector, by definition regulated by law and supposed to work in the public interest.

With reference to Dzur’s professional characteristics, commissioners certainly share with evaluators the need for training in the specific activity that is evaluation; hence, our concern for training commissioners on the potentialities of evaluation (see chapters in the third section of this volume).

But what about commissioners’ relationship to the public interest? In contemporary welfare states, where policies are evaluated based on their ability to improve population well-being, reference to public interest is generally equated with policy success (efficiency plus effectiveness), reflected in how evaluations are commissioned. Commissioners’ activities are regulated by public service laws, and they are expected to be loyal to political elites. This reduces their freedom of action, but it doesn’t preclude their autonomous agency within the evaluation process, especially when they collaborate with external actors and broaden the evaluation’s perspective. While evaluation professionals are free to decide which values to honor, setting themselves as technocratic or democratic evaluators, commissioners can play a subordinate role, feeling bound by strict regulations and dependent on upper-level political decisions. Alternatively, they can play a more autonomous role, seizing opportunities in open situations to avoid over-bureaucratization, over-standardization and excessive rigidity. In doing so, they act in the name of stakeholder collaboration and public interest.

Thus, there may be various degrees of alignment between the evaluation policy and commissioning. Commissioners may adhere completely to the evaluation policy or exert some autonomy. For example, they may pursue democratic goals that are distant from the evaluation policy. They may also interpret their role in a legalistic way, identifying commissioning with contracting, even in the presence of a democratic policy. All of this is a matter of strategy, which we will discuss below.

Adherence to the evaluation policy of an organization will showcase corresponding kinds of **commissioning**:

- **Bureaucratic** commissioning: asking token reports for mere compliance (purely submitted to the internal power unbalance, as in the concept of “evaluation capture” (Raimondo & Leeuw, 2021))
- **Technocratic** commissioning: tendering for “robust research” to offer decision-makers as in the idea of evidence-based movement
- **Democratic** commissioning: involving stakeholders and their empowerment

Our chapters provide abundant exemplification on these different types of commissioning, through the analysis of how the various phases and tasks of commissioning are tackled in the cases, or topics, under scrutiny. In some cases, they are even presented in rubrics in Chapters 3 and 4 (Rogers & Montague, 2027; Ling & Millard, 2027).

Chapter 2's identification (Perrin & Stame, 2027) of the main aspects of commissioning is an account of the differences between bureaucratic and democratic commissioning. The latter is premised on the idea that commissioners' power should push them to take responsibility in getting a useful evaluation, throughout all its process. The guiding principles of a democratic evaluation would imply engaging other stakeholders and collaborating with them, utilizing a bespoke approach, able to allow for changes in the situation and iteratively readapting the evaluation questions, being open to the bigger picture, and "building in considerations regarding utilization from the beginning of the evaluation planning". Such principles are based on a lived experience of contrasting an all-too frequent practice of bureaucratic commissioning, characterized by repetitive requests – for the same programs, with the same ToR, the same pick of evaluators, the same "one size fits all" approach – and a mechanical use of guidelines and standards.

Rogers & Montague, 2027 delineates three requisites for *democratic* commissioning: transparency, engagement and heterogeneity. On top of the authors' rubric, these characteristics may also work for distinguishing democratic commissioning from other types. Transparency implies explicating evaluation plans and reporting achievements both in terms of methods for data collection and analysis, and processes for governance and quality assurance; but the authors have experienced cases of censure of non-favorable data, a trait of bureaucratic commissioning. Engagement would imply providing resources, authority, incentives and opportunities to support meaningful engagement with participants, not considered as simple sources of data, as would happen in bureaucratic commissioning. Heterogeneity would mean identifying and seeking to understand variations and explicitly paying attention to this in recommendations, instead of looking for an average effect, that would lead to the same recommendations for a varied population, a typical trait of technocratic commissioning.

In their turn, Ling & Millard, 2027 draw a rubric of a "Right Holder-Centric Approach" stating to what extent a situation meets the criteria for democratic commissioning: accessibility and inclusiveness ("comprehensive strategies are in place to ensure the inclusion of all relevant vulnerable or hard-to-reach groups"); engagement in data interpretation ("right holders are actively involved in interpreting the data, and their feedback is integral to the evaluation process"); access to results ("results are fully shared with right holders, ensuring transparency and access for all stakeholders"); operationalization of results ("right holders are actively engaged in discussions and decision-making on how results are operationalized"). For each of these criteria, fully meeting them could be considered tantamount as democratic commissioning, while not meeting them as bureaucratic commissioning.

In Chapter 5 (Koleros, Liston & McGhee, 2027), the "equitable evaluation platform" opposes *democratic* commissioning ("critical practices of collaboration

Table 15.1 Traits of different types of commissioning

	<i>Bureaucratic</i>	<i>Technocratic</i>	<i>democratic</i>
Rogers & Montague	• Censure on report • No resources for engagement of beneficiaries	Requiring RCT (homogeneity)	• Transparency • Engagement • heterogeneity
Ling & Millard	• Echo chamber effect • Right-holders not consulted • Right-holders not owning the results		• accessibility and inclusiveness • engagement in data interpretation • access to results • operationalization of results
Koleros, Liston & McGhee		Mere operationalization of equitable goals	• Equity • Cooperation • dialogue
Perrin & Stame	• Serial ToR • Mechanical use of guidelines • Standardized approach	Intruding in methodologies	• Taking responsibility • Cooperation • Responsibility for changes in situation (iterative approach) • bespoke approach • engage stakeholders in all aspects of the evaluation

for equitable evaluation”) to *technocratic* commissioning (demands for actions performed by external evaluators). This distinction overlaps that between double-loop learning (inquiring on the assumptions and beliefs of the foundation, addressing power asymmetries and root causes of complex societal issues) and single-loop learning (operationalizing equitable evaluation principles through a set of actions that make a more conventional evaluation “more equitable”). Collaboration with evaluators (and other stakeholders, as grantees and foundation staff) requires negotiating differing perspectives and expectations, but it also offers the chance to build a more inclusive and aligned understanding of what equitable evaluations entail, by empowering stakeholders to “co-create solutions that address systemic inequities effectively”.

Table 15.1 shows how the traits of commissioning that have been analyzed in the aforementioned chapters compare with the typology of commissioning.

An intermediate summing up. What do commissioners want to be?

We firmly believe in the “democratic purpose of evaluation” (Schwandt, 2023). However, we are aware that evaluations may be conducted in ways that undermine democracy, or worse, become tools for non-democratic governance within rigid evaluation systems. This underscores the importance of identifying the characteristics of a democratic evaluation. MacDonald pioneered this line of thinking when he distinguished democratic evaluation from bureaucratic and

Table 15.2 Typology of democracy and its adaptation to evaluation

<i>Types of democracy</i>	<i>Types of evaluation</i>	<i>Evaluation policy</i>	<i>Evaluation professionalism</i>	<i>Commissioning role</i>
(Hanberger)	(Fishkin)	(MacDonald)	(Rogers Montague)	(Schwandt)
	Competitive	Bureaucratic	Absent token	Bureaucratic
Elitist (To the people)	Elitist	Autocratic	For evidence-based policy	Social Trustee Technocratic
Participatory (For the people)	Participatory			
Democratic (With the people)	Deliberative	Democratic	Deliberative engagement	Democratic

autocratic evaluation. This typology continues to inspire evaluators and other practitioners.

We claim that for an evaluation to be democratic, all actors involved in the process, including commissioners, should act according to democratic principles. We found that MacDonald's typology, which has already been used for other evaluation-related aspects, can also be applied to commissioners. Table 15.2 shows the full range of applications of that typology and its variations.

As with all typologies, Table 15.2 represents alternative poles set as good/discrete/bad or white/gray/black. These are abstract concepts of net alternatives that may not seem realistic in everyday life with all its complexities and uncertainties.

The typology helps us understand that, like evaluation, commissioning is not a neutral activity and has consequences based on how it is performed. What the typology means is that commissioning, as evaluation, is not a neutral activity, and has its consequences in whatever way it is performed. The typology helps us understand what commissioners want to be (the same question Schwandt asked evaluators), what are the values they pursue with their activity, what they are acting for. In this way, the typology complements our discourse on the three headings. It is a matter of agency, as commissioners are confronted with the consequences of their decisions. It is also a matter of approaches, ranging from positivist to constructivist. Some approaches ignore the role of the commissioner, while others engage them. All these approaches imply a related kind of commissioning. Finally, it is a matter of evaluation systems, ranging from rigid to flexible.

However, democratic commissioners will continuously encounter situations in which it is unclear how to align with their aspirations. As Tom Ling suggested during a session at the 2025 INTEVAL meeting, they may need to "dance" between technocratic and democratic solutions, developing strategies to take advantage of favorable situations or limit risks.

This discourse leads us to address a different question: What can or should commissioners who want to pursue democratic values do? This is a matter of practical guidance, which we will address below.

Clearly, commissioners play a pivotal role in evaluation processes. This role is political and strategic, not technocratic. They are not merely managers looking for compliance with norms and terms of reference (ToRs): they must exert their agency and recognize their pivotal role between evaluators, stakeholders of the evaluands and policymakers who will use the evaluations.

Taking into account the content of the various chapters, as well as our discussions and debates at the INTEVAL meetings in Rome (2023), The Hague (2024) and Washington (2025), we can identify two main dimensions that characterize the pivotal role of strategic, political and democratic commissioners. One dimension relates to the individual and organizational styles through which commissioners exercise their authority. A flexible approach that moves away from rigid adherence to rules and norms can facilitate adaptation to each evaluation context. This style involves considering the context beyond the organization or institution to reach society as a whole, based on the idea that democratic commissioners should be accountable to the public interest, not just institutional compliance. This approach should incorporate ethics oriented toward the concepts of “ethos” and professionalism rather than focusing solely on professionalization (Schwandt, 2017). The second dimension relates to the inclusion and involvement of stakeholders, including evaluators, policymakers and other evaluand’s voices, in the commissioning processes. A co-responsibility and co-creation scheme between commissioners and evaluators, rather than a hierarchical one, is key to democratic commissioning. This means not leaving the identification and engagement of stakeholders as the sole responsibility of evaluators but rather considering it a shared responsibility with commissioners. Commissioners could play a more active role by exercising their authority and agency while respecting and recognizing the authority and agency of evaluators and other stakeholders. This would entail stronger involvement in evaluation processes and go far beyond ensuring compliance with terms of reference.

Part II Practical guidance for democratic commissioning: what can/should commissioners do if they want to act according to democratic values?

To organize the practical advice that can be extracted from the book, we used an adaptation of the typology that Bustelo, Crespo, & Ligerio (2027) used in Chapter 12 to identify evaluation competencies (see Table 15.3). This adaptation reflects a specific commissioner’s perspective. These are seven categories or areas: 1. Understanding the discipline; 2. Strategic & democratic evaluation vision; 3. Technical and methodological knowledge and skill; 4. Management capacity; 5. Interpersonal and organizational skills; 6. Ethics and transformative vision; and 7. Promoting evaluation use and influence. Although 2, 4, 6 and 7 are specifically key areas for commissioners, this does not mean they should also be attentive to

and develop 1 and 3 (more specifically important for evaluators) and 5, a key area for any role.

Considering the three headings set up in Chapter 1 (Stame & Bustelo, 2027) for analyzing the encounter between commissioning and democratic principles, we have developed these seven areas: the responsibility of commissioners and their exercise of agency; the influence of different evaluation approaches (methods, values and relationships) on the commissioning process; and the necessary consideration of the context in which evaluation occurs, that is, evaluation systems. For example, an understanding of the discipline and the theory and practice of evaluation is necessary for the “agency” we claim for commissioners who commit to and contribute to democratic commissioning processes. The more evaluation education, knowledge and experience commissioners have, the greater their capacity to act. Having a good knowledge of evaluation systems is necessary for commissioners to develop a strategic and democratic evaluation vision (2), management capacity (4), methodological skills (3), and ethics and a transformative vision (6) of values, as well as interpersonal and organizational skills (5) for relationships. Finally, to promote evaluation use and influence, commissioners need a combination of agency and knowledge of evaluation approaches and systems.

1 Understanding the discipline, the theory and practice of evaluation. *This includes appreciating the role of evaluation in society, and being familiar with evaluation theories, schools, concepts and approaches, as well as evaluation links with other social sciences.*

Strictly speaking, commissioners do not need to be evaluation experts, but they should have a basic understanding of the discipline. In this book’s third section, Chapter 12 (Bustelo, Crespo & Ligeró, 2027), Chapter 13 (Bastøe, Høydal & Fossetøl, 2027), and Chapter 14 (Olejniczak, Strzęboszewski & Stępień, 2027)) advocate for a certain type of evaluation training for commissioners, and the title of the section – “Educating the commissioner, through a dialogue between commissioners and evaluators”- reflects positive experiences with training and capacity building for commissioners. All of these experiences involve dialogue between commissioners and evaluators. Dialogue is always more enriching when both parties have a similar starting point. In discussions with INTEVAL members who have been evaluators and commissioners, everyone recognizes the usefulness of understanding the theory and practice of evaluation to be a good commissioner.

This volume’s co-authors propose basic “dos and don’ts” related to the theory of evaluation. For example, there is the concept of evaluative thinking, which is critical thinking applied to evaluation contexts (Bastøe, Høydal, & Fossetøl, 2027), and the importance of context. This means designing the evaluation around the specific context, information needs, and the context in which the evaluation is taking place (Perrin & Stame, 2027). For this reason, Bustelo, Crespo, & Ligeró (2027) caution commissioners against imposing any concrete methodological or theoretical perspective and recommend negotiating it with the context, the needs of the evaluand, the evaluators, the users and other stakeholders. Additionally, an

important feature of evaluation theory is its political nature. This implies acknowledging evaluation as a social and political practice, not just technical work (Bustelo, Crespo, & Ligeró), and recognizing that evaluation – and thus commissioning – is shaped by power and interests (Ling & Millard, 2027). Additionally, some co-authors argue that commissioners should use theory-driven approaches to better understand the evaluation process and link program design, implementation and observed outcomes (Chapter 9: Marra & Buscemi, 2027). In sum, if commissioners are equipped with an education in, and competencies related to, the discipline's recognized theoretical principles, they will be better prepared to fulfill their role as commissioners. This idea is supported by the three chapters in the third block (Bustelo, Crespo & Ligeró, 2027; Bastøe, Høydal & Fossestøl, 2027; Olejniczak, Strzęboszewski & Stępień, 2027).

2 Strategic & democratic evaluation vision: *ability to identify relevant evaluation questions, use of evaluative thinking, constructively engage with stakeholders, understand the value of diverse evaluation approaches, and adapt designs and methods to the context.*

This is a key area for commissioners. If anyone should have a strategic vision for evaluation systems, it should be them. Therefore, they must lead, enable and promote this strategic, democratic vision within organizations that commission evaluations. To this end, in Chapter 13, Bastøe, Høydal, & Fossestøl (2027) encourage critical evaluative thinking because commissioners must reflect on why an evaluation is initiated and how the results can promote democratic values, such as transparency and participation. According to Buckley et al. (2015), evaluative thinking is critical thinking applied to the context of evaluation. It is motivated by an attitude of curiosity and a belief in the value of evidence. It involves identifying assumptions and presuppositions, understanding the evaluand's program theory and theories of change, asking reasoned questions, setting relevant evaluation questions strategically, pursuing deep understanding through reflection from multiple perspectives, acknowledging democratic values and the diversity of evaluation approaches, and offering informed decisions in preparation for action. Evaluation systems where commissioning processes take place should allow evaluative thinking to reach its full potential.

The four chapters in the first block of this volume (Chapter 2: Perrin & Stame, 2027; Chapter 3: Rogers & Montague, 2027; Chapter 4: Ling & Millard, 2027; and Chapter 5: Koleros, Liston & McGhee, 2027) present this vision of strategic and democratic evaluation. Additionally, the six chapters reflecting on experiences where this strategic vision (or lack thereof) is key to understanding the challenges and opportunities of commissioning in different contexts (Chapter 6: Catsambas, 2027; Chapter 7: Lahey & Lucks, 2027; Chapter 8: Olejniczak, Kupiec & Park, 2027; Chapter 9: Marra & Buscemi, 2027; Chapter 10: Laaksonen & Palenberg, 2027; Chapter 11: Lindkvist & Lloyd, 2027) present this vision.

Catsambas (2027) sums up her advice with the phrase, "Map and strategically navigate context". According to her, commissioners must assess the political,

institutional and organizational context in which they operate in order to identify opportunities and constraints. This mapping helps align evaluations with democratic values, such as inclusion, equity and transparency, even in restrictive or politically sensitive settings.

Perrin & Stame (2027) argue that commissioners must seek meaningful stakeholder engagement throughout all phases of the evaluation process to bring a strategically democratic vision to evaluation processes. This engagement should start with conception and continue through utilization. Rogers & Montague (2027) encourage commissioners to consider using “community engagement-oriented” guiding principles to complement conventional evaluation principles by supporting dialogue across differences, deliberation processes and community members and stakeholder groups taking in new information, understanding different viewpoints and making informed decisions. Ling & Millard (2027) propose framing evaluations to empower right holders and vulnerable groups. They suggest asking who benefits from each evaluation and who is included or excluded. They also advise avoiding the temptation to reduce democratic debate to elite-driven processes and adapting evaluation purposes to support democratic deliberation. Similarly, Bustelo, Crespo & Ligerio (2027) claim that commissioners should exercise their agency and power strategically, but not at the expense of evaluators’ and other stakeholders’ agency. Koleros, Liston, & McGhee (2027) argue that commissioners should enhance an equitable evaluation framework by incorporating self-reflexive practices that examine their organization’s assumptions, power dynamics and internal politics rather than commissioning external “equitable” actions. In essence, they imply that the traditional contractor-client power dynamic is incompatible with equitable evaluation principles.

Even when outsourcing operational functions, this strategic perspective and oversight should be retained, as full delegation can risk the direct relationship between commissioners and evaluators. Commissioners must be decision-makers, not merely contract managers. Democratic commissioning depends on dialogue, not delegation alone (Laaksonen & Palenberg, 2027). Commissioners should act as knowledge brokers, linking evidence to organizational needs. To do so, they must balance accountability with learning objectives and engage stakeholders and external experts early in the process (Lindkvist & Lloyd, 2027). It is also important to commission evaluations that clearly demonstrate how the findings will inform programming and accountability (Marra & Buscemi, 2027). The commissioner role should therefore evolve from technical to strategic (Olejniczak, Kupiec & Park, 2027).

Bastøe, Høydal & Fossetøl (2027) warn us about the risks of standardizing evaluation practices in public administration to the point that they lose their potential to provide real learning and their ability to challenge power. Lindkvist & Lloyd (2027) discuss the danger of isolating evaluations from broader debates or organizational reform needs. Laaksonen & Palenberg (2027) discuss the danger of outsourcing to the extent that it might erode the commissioner’s sense of ownership over the evaluation’s purpose and process. Conversely, Bustelo, Crespo & Ligerio (2027) argue that commissioners should not act as if they are the sole owners of the

commissioning process, implying that evaluation is a process that belongs exclusively to the organization that commissions it, and evaluators are only external consultants who do a job that does not belong to them.

3 Technical and methodological knowledge and skills: *understanding the strengths and challenges, as well as the formative and summative potential of various methods, techniques and instruments, including their validity and reliability.*

Almost all co-authors acknowledge the implications of method selection. From there, in Bustelo, Crespo & Ligerio (2027) assert that commissioners should have some methodological expertise to understand that methodological choices are not neutral and reflect broader ethical, theoretical, conceptual and visionary orientations. In Rogers & Montague (2027) emphasize the importance of paying attention to heterogeneity in democratic commissioning. They argue that it is necessary to investigate and report on the differential experiences of key groups, resources, effects and implications, rather than simply reporting average effects or treating differential effects as “nuances”. This is often overlooked when using randomized controlled trials (RCTs) and experimental approaches exclusively. Ling is more specific, talking about the usefulness of participatory and discursive approaches to surface diverse perspectives. Perrin & Stame (2027) claim the flexibility of the outcome approach to evaluation commissioning.

In the context of Sustainable Development Goals (SDG) implementation in Lahey & Lucks (2027) warn us about the common mistake of relying solely on traditional linear evaluation approaches in contexts of high uncertainty or multiple stakeholders. They also caution against assuming that one-size-fits-all evaluation methods or timelines are appropriate, and they emphasize the importance of not letting methodological rigidity prevent the timely delivery of findings relevant to decision-making (“don’t let the perfect be the enemy of the good”). Lindkvist & Lloyd (2027) believe that commissioners should allow for innovation by not prescribing methodologies too rigidly.

4 Management capacity: *ability to respond to the needs and concerns of various stakeholders, anticipate opportunities and challenges throughout the process, and prudently manage resources to achieve evaluation objectives.*

Along with category 2, this represents the most specific and key role for commissioners. For this category, we were inspired by the chapter on “Commissioning Great Evaluations” in Catsambas & Davidson (2024) for scoping evaluation criteria and questions, determining the best way to contract an external evaluation, and writing Terms of Reference (ToRs). We also include resource management, including timing and calendars, in this category.

Regarding their management capacity in general, Marra & Buscemi (2027) warn commissioners to allocate sufficient resources so that evaluation funding reflects the scale of public investment. Bustelo, Crespo & Ligerio (2027) identify

an effective commissioner as one who is fair, just, and inclusive. This means that playing this role involves taking proactive actions to develop procedures, contracts, management plans and monitoring systems that respect the different roles assigned to the various actors in the evaluation process.

Regarding the scope of evaluation criteria and questions, all co-authors emphasize the pivotal role of commissioners. Rogers & Montague (2027) emphasize securing evaluation questions that promote collective learning and stakeholder engagement. In Chapter 9, Marra & Buscemi (2027) emphasize providing clear, context-sensitive evaluation questions and maintaining open dialogue with contractors throughout the process.

To frame adequate evaluation questions, the co-authors offer practical advice related to the flexibility and adaptability of the management process, as well as allowing spaces for coordination, teamwork, participation and co-creation. Perrin & Stame (2027) warn commissioners against following rigid templates, or “ticking bureaucratic boxes”, while Ling & Millard (2027) proposes anticipating contextual changes and allowing for iterative adjustments. This includes not scheduling evaluations only around political cycles. Laaksonen & Palenberg (2027) also caution against confusing standardization with rigidity. Rather than replacing context-sensitive evaluation design and relationships, uniform procedures should enhance them. Bustelo, Crespo & Ligeró (2027) advocate for the creation of governance spaces, such as steering groups that promote participation and commitment. Koleros, Liston & McGhee (2027) advocate for the creation of shared decision-making processes in which commissioners, evaluators and grantees have an equal voice in shaping evaluation goals and methods. These steering committees should also include technical, thematic and evaluation experts to ensure quality and learning (Marra & Buscemi, 2027). In the case of outsourcing some commission services, Laaksonen & Palenberg (2027) advise promoting structured formats and toolkits that support process clarity and mutual expectations. They also advise using intermediary roles (e.g., service coordinators) to support, not substitute, commissioner engagement and responsibility. Lindkvist & Lloyd (2027) advise using interim deliverables to maintain alignment between commissioners and evaluators.

Some authors warn commissioners about the importance of balancing internal actions and outsourcing and the danger of using outsourcing to avoid internal responsibilities. Koleros, Liston & McGhee (2027) argue that equitable evaluation should not be merely a checklist of external actions. Rather, it should be fundamental internal practices that guide every step in the commissioning process. Marra & Buscemi (2027) advise against treating in-house and outsourced evaluations as interchangeable without first assessing internal capacity.

To do a good job scoping evaluation criteria and questions, it is necessary to provide sufficient time, resources and opportunities to meaningfully involve all key stakeholders in the evaluation process from the start. This involves not underestimating the budget and timeline needed for truly collaborative, equity-centered approaches, managing complex, multi-sectoral or horizontal evaluations and supporting flexible evaluation processes, especially under time or resource

constraints. This ensures relevance and the need for structured dialogue and coordination mechanisms in complex evaluations.

Determining the best way to contract an external evaluation should allow for the establishment of contracts that guarantee evaluators' independence, and the decision-making autonomy of the various actors involved, as well as the specification of responsibilities and tasks to ensure respect of different roles (evaluators, commissioners, users and other stakeholders) (Bustelo, Crespo & Ligerio, 2027). In this regard, the co-authors caution us to ensure, through distinct procedures, the autonomy of the entities responsible for payment from those accountable for managing and monitoring evaluations. This will prevent payment delays from being used as a means of controlling evaluators. Laaksonen & Palenberg (2027) also emphasize the importance of introducing transparent qualification criteria and participatory team selection processes that include space for commissioner input and stakeholder balance. In any case, these authors argue that evaluation team selection cannot be treated as a purely administrative process because it is a relational act with strategic implications.

In their analysis of the Italian Cohesion Policies' evaluation market, Marra & Buscemi (2027) provide concrete examples and advice for managing a truly democratic and strategic procurement process that yields better results. The authors advise diversifying contracting to include large and small firms to promote competition, methodological pluralism and policy relevance. This includes facilitating partnerships between large and small firms to balance capacity and local knowledge. This diversity should prompt commissioners to divide contracts into specialized tasks, thereby increasing transparency and enabling more nuanced evaluations. However, as they point out, the market currently operates based solely on the reputation of evaluation firms, ignoring their actual value based on substantive contributions. The authors warn commissioners against relying on a single contractor for an entire programming period, unless the contractor demonstrably supports participation, continuity and learning. They also caution against excessive auction rebates, which they believe should be limited, as these rebates can undermine evaluation quality and perpetuate market concentration. They emphasize the importance of not prioritizing price over content quality, particularly in complex, multi-year evaluations. They also advise against over-bundling contracts with multiple deliverables that dilute focus and compromise methodological rigor.

Regarding the writing of Terms of Reference (ToRs), Laaksonen & Palenberg (2027) propose that commissioners co-design them with evaluation teams during a flexible pre-contracting phase to foster alignment, ownership and quality. However, Lindkvist & Lloyd (2027) advise commissioners to define clear, tightly scoped ToRs. Ling & Millard (2027) argue that ToRs should allow for causal complexity ("causal density") and should not be overloaded with too many disconnected or unfocused questions (Lindkvist & Lloyd, 2027).

5 Interpersonal and organizational skills: *flexibility and capacity to adapt, ability to clearly communicate objectives and results, promote teamwork, assertive negotiation and positive conflict resolution. Especially important for*

the relationship between commissioner and evaluator. This also includes cultural and gender sensitivity.

Almost all of the authors, as well as our interviewed INTEVAL colleagues, emphasize the importance of interpersonal skills and a practical approach on the part of the commissioners. This approach should not be solely driven by norms or by what should be done, but rather by flexibility in order to understand and adapt to the context (Bastøe, Høydal & Fossetøl, 2027; Perrin & Stame, 2027). Catsambas (2027) believes that commissioners should exercise influence beyond their formal roles. This means that even when evaluation is not their primary responsibility, they can go beyond procedural tasks to shape evaluations that elevate marginalized voices, encourage open dialogue and promote civic engagement. In this sense, Olejniczak, Kupiec & Park (2027) conclude that, in addition to being contracting authorities, commissioners should be facilitators. However, Laaksonen & Palenberg (2027) highlight that we should not underestimate the influence of staffing limitations and information management systems on the practical possibilities of commissioning.

Pollard (Appendix: Pollard, 2027), in her analysis of the model of the Neo-Weberian State, argues that organizational rules and requirements about commissioning evaluations should be interpreted to provide maximum opportunities to give voice to citizens, particularly direct beneficiaries, and enable deliberative engagement. Nearly all of the authors advocate for commissioners to actively promote stakeholder participation by engaging civic stakeholders early and throughout the process and by not reinforcing echo chambers where only dominant voices are heard (Ling & Millard, 2027).

Lindkvist & Lloyd (2027) make a special mention of the relationship between commissioners and evaluators, arguing that commissioners should foster trust and long-term collaboration with evaluators. According to Bustelo, Crespo & Ligeró (2027), a good commissioner executes their agency and power in a negotiated, horizontal, collaborative and collegial manner. This approach recognizes evaluators' roles as not only analysts, but also facilitators and communicators in complex systems (Lahey & Lucks, 2027). Consequently, it facilitates evaluators' access to diverse participants, ensuring that service users are not excluded or treated as passive data providers (Ling & Millard, 2027). Additionally, it prevents gatekeeping of information and limits evaluators' access to internal organizational dynamics only when relevant to understanding the work (Kolerós, Liston & McGhee, 2027). Lindkvist & Lloyd (2027) emphasize safeguarding evaluator independence while monitoring quality. For Ling & Millard (2027), it is important to support adaptation, dialogue and review moments.

- 6 **Ethics and transformative vision:** *ability to uphold ethical standards and democratic values in evaluations, maintain independence in data interpretation and judgment, use a collaborative approach with evaluators and other stakeholders respect and recognition of evaluation teams, and contribution to the professional evaluation community.*

Catsambas (2027) asks the commissioners to balance risk with ethical responsibility. This means they should be aware that promoting democracy through evaluation may entail personal and professional risks. They must make calculated decisions balancing bold action with ethical considerations, including protecting participants and themselves.

Rogers & Montague (2027) argue that commissioners should expect transparency in evaluation processes and products, ensuring that reporting is full and frank except where privacy commitments require otherwise. They imply that simply reporting evaluation findings does not achieve transparency without information about processes or opportunities to engage. Koleros, Liston & McGhee (2027) also recommend establishing clear communication protocols to foster transparency and democratic participation throughout the evaluation process. Ling & Millard (2027) recommend sharing results transparently and accessibly.

Lahey & Lucks (2027) caution commissioners to use evaluations as facilitative tools to navigate competing interests and potential trade-offs, producing credible insights into future opportunities for improved results for all, including the most vulnerable. Pollard's appendix demonstrates this commitment to inclusion when she instructs commissioners to consider how commissioning activities can reflect national commitments to human rights and equality, enabling minority groups' rights, voice and interests to be heard. Ling & Millard urge us not to overlook how evaluations privilege some voices while excluding others. They also emphasize the importance of prioritizing the inclusion of vulnerable and hard-to-reach groups and respecting right holders as active interpreters of findings.

Upholding ethical standards and democratic values in evaluations means safeguarding the commissioning process from bad practices caused by vested interests. This includes the interests of the commissioners themselves, as well as those created by political or organizational defensiveness. Lindkvist & Lloyd (2027) clearly warn commissioners against diluting findings due to these interests. Ling & Millard caution against designing evaluations solely to satisfy policymakers. Bustelo, Crespo & Ligeró (2027) urge commissioners to pay special attention to interference in the drafting of final evaluation reports. These authors offer practical advice on how to avoid bad commissioning practices identified by evaluation practitioners who are alumni of the Complutense University of Madrid master's program and have experienced these practices firsthand over the last 25 years. They emphasize the importance of commissioners not interfering with evaluators' work outside established channels, particularly in methodological decisions, such as which aspects of the program should be evaluated, the weight given to certain interpretations, and the modification or deletion of judgments in final reports. Following Ernest House, who warned us in *Evaluating with Validity* (1980) about discrepancies in evaluation content, they propose that commissioners who disagree with interpretations, judgments or recommendations given by evaluators might add a note explaining their reasons for disagreement instead of trying to modify evaluation reports. More specifically, they argue that commissioners should recognize evaluators' roles and report authorship, and should not use practical issues, such as payments, management or administration, to modify the evaluation process.

- 7 Promoting evaluation use and influence:** *referring specifically to the commissioners' role related to understanding the general and the organizational context in which the evaluation takes place, the identification of policy users' information needs, and continuous engagement and communication with stakeholders. This also includes promoting Evaluation Capacity Building for the promotion of evaluation use and influence.*

Interestingly, after working with competencies from the perspective of the commissioners, Bustelo, Crespo & Ligeró (2027) propose adding this category to the list. It is embedded in the other categories and refers specifically to the commissioners' role in promoting the use and influence of evaluations in their organizations.

Commissioners should first consider the issue of understanding the general and organizational context in which the evaluation takes place and enhancing a continuous organizational learning process in that context. To do so, Lahey & Lucks (2027) argue that commissioners should first ensure that knowledge generated through evaluations remains accessible and relevant over time. They promote evaluations as not only real-time learning tools, but also long-term knowledge resources. More specifically, Laaksonen & Palenberg (2027) recommend investing in system-level infrastructure, such as knowledge repositories, to ensure continuity, reduce duplication and promote learning. They also recommend using framework contracts to create continuity across evaluations and enable learning loops and shared institutional memory. However, they caution against assuming that improved evaluation management systems automatically lead to improved knowledge use. They argue that without systemic change in how institutions handle evaluation results, democratic gains will remain limited. Additionally, Lindkvist & Lloyd (2027) recommend building continuity and institutional memory across evaluations and linking new evaluations to previous ones to reinforce learning.

Another important issue related to enhancing use is, once again, stakeholder involvement. According to Marra & Buscemi (2027), stakeholder involvement can improve the formative use of findings, especially in contexts with a weak evaluation culture. Lahey & Lucks (2027) argue that the most effective way to promote the use of evaluations is to establish open communication and productive relationships between commissioners, evaluators and other partners. It is crucial not to exclude important sectors or institutions that will be essential for implementing evaluation recommendations. Koleros, Liston & McGhee (2027) emphasize the importance of allocating sufficient time and resources for authentic relationship-building among all evaluation stakeholders from the outset. Finally, Ling & Millard (2027) caution against limiting the use of results to the internal needs of commissioners, especially by empowering right holders to act on findings and holding duty-bearers accountable.

Communication of results is key to enhancing use, and Lindkvist & Lloyd (2027) recommend planning it strategically instead of treating it as an afterthought. For that, it is necessary to invest in dissemination and communication strategies. Commissioners should not rely solely on long technical reports. They

Table 15.3 Summary of dos and don'ts

Category	Dos	Don'ts
1. Understanding the discipline	- Acknowledge evaluation as a social and political practice, not just technical work. - Be familiar with evaluation theories (evaluative thinking, attention to context, evaluation political nature), schools, and approaches. - Acknowledge a diversity of methodological and theoretical perspectives	- Don't attempt to impose a standardized framework or approach - Don't treat evaluation commissioning as a neutral, purely administrative exercise. - Don't ignore power dynamics shaping evaluation agendas.
2. Strategic & democratic evaluation vision	- Acknowledge commissioners' agency as central to evaluation quality and democracy. - Frame evaluations to address relevant social needs and democratic purposes: inclusion, equity, and transparency. - Map and navigate context, always negotiating evaluand's needs, evaluators, users and other stakeholders. - Constructively engage stakeholders, ensuring inclusion of diverse perspectives and voices. - Balance accountability with learning and improvement.	- Don't isolate evaluations from broader societal debates. - Don't commission evaluations only for bureaucratic compliance. - Don't reduce democratic debate to elite-driven processes. - Don't act as the only owners of the commissioning process
3. Technical & methodological knowledge and skills	- Understand that methodological choices are not neutral - Be aware of evaluation methods and tools, and when they are more effective. - Ensure methodological appropriateness and rigor but also leave space for innovation. - Promote the attention, beyond average effects, to the differential experiences of key groups.	- Don't pretend that methodological choices are technically objective - Don't dictate methods without considering context and validity. - Don't rely in "one size fits all" evaluation approaches in contexts of multiple stakeholders - Do not allow methodological rigidity impeding timely decision-relevant findings
4. Management capacity	- Play a fair, just and inclusive management role, respecting the different roles assigned to the various actors in the evaluation process. - Introduce transparent qualification criteria and participatory team selection processes.	- Don't plan evaluations with insufficient timelines and resourcing for engagement and participation. - Don't only follow rigid templates

	- Co-design ToRs with evaluation teams to foster alignment, ownership, and quality. - Formulate evaluation questions for collective learning and stakeholder engagement (provide sufficient time, resources, and opportunities for it) - Support and create governance spaces, like steering groups, that promote participation and commitment.	- Don't overload ToRs with too many disconnected, unfocused or unrealistic evaluation questions. - Don't treat evaluation team selection as a purely administrative process. - Don't over-bundle contracts with multiple deliverables that dilute focus and compromise methodological rigor.
5. Interpersonal & organizational skills	- Engage civic stakeholders and enable deliberative engagement with citizens - Exercise influence beyond formal roles - Build trust and collaborative relationships with evaluators and stakeholders. - Communicate clearly and promote teamwork. - Safeguard evaluator independence while monitoring quality - Exercise negotiation and conflict resolution with sensitivity	- Don't isolate the evaluation process from stakeholders. - Don't exclude service users or treat them as passive data providers. - Don't let organizational silos undermine collaboration. - Don't limit evaluator's access to internal organizational dynamics when they're relevant to understanding the work
6. Ethics & transformative vision	- Balance risk with ethical responsibility, committing to equity, cultural sensitivity, and democratic values. - Foster transparency and democratic participation with communication protocols throughout the evaluation process. - Prioritize inclusion of vulnerable groups and respect right holders as active interpreters of findings. - Respect evaluators' expertise and other stakeholders' contributions.	- Don't allow political pressure or organizational defensiveness to distort findings. - Don't perpetuate exclusion or inequity through evaluation practices. - Don't design evaluations only to satisfy policymakers - Don't not interfere with evaluators' work outside of the established channels, especially regarding methodological decisions.
7. Promoting evaluation use & influence	- Invest in and ensure effective dissemination and communication strategies tailored to audiences. - Invest adequate time and resources for relationship-building among all evaluation stakeholders from the start - Create conditions for stakeholders to use results. - Link evaluations across time to build organizational learning. - Build evaluation capacity and engage yourself in continuous professional development	- Don't relegate dissemination to an afterthought. - Don't rely solely on long technical reports. - Don't allow evaluation findings to remain unused or hidden. - Don't leave out sectors critical to implement evaluation recommendations.

should prepare audiences in advance, tailor communication and use multiple tools, including events, briefs, podcasts and visuals, to reach wider publics (Ling & Millard, 2027).

Finally, building evaluation capacity is necessary to promote the use, influence and sustainability of democratic practices. According to Catsambas (2027), commissioners can foster long-term democratic impact by strengthening evaluation systems and building capacity within institutions and communities. This includes investing in the skills, resources and organizational culture that support participatory and rights-based approaches. Along with other authors, Lindkvist & Lloyd (2027) caution against neglecting evaluation capacity within organizations because sustainable change requires institutional support. Lahey & Lucks (2027) argue that commissioners should also support the development of evaluation awareness and capacity among decision-makers, including those beyond the direct process of designing and managing evaluations. The third section of this volume extensively covers the necessary evaluation capacity building for commissioners, emphasizing the promotion of evaluative thinking (Bastøe, Høydal & Fossetøl, 2027), identifying specific competencies of commissioners (Bustelo, Crespo & Ligeró, 2027), and recommending long-term, comprehensive capacity building, as well as systemic networks and communities of practice (Olejniczak, Strzęboszewski, & Stępień, 2027).

Notes

- 1 The current meaning of the term “autocratic” is different from when MacDonald conceived his typology. His concept of autocracy referred to the preeminence of academics over policymakers and the public. Today we refer as autocracy as to those systems of government based on personal power over their people and control against democratic and liberal countries (see Anne Applebaum *Autocracy, Inc.*, 2024, for a thorough account of contemporary incarnations of it). We keep the term “autocratic” here, while we change it to “technocratic” (following Schwandt) when we draw analogous typologies.
- 2 Rogers and Montague do not label this type of policy, but in the context of their argument we may well use this name.
- 3 Dzur (2008, p. 41) draws on Dewey, according to whom democracy is defined by “social equality, communication, collaboration, reflection”.

References

- Andersen, Niels Åkerstrøm. 2021. “The Power and Paradoxes of Evaluation Systems: Increasing Use but Impeding Change.” *Scandinavian Journal of Public Administration* 25 (3–4): 39–59. <https://doi.org/10.58235/sjpa.v25i3-4.7063>
- Bastøe, Per Øyvind, Øyunn Syrstad Høydal and Knut Fossetøl. 2027. “Training for Critical Evaluative Thinking.” In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Benerdal, Meike, and Malin Larsson. 2021. “Democratic Values in Evaluation Systems: A Circle That Can Be Squared?” *Scandinavian Journal of Public Administration* 25 (3–4): 63–83. <https://doi.org/10.58235/sjpa.v25i3-4.7069>

- Brousselle, Astrid. 2022. "Planting Seeds for Change." *What Should Evaluation Learn from COP 26? Views of Evaluation Practitioners*, by Van der Berg et al., *Evaluation* 28 (1): 9–10. <https://doi.org/10.1177/13563890221074173>
- Buckley, Jack, Thomas Archibald, Megan Hargraves, and William M. Trochim. 2015. "Defining and Teaching Evaluative Thinking: Insights from Research on Critical Thinking." *American Journal of Evaluation* 36 (3): 375–388. <https://doi.org/10.1177/1098214015581706>
- Bustelo, Maria, Ramón Crespo and Juan Andrés Ligeró. 2027. "The Good Commissioner: Competencies for Enhancing Democracy in Commissioning Processes." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Bustelo, María, and Steve Jacob. 2025. "From Studies to Systems: Ray Rist's Influence on Evaluation Systems—Insights from the International Research Group on Policy and Program Evaluation (INTEVAL)." *Journal of MultiDisciplinary Evaluation* 21 (50): 123–129.
- Bustelo, Maria, and Nicoletta Stame. 2027. "Democratic Evaluation Commissioning: Lessons from Theory and Practice." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Catsambas, Tessie Tzavaras. 2027. "The Influence of Context for Different Types of Commissioners of Evaluation: Challenges and Opportunities." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Catsambas, Tessie, and Jane Davidson. 2024. *Evaluation Management*. Thousand Oaks, CA: Sage.
- Cox, Julian, and Peter Barbrook-Johnson. 2021. "How Does the Commissioning Process Hinder the Uptake of Complexity-Appropriate Evaluation?" *Evaluation* 27 (1): 32–56. <https://doi.org/10.1177/1356389020976157>
- Dzur, Albert W. 2008. *Democratic Professionalism*. University Park, PA: Pennsylvania State University Press.
- Fishkin, James S. 2009. *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford: Oxford University Press.
- Hanberger, Anders. 2006. "Evaluation of and for Democracy." *Evaluation* 12 (1): 17–37. <https://doi.org/10.1177/1356389006064194>
- House, Ernest R. 1980. *Evaluating with Validity*. Beverly Hills, CA: Sage.
- Koleros, Drew, Monique Liston and Ray McGhee. 2027. "How Evaluation Commissioners and Evaluators Can Collaboratively Move Equitable Evaluation from a Set of External Actions to Internal Practices." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Laaksonen, Sari, and Markus Palenberg. 2027. "Evaluation Management Services at the MFA of Finland." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Laaksonen, Sari, and Markus Palenberg. 2027. "Evaluation Management Services at the MFA of Finland." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Lahey, Robert, and Dorothy Lucks. 2027. "Democracy, SDGs, Evaluation and Commissioners—Connecting the Dots." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Leeuw, Frans, and Jan-Eric Furubo. 2008. "Evaluation Systems: What Are They and Why Study Them?" *Evaluation*, 14 (2): 157–169.

- Lindkvist, Ida, and Rob Lloyd. 2027. "The Art of Commissioning. Lessons from Commissioner-Evaluator Collaboration." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Ling, Tom, and Ananda Millard. 2027. "Commissioning with Clarity: Empowering Democratic Spaces by Drawing on the Principles of Deliberative Democracy." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Perrin, Burt, and Nicoletta Stame. 2027. "How to Make Commissioning Matter in Times of Complex Fast-Changing Reality." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Lipsky, Michael. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.
- MacDonald, Barry. 1977. "A Political Classification of Evaluation Studies." In *Beyond the Numbers Game: A Reader in Educational Evaluation*, edited by David Hamilton, Denis Jenkins, Charles King, Barry MacDonald, and Malcolm Parlett. London: Macmillan Education, 105–108.
- Marra, Mita, and Virgilio Buscemi. 2027. "Democratic Implications of Commissioning Market-Based Evaluations: Insights from Italy's Cohesion Policy." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo, New York: Routledge.
- Olejniczak, Karol, Tomasz Kupiec and Yaerin Park. 2027. "Learning Agendas – From Studies to Streams of Evidence." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Olejniczak, Karol, Piotr Strzęboszewski and Marlena Stępień, 2027. "Academy of Evaluation – A capacity-building Initiative for Commissioning Authorities." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Pollard, Alison. 2027. "Applying Insights from the Model of the Neo-Weberian State to Commission More Democratic and Valuable Evaluations." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Raimondo, Estelle, and Frans Leeuw. 2021. "Evaluation Systems and Bureaucratic Capture: Locked." In *Changing Bureaucracies*, edited by Burt Perrin and Tony Tyrrell, 143–160. New York: Routledge.
- Rogers, Patricia, and Montague. 2027. "How Evaluation Commissioning Processes Can Support or Undermine Democratic Processes." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.
- Schwandt, Thomas A. 2017. "Professionalization, Ethics, and Fidelity to an Evaluation Ethos." *American Journal of Evaluation* 38 (4): 546–553.
- Schwandt, Thomas A. 2018. "The Concerted Effort to Professionalize Evaluation Practice: Whither Are We Bound?" In *The Evaluation Enterprise: A Critical View*, edited by Jan-Eric Furubo and Nicoletta Stame, 173–189. New York: Routledge.
- Schwandt, Thomas A. 2023. "Competencies for Evaluation as a Civic Science." *Journal of MultiDisciplinary Evaluation* 19 (46): 23–30.
- Stame, Nicoletta, and Maria Bustelo. 2027. "Commissioning Matters: Considerations and Challenges for Evaluation Commissioners Who Intend to Promote Democracy." In *The Power of Commissioning: Enabling Democratic Evaluation*, edited by Nicoletta Stame and Maria Bustelo. New York: Routledge.

Appendix

Applying insights from the model of the Neo-Weberian State to commission more democratic and valuable evaluations¹

Alison Pollard

This appendix reflects on the merit of applying insights, approaches and tools that are aligned with the Neo-Weberian State (NWS) model to commission evaluations that are more democratic and useful for public administrations. Analysis of European public administrative reforms since 2000 shows that many have been informed by the NWS model. This model has hierarchy as a principal form of co-ordination and incentivizes civil servants to meet citizens' needs, apply professional leadership styles and achieve results. When it is interpreted in relation to commissioning evaluations, it highlights the importance of applying participatory approaches to engage citizens – particularly direct beneficiaries – and exploring the extent to which results frameworks are valuable and sufficiently consider design, context, contribution and beneficiary needs. It recognizes the importance of professional leadership styles to support creativity and agility within commissioning activities. The model also shows the value of applying reflexive rationality – the logic of consequences and appropriateness – as well as results. When interpreted in relation to commissioning evaluations, this highlights the importance of reflecting on the extent to which policies support the achievement of national and global commitments to respect for human rights and equality, which are essential elements of democracy.

Bureaucracy, public administrations and Weber

Public administrations are the organizations with delegated responsibility for assessing the design and value of different policies and areas of government expenditure. The evaluations they commission support transparency, accountability and have the potential to inform organizational learning and decision-making at the regional, national and international level. The organizational structure of public administrations is inherently bureaucratic: they follow a rules-based structure that is operated by officials, with clearly defined hierarchical reporting and accountability structures. Logic, rationality and rules inform decision-making. Given the important role that public administrations play in commissioning evaluations, it is valuable to reflect on if and how conceptual models that analyze the structure and dynamics of bureaucracies can provide insights to help us understand the

evaluation commissioning process. Although Max Weber wrote about bureaucracies in the early twentieth century, his work remains highly relevant to contemporary analysis of public administrations.

Weber is widely recognized as creating a valuable ideal type – a methodological tool that can be used to interpret reality – that supports understanding of the characteristics and dynamics of bureaucracies. Weber (1997) argues that bureaucracy is critical to the functioning of public administrations in modern societies. The Weberian ideal type of bureaucracy identifies rationality and logic as driving forces for decision-making, with clearly defined organizational rules and chains of command informing working practices. Tasks are divided among staff based on their skills and expertise. This is seen to maximize effectiveness and efficiency as people work in consistent and coherent ways with roles that are specialized, structured and complementary. However, Weber (1997) notes that this can limit collaboration and creativity, create siloed working and lead to excessive adherence to bureaucratic rules. This way of working is problematic in environments where evaluations are commissioned because it incentivizes a high level of standardization and evaluation capture within the commissioning process. This limits opportunities for evaluations to be agile, responsive to emerging needs for evaluative evidence by public administrations. It also limits opportunities for evaluations to reflect the voice of different groups of citizens.

The Neo-Weberian State: a model to analyze bureaucracies

In contrast to Weber's ideal type of bureaucracy, Pollitt and Bouckaert (2011) argue that the NWS model places a greater emphasis upon public administrations being incentivized by:

- Understanding and meeting citizens' needs to support democratic accountability and responsiveness;
- Complementary representative democracy activities that increase voice and engagement;
- An orientation toward results and accountability of results; and
- Increasing professionalization of management and leadership.

These developments, along with modernization, development and digitalization provide public administrations with incentives and tools that mitigate against standardized, siloed and excessively bureaucratic ways of working. Instead, they support greater creativity and collaboration. Pollitt and Bouckaert (2011) assert that the emergence of NWS-oriented public administrations in many countries has enabled states to work to develop relevant, efficient and effective responses to complex global challenges including the climate crisis, global security and artificial intelligence.

In his recent working paper, Bouckaert (2022) notes that there are two variations of the NWS ideal type. The first highlights “top down” professional and results-orientated management that is led by civil servants when they are not constrained by rigid bureaucracy and hierarchy. The second focuses on the value and role of citizen engagement and “bottom up” influence from citizens and civil society. In reality, Bouckaert argues that there are hybrids and combinations of the two variations within public administrations and an additional influence of the role of markets, which engage and collaborate with the public sector.

New Public Management (NPM) and New Public Governance (NPG) are the two other main conceptual models that have been applied and discussed when analyzing recent public sector reforms. NPM emphasizes the value of applying the principles and practices of the free market and competition within the public sector to increase efficiency, with citizens seen as customers. The creation of quasi-markets is seen as valuable because they support agility and customer focus while also increasing value for money. Specific examples of NPM approaches within public administrations include the increasing use of private contractors and competitive tendering to deliver outsourced services through professionalized procurement activities. The NPG model emphasizes the importance of governance networks, which are multiple inter-dependent actors, stakeholder groups and partnerships to delivering effective public services. This includes organizations and networks across the private, voluntary and public sector. The use of these networks and relationships is seen to lead to enhanced public–private and public–non-profit collaboration that can address complex problems and subsequently deliver public value (Osborne, 2006 and 2010).

Andrews, Bezes and Van de Walle (2016) argue that between 2008 and 2013 most Western European public reforms were primarily NWS or NPG-based approaches, rather than NPM. They cite analysis of reforms within public governance that focus on improved collaboration, transparency and e-governance as rooted in the development of the NWS, based on a model of NPG or synergies between the two. Similarly, Torfing (2020) draws on analysis of public sector reforms in OECD member states to show that different governments have identified different aspects of the NWS, NPM and NPG models that are most operationally, politically and contextually appropriate and combined them within reforms to public administrations.

There are clearly challenges with isolating one specific model when analyzing the structure and functioning of public sector bureaucracies. However, I agree with Krogh and Triantafyllou (2024) that there is value in understanding these reform models as conceptual models to explore structures and dynamics within public administrations, including to understand how evaluations are commissioned. In the next section, I identify specific characteristics of the NWS model that are relevant to public administrative environments where evaluations are commissioned and consider how they can provide opportunities to commission more democratic and useful evaluations.

Neo-Weberian characteristics of bureaucracy

The Neo-Weberian model of the state emphasizes the continued relevance of the following aspects of Weber's ideal type of bureaucracy:

- the important and irreplaceable role of the state;
- the value of administrative and representative law;
- a distinctive professionalized civil service with task specialization;
- the centrality of formal rules and bureaucratic regulation; and
- clearly defined organizational structures and lines of accountability.

Pollitt and Bouckaert (2017) identify distinctive “neo aspects” within the NWS model, which respond to significant recent changes in the digital, societal and environmental contexts in which public sector bureaucracies work. Three “neo aspects” are particularly relevant to commissioning evaluations: a focus within public administrations on understanding and meeting citizens' needs, professionalization of leadership and management; and a motivation to deliver results for citizens.

Increasing focus on understanding and meeting citizens' needs

Citizens are a heterogeneous group of people. They include direct beneficiaries; indirect beneficiaries; and others who have an interest in or relationship with a policy or program, such as through their status as taxpayers or residence within a country. In the NWS model one of the primary motivations within contemporary public administrations is meeting the needs of citizens. This increasing orientation toward meeting citizens' needs has been influenced by changing societal norms about rights, roles and responsibilities of citizens and the emergence of representative democracy activities. This includes increasing opportunities and societal motivations for citizens to articulate their needs, share their views and hold elected governments to account through digital technology, community-based activities and direct action.

Pollitt and Bouckaert (2017) argue that public administrative reforms that are informed by the NWS model facilitate a culture of quality and public service that focuses on understanding and meeting the needs of citizens and responding to their feedback. This in turn, facilitates an environment where civil servants are motivated to work in more creative, agile and dynamic ways to better understand and respond to citizens voice and needs. This has specific implications for commissioning evaluations. It is conducive to supporting the development of working environments where commissioners are incentivized to develop in-depth knowledge about interventions being evaluated and evaluation as a discipline. It also incentivizes commissioners to listen to feedback from citizens about their concerns, interests and better understand the perspectives of direct beneficiaries in particular, as they have the most in-depth knowledge of the policy or program being evaluated. As such, reforms in public administrations that are informed by

the NWS model are conducive to organizational cultures where staff are supported to commission evaluations that reflect the concerns, needs and interests of citizens and generate timely, relevant and insightful findings in response to these needs.

Citizens panels have been operationalized in France since 2020 as a method of increasing direct democracy in a range of policy areas, including the development of a national strategy to reduce greenhouse gas emissions. In 2021 Paris created a permanent citizens' council to ensure citizen-led voice and direct democracy inform local legislation affecting Parisians. This approach is aligned with the French Republic's *Declaration of the Rights of Man and Citizens* (1789) which states "society has the right to require of every public agent an account of its administration" and the country's aim of increasing citizen engagement in democratic decision-making. As part of France's national strategy for the prevention and fight against poverty, the government created a citizens' panel of people who had or were currently experiencing poverty to ensure unique insight, lived experience and expertise from those affected by poverty informed the government's actions. A subgroup of eleven people from this panel were identified to create an evaluation citizens' panel with responsibility for identifying evaluation priorities and providing advice about evaluation approaches. Members of the panel were specifically asked to reflect on their identity as French citizens who had experiences of poverty, policies that had aimed to tackle poverty and apply their values, experiences and views to inform decisions about evaluations. All members received training about the role of evaluation, evaluation approaches, questions and techniques prior to their participation in the panel to support their ability to provide informed, relevant and useful feedback.

The panel highlighted the value of the evaluation strategy including a qualitative approach and reflecting on the contribution of different community groups, voluntary organizations and family networks to preventing and addressing poverty. The panel also noted the importance of the evaluation strategy exploring the role of policies that were not specifically included in the national poverty strategy but may have contributed to a reduction in poverty. The French Government's assessment of this approach concluded that it played an important role in ensuring the evaluation activities reflected the voices, needs and interests of citizens (République Française, 2024). It was seen as a form of direct democracy within the design of the evaluation strategy and approach. While this way of working did not solely focus on commissioning evaluations, it can be applied to democratize and increase citizen voice and engagement – particularly for direct beneficiaries – in the commissioning process.

Other participatory strategies can be applied to increase the voice and role of citizens, including direct beneficiaries and groups that represent them, within commissioning activities. Web-based consultations with stakeholders, co-creation within evaluation commissioning processes, and joint evaluation activities can help to orient civil servants to commission evaluations that better understand and reflect the voice and needs of citizens, particularly direct beneficiaries. For example, the United Kingdom's Independent Commission for Aid Impact (ICAI) organizes regular online public consultations to seek feedback from citizens about which

aspects of the UK's international development program should be reviewed (United Kingdom, 2024). This way of working aims to increase citizen voice, inclusion, engagement and insights from the voluntary sector, international development research institutes and beneficiary communities to improve relevance, democratic engagement and the utility of ICAI's evaluative reviews.

UNICEF's Participatory Approaches Methodological Brief (Guijt, 2014) provides detailed guidance on different participatory approaches that can be utilized within evaluations to engage and empower direct beneficiaries of a program or policy. This guidance notes the importance of asking the reflective question of "who should be involved, why and how" at each stage of the design and development of an evaluation. While this guidance is focused on evaluation in general, these reflective questions can be applied within all the stages of commissioning evaluations to support timely and appropriate engagement with citizens that are direct beneficiaries. Commissioning evaluations that reflect and the voice and values of citizens, and particularly direct beneficiaries, is highly aligned with the NWS model's focus on understanding and responding to the needs and feedback from citizens. Direct beneficiaries are able to provide uniquely valuable contributions to commissioning activities because they have lived experience and detailed knowledge of the context, their own needs and the policy or program.

Increasing orientation toward understanding the achievement of results in ways that are contextualized, holistic and meaningful

Pollitt and Bouckaert (2017) argue that within NWS-oriented public administrations, staff are managed and incentivized to achieve results that are relevant to the strategic goals of the organization. Structuring and delivering work in accordance with results targets and frameworks is seen to incentivize civil servants to design and deliver relevant, effective and efficient public services. Evaluation commissioners can play a critical role in supporting public administrations to understand the value, relevance and impact of results for different groups of citizens as well as exploring design, process and context. Evaluations can be commissioned to ask critical evaluative questions and to contextualize analysis of results rather than focus on traditional results-based questions. For example, the commissioning process can ask evaluative questions about whether and why results frameworks and targets are realistic, appropriate in relation to different contexts, the needs and perspectives of beneficiary communities and how they could be improved.

Similarly, commissioning activities can support public administrations to gather insights and support interpretation in relation to the achievement of results. Evaluations and evaluative activities that consider these questions have an important role to play in enabling organizations to develop relevant and tailored results frameworks and deliver results in ways that are meaningful and relevant to their citizens and direct beneficiaries in particular. Supporting commissioners to play this valuable role, rather than simply designing evaluations to assess whether results targets have been achieved requires organizations to provide training and

support commissioners to develop critical evaluative thinking skills and evaluation approaches.

Two examples of evaluations that have provided insightful analyses of organizational results frameworks and results-based management systems are: the FAO's evaluation of its Strategic Results Framework (2019) and the Council of Europe's evaluation of its Results Based Management system (Balogun and Kabell, 2021). Both evaluations conclude that the RBM systems in these organizations developed and matured significantly in recent years due to organizational commitments and investments. They also found that these systems were primarily focused on and successfully meeting accountability needs and in the case of the Council of Europe organizational communications needs. Balogun and Kabell (2021) note that due to the mandate and structure of the Council of Europe, meeting accountability and communication needs are critical outcomes for the organization's RBM system. Similarly, the independent evaluation of the FAO's RBM system (FAO, 2019) recognizes the importance of the RBM system enabling the organization to be accountable to its members.

These evaluations identify opportunities for the organizations to critically reflect on how their RBM systems could be developed to support strategic learning and contribute to more reflective and evaluative organization-wide cultures. They highlight the merit of the organizations developing their systems to be more contextualized, and capture data to support better understanding of whether, how and why outcomes are being achieved for different communities. Balogun and Kabell (2021) note that there is value in the Council of Europe developing standardized indicators that are aligned with an over-arching organizational theory or theories of change to improve organizational understanding about how results are contributing to strategic objectives. They also identify the benefits of applying contribution analysis to help the Council of Europe understand if, how and why delivery of results has led to changes for different groups and in different contexts at the intermediate objective level. The evaluation notes that this would support the RBM system to be more useful for organizational learning, informed decision-making and wider knowledge generation.

Similarly, the evaluation of the FAOs' RBM system (FAO, 2019) notes that while the organizational results framework has become more strategic and aligned with the Sustainable Development Goals (SDGs) it lacks alignment with wider organizational narrative and theory of change. The evaluation shows that there are opportunities for this system to identify contribution pathways and inform wider organizational analysis about the extent to which, how and why organizational intermediate outcomes have been achieved. The evaluation concludes that this type of reorientation would enable the RBM system to collect evaluative evidence to help the FAO understand how, why and the extent to which meaningful change occurred and in which contexts. It also recognizes the value of RBM systems providing qualitative insights about design issues, contribution, causal pathways and change rather than aggregate reporting to demonstrate results.

These evaluations illustrate the valuable role that public administrations can play by commissioning evaluations that critically reflect on results frameworks and RBM systems and exploring the extent to which they capture the needs, voice and priorities of citizens and sufficiently consider context, contribution and design. This can enhance the voice of beneficiaries within the commissioning process and enable organizations to collect and use evidence about program design and results that are credible, contextualized and reflect democratic voice and values.

Professionalization of leadership and management

A key feature of Weber's ideal type of bureaucracy is the use of formal rationality to make decisions. Weber defines formal rationality as actions that happen as a result of a means-end rational calculation – rational steps informed by evidence, logic and science to achieve a goal (Weber, 1997). However, March and Olson (2011) argue that in response to technological, societal and environment developments decision-making within NWS oriented public administrations is grounded in Weber's (1997) logic of rationalization with the logic of consequences and of appropriateness. This includes consideration of values embedded within legal frameworks such as human rights and equality. According to the United Nations Commission on Human Rights (2002), equality and respect for human rights are essential elements of democracy. Applying a logic of rationality within the evaluation commissioning process involves considering the value of results achieved, if and how an intervention has considered human rights and equality as well as the achievement of its goals. As such, it is conducive to commissioning more holistic, democratic and valuable evaluations. This approach is often termed "reflexive rationality" and applying this logic of rationality can support commissioners to ensure human rights, diversity and equality commitments that are articulated within globally ratified agreements and national laws are considered in their work. This supports the commissioning process to give voice and representation to marginalized, more vulnerable and excluded communities. It also enables evaluation to play a role in holding democratically elected governments to account for meeting their global commitments and national laws on equality and human rights.

The NWS model highlights the role that increasing professionalization of leaders and managers plays in limiting the risk of NWS-based public administrations being overly process-focused and applying rules in rigid ways. Coaching styles of management, democratic and consultative leadership styles, and inclusive decision-making practices embrace democratic values. They emphasize the importance of human-centered management, equity and inclusivity within decision-making. They also aim to empower staff and delegate decision-making with strategic guidance (Grant, 2021). Large public administrations including the United Kingdom's Civil Service (UK Government, 2018) and the Organisation for Economic Co-operation and Development (OECD, 2022) identify these ways of working within their

professional competency frameworks. While these leadership styles and behaviors cannot be assumed to be practiced within NWS-based public administrations, the structures and goals of these organizations are more likely to reward and support them. Increasing professionalization within public administrations facilitates working environments where staff are motivated and empowered to apply organizational rules, including those about evaluation commissioning processes and practices, in pragmatic and contextually relevant ways.

Practical ways to apply an NWS lens to commission more democratic and useful evaluations

Understanding and reflecting on the model of the NWS provides a range of insights about how evaluations can be commissioned to be more democratic and valuable to public administrations. In particular, this appendix highlights the following four approaches that can be employed within commissioning activities:

Be citizen-led – and direct beneficiary led in particular – when commissioning evaluations

Engage with citizens and direct beneficiaries in particular when designing and implementing commissioning activities, including work to develop evaluation priorities and questions. This is conducive to commissioning activities being more democratic as they reflect the voices, views and perspectives of citizens. Given their in-depth experience of policies or programs it is particularly important to engage direct beneficiaries as a distinct group of citizens during commissioning activities.

Explore results frameworks and systems, including the extent to which they consider context, design and processes within commissioning activities

Reflect on the appropriateness and value of organizational results targets, frameworks and systems, including how they consider context, beneficiary needs and design issues when commissioning evaluations. Reflecting on these organizational tools when designing and implementing commissioning activities can provide valuable insights about they could be more meaningful, useful and relevant to beneficiaries' needs, priorities and values.

Apply reflexive rationality to consider if and how policies support human rights and equality within the context of national laws and global commitments

Explore how the commissioning process can give voice and representation to more marginalized and excluded communities by reflecting on whether and how policies support equality and respect for human rights, which are essential elements of democracy. This can support public administrations to understand how they can meet

commitments and enable citizens to hold their governments to account on these elements of democracy.

Explore how professional leadership and management styles within the organization can support evaluation commissioning activities to be creative and agile

Understand leadership approaches, management styles and subsequent opportunities to commission tailored, contextualized and creative evaluations that support democratic voice and accountability. This involves commissioners being trained in evaluation as a discipline, evaluative thinking and participating communities of practice to share learning and good practice.

Conclusion

The NWS model's focus on meeting citizens' needs, delivering results and professional leadership skills can be applied to support evaluation commissioning activities to be more democratic and valuable to public administrations and the democracies that they serve. Techniques such as citizens panels, public consultations and participatory approaches can be integrated into the design and delivery of commissioning activities. These ways of working provide opportunities for citizens – particularly those with valuable lived experience of the program or policy being evaluated – to play an active role in the commissioning process. They support democratic processes and values to be embedded within evaluation commissioning activities by creating opportunities for knowledge, views and expertise of groups of citizens with the most insightful knowledge to inform the design and implementation of commissioning activities.

Commissioning evaluations in ways that support organizations to critically reflect on results targets, frameworks and results-based management, including how they consider context, design issues and contribution, can help these tools become more beneficiary-centric. They can support decision-makers to understand how results-based management systems can better understand program design, processes and results being achieved in relation to the needs and priorities of their citizens and direct beneficiaries in particular.

The NWS model also provides valuable insights about how commissioners of evaluations can apply reflexive rationality to support evaluations to give voice and representation to more marginalized communities. By applying the logic of consequences and appropriateness, evaluation commissioning activities can explore whether and how policies are supporting the achievement of national and global commitments to equality and respect for human rights which are fundamental aspects of democracies. Increasing professionalization of leadership, more democratic management approaches and higher levels of autonomy for staff are also aligned with the NSW model. They are conducive to creating environments that facilitate more creative and agile evaluation commissioning activities which in turn support the design and delivery of evaluations that are more democratic, tailored and useful.

The influence of the NWS model has been evident in many recent reforms of public administrations in economically developed countries. This appendix highlights the merit of applying insights from the model to increase democratic voice, processes and values within commissioning activities. This can support the generation of valuable evaluative evidence to inform decision-making, improve transparency and enhance accountability of democratically elected governments.

Note

- 1 The opinions expressed in this chapter are the author's own and do not reflect the views of the OECD.

References

- Andrews, Rhys, Phillipe Bezes, Gerhard Hammerschimid, and Steven Van de Walle. 2016. *Public Administration Reforms in Europe: The View from the Top, Conclusion: A Kaleidoscope of Administrative Reforms in Europe*. United Kingdom: Edward Elgar Publishing.
- Balogun, Paul, and Dorte Kabell. 2021. *Evaluation of Results Based Management in the Council of Europe*. Strasbourg, France: Council of Europe.
- Bouckaert, Geert. 2022. *The Neo-Weberian State: From Ideal Type Model to Reality?* London: Institute for Innovation and Public Purpose, University College London.
- Food and Agricultural Organisation (FAO). 2019. *Evaluation of FAO's Strategic Results Framework*. Rome, Italy: Food and Agricultural Organisation.
- France. 1789. *Declaration of the Rights of Man and of the Citizen*. 26 August (AE/II/1129).
- Grant, Adam. 2021. *Think Again: The Power of Knowing What You Don't Know*. New York: WH Allen Publishing.
- Guijt, Irene. 2014. *Participatory Approaches, Methodological Briefs: Impact Evaluation 5*. Florence, Italy: UNICEF Office of Research.
- Krogh, Andreas Hagedorn, and Peter Triantafillou. 2024. "Developing New Public Governance as a Public Management Reform Model." *Public Management Review* 26: 1–17.
- March, James, and Johan Olson. 2011. "The Logic of Appropriateness." *The Oxford Handbook of Political Science*, edited by Robert Goodin. Oxford: Oxford University Press.
- Organisation for Economic Co-operation and Development (OECD). 2022. *Core Competencies Recruitment Framework*. Paris: OECD.
- Osborne, Stephen. 2006. "The New Public Governance?" *Public Management Review* 8(3): 377–387.
- Osborne, Stephen, ed. 2010. *The New Public Governance: Emerging Perspectives on the Theory and Practice of Public Governance*. New York: Routledge.
- Pollitt, Christopher, and Geert Bouckaert. 2011. *Public Management Reform: A Comparative Analysis – NPM, Governance and the Neo-Weberian State*. 3rd ed. Oxford: Oxford University Press.
- Pollitt, Christopher, and Geert Bouckaert. 2017. *Public Management Reform: A Comparative Analysis – Into the Age of Austerity*. 4th ed. Oxford: Oxford University Press.
- République Française. 2024. *Participatory Evaluation: The Voice of People in Poverty*. Paris: République Française.

- Torring, Jacob, Lotte Andersen Bogh, Carsten Greve, and Kurt Klaudi Klausen. 2020. *Public Governance Paradigms: Competing and Co-Existing*. United Kingdom: Edward Elgar Publishing.
- UK Government. 2018. *Civil Service Competency Framework*. United Kingdom: UK Government.
- UK Government. 2024. *International Commission for Aid Impact (ICAI) Public Consultation about Review Priorities*. United Kingdom: UK Government.
- United Nations. 2002. *Democracy and Human Rights: The Human Rights Normative Framework*. Accessed December 20, 2024.
- Weber, Max. 1997. *The Theory of Social and Economic Organisation*. New York: Free Press.

Index

Note: Page locators in **bold** and *italics* represents tables and figures, respectively.

- Academy of Evaluation (Poland): origins and founding of 189–190; target audience and participants of 191–193; pedagogical model of 194–196; curriculum of 197–198; impact on evaluation system 199–200; cited in chapter 15 201–202
- Academy of Regional Development (Poland) 190
- accountability: downward vs. upward 31; and learning, balancing 44, 50, 53, 57, 147–148, 150, 156; democratic 104, 106, 122, 127, 207, 219, 226, 234, 237, 240; results-based 122, 127
- agency: of commissioners 6, 9, 16–17, 161, 163–165, 171; feminist and political theory of 164
- Agency for Development Cooperation (Norway) 187
- Agriconsulting 124
- American Evaluation Association (AEA) 19
- Andrews, Rhys 233
- Annual Seminar on Experiences of Evaluation (Spain) 165
- Anti-Homosexuality Act of 2023 (Uganda) 86
- Applebaum, Anne 228
- Applying Evaluation Criteria Thoughtfully* (OECD/DAC, 2021) 30
- appreciative inquiry 33–34
- APROEVAL (Iberian Association of Evaluation Professionals) 165, 172
- Argyris, Chris 66, 73
- Artificial intelligence in evaluation 25, 58–59, 98, 200
- audit society 7
- AusTender (Australia) 38
- Australasian Evaluation Society (AES) 7, 19
- Australia 92
- Autocracy, Inc.* (Applebaum, 2024) 228
- autocratic evaluation *see* MacDonald’s typology of evaluation
- average effects, focus on *see* heterogeneity, attention to
- backward design (pedagogy) 195
- Balogun, Paul 237
- Bastøe, Per Øyvind cited in chapter 1 9, 13, 16, 19; co-author of Chapter 13 (Training for Critical Evaluative Thinking) 177, 187; cited in chapter 15 208, 210–211, 218–220, 223, 224, 228
- beneficiaries *see* right holders
- Bernard et al. (2009) 196
- bespoke evaluation (“made-in-measure” approach) 10, 28–29, 33
- Bezes, Philippe 233
- BIPOC (Black, Indigenous and People of Color) 67
- Blair, Tony 96
- Boggs, Grace Lee 68
- Boschen, Kathryn A. 7, 26, 28
- Bouckaert, Geert 232–234, 236
- Boyte, Harry C. 164, 172
- Brazil 82–83
- bureaucratic evaluation *see* MacDonald’s typology of evaluation
- Bureau of International Information Programs (US) 42
- Buscemi, Virgilio: cited in chapter 1 4, 9, 13–14, 16, 18; co-author of Chapter 9

- (Italy's Cohesion Policy) 116–118; cited in chapter 15 218–222, 225
- Bustelo, Maria: as co-editor, in chapter 1 1–9; cited elsewhere 132, 139, 189; co-author of Chapter 12 ('The good commissioner') 161–172; co-author of Chapter 15 207–228
- Camel train (evaluation questions metaphor) 57
- Canada 41, 43, 95–96
- CanadaBuys (open.canada.ca) 38
- Canadian Evaluation Society (CES) 7, 19
- Canadian Journal of Program Evaluation 100
- capacity-building: of commissioners and commissioning authorities 42, 87, 146–147, 156, 171, 189–200, 225, 228; evaluation systems, of 177–182, 186
- Carnegie UK Trust 61
- Catsambas, Tessie (Tzavaras): cited in chapter 1 2, 5–9, 11–14, 19; on commissioning stages (Perrin & Boschen tradition) 26–27, 34; co-author of Chapter 6 ('The influence of context...') 84–86, 88; cited in chapter 15 208, 210–211, 218–220, 223, 224, 228
- causal density 52–53, 222
- Center for Global Development 147, 158
- Champions (evaluation leadership) 95
- Changing Bureaucracies* (Perrin & Tyrrell, 2020) 16
- Chelimsky, Eleonor 1, 6
- Chile 83
- Chouinard, Jill A. 66
- citizens' juries and panels 13, 54, 235, 240
- civic agency 1, 164
- civil society 31, 80, 201, 212
- client-service relationship 5–6, 11, 117
- Cochrane Collaboration 44
- co-design (Terms of Reference) *see* Terms of Reference (ToRs)
- Cohesion Policy (EU): Learning Agendas and (Chapter 8) 106; evaluation contracting under, Italy 116–128; Academy of Evaluation and (Chapter 14) 189–190, 192, 198–200; cited in chapter 15 222
- Collington, Rosie 128
- COM-B model (Capability-Opportunity-Motivation-Behavior) 102, 104
- commissioning authorities: Learning Agendas and (Chapter 8) 102, 104, 109–110, 113; Academy of Evaluation and (Chapter 14) 189–190, 192–193, 199–200
- commissioning (evaluation): agency of commissioners in 6, 9, 16–17, 161, 163–165, 171; authority and power in 25–26, 28, 34–35, 92–93; bureaucratic 207–209, 212, 213, **214**; collaboration between commissioners and evaluators 5–6, 9, 11–12, 68–73, 146–158; contracting distinguished from 3–4, 25, 117; definition of 3–4, 25, 41; democratic, *see* democratic commissioning; flexibility in 1–2, 5, 9–10, 14, 28–29; four-stage model of (Perrin & Boschen) 7, 26; good practice in, Table 15.1 traits of **214**; market-based 116–131; outsourcing vs. in-house 4, 121–128, 132–145; as a relationship, not a contract 3–4, 25, 208; technocratic 210–211–13, **214**; Terms of Reference and *see* Terms of Reference (ToRs)
- "Commissioning Great Evaluations" (Catsambas & Davidson, 2024) 208, 210–11, 218, 220
- Committee for Development Policy (UN) 96
- Common Agricultural Policy (CAP) 126–127
- communications strategy (evaluation dissemination) 56, 58, 73
- community power (RWJF initiative) 67
- community scorecards 37
- competencies frameworks (evaluators and commissioners): cited in chapter 1 2, 7–9, 16; Bustelo, Crespo & Ligerio framework of, incl. Spanish national framework (Chapter 12) 161–176; cited in chapter 15, generally 207–228; seven categories of, dos and don'ts (Table 15.3) **226–227**
- complexity (in evaluation) 1, 4, 10, 90–91, 98
- Complutense University of Madrid (UCM) 9, 162, 165, 224
- confidentiality 39
- conflicts of interest (consortium companies) 138
- contribution analysis 10, 234
- Council of Europe 237
- Council of the European Union 190
- COVID-19 pandemic 69, 89, 96–97, 99
- Crespo, Ramon: cited in chapter 1 9; co-author of Chapter 12 ('The good

- commissioner') 161–172; cited in chapter 15 216–228
- critical evaluative thinking: as chapter topic (Chapter 13) 177–188; cited elsewhere 9, 218
- Cronbach, Lee 1–3, 6
- culturally responsive evaluation (CRE) 64
- Dahler-Larsen, Peter 7, 83, 184–185
- Data Foundation (US) 106, 114
- Datta, Lois-Ellin 2, 6, 18–19
- Davidson, Jane 26–27, 34, 88, 208, 210–11, 220
- decolonizing evaluation 18
- deliberative democracy: as democratic evaluation framework (Chapter 3) 37–38, 42; in commissioning with clarity (Chapter 4) 49–50, 53–54, 57, 61; cited elsewhere 11, 209, 212, **216**
- Democracy indicators 61, 83
- democratic accountability *see* accountability
- democratic commissioning: in Finland EMS case study (Chapter 10) 132, 137, 139, 141–142, 144; central concept in chapter 15 207–228
- democratic evaluation: definition and traditions of (populist and public interest) 1–2, 6, 19; three essential elements of (House & Howe) 37–38; typology of *see* MacDonald's typology of evaluation; cited across commissioning chapters 25, 34–35, 61, 79–80, 111, 113, 161, 171–172
- democratic professionalism 208, 211–212, 215, 228
- Democratic Professionalism* (Dzur, 2008) 228
- democratic space(s) 49–53, 58
- Department of Cohesion Policy (Italy) 118–120
- Development Evaluation Norm (Finland MFA, 2015) 144–145
- Development Evaluation Unit (Finland MFA): cited in chapter 1 14; role within Evaluation Management Services 132–140; knowledge management by 141–145
- Dewey, John 211, 228
- DFØ (Norwegian Directorate for Financial Management): role in Norad evaluation series 150; co-organizer of critical evaluative thinking course 178–182, 184–185, 187
- digital divide 43
- discourse theory 52
- discursive evaluation 38, 51–52, 217
- do-no-harm principle 87
- double-loop learning *see* single, double and triple loop learning
- Duffy 179
- Dzur, Albert (W.) 208, 211–212, 228
- Early Head Start Program 43
- echo chamber effect 59–60
- Economist* Democracy Index 83
- Ecorys 124
- EES (European Evaluation Society) 163, 166
- elitist democracy-oriented evaluation 38, 51, 215
- emerging evaluators 14, 136–137, 140, 143–144
- engagement *see* stakeholder engagement
- equitable evaluation: framework and principles of (Equitable Evaluation Framework) 63, 65–66; housing justice case study (Chapter 5) 64, 67–73; cited in chapter 15 210, 213, 214, **214**, 219, 221
- Equitable Evaluation Framework 63–65, 67
- Equitable Evaluation Initiative (EEI): framework and principles of 63–66, 73; cited in chapter 15 210
- ethics (in evaluation) 87, 167, 170, 227
- EUROREG (University of Warsaw) 190, 195, 202
- EvalForEarth 97
- EvalNet 97
- EvalPartners 87, 97
- EVALSDGs 97
- Evaluating and Valuing in Social Research* (Schwandt & Gates, 2021) 16, 66–67
- Evaluating with Validity* (House, 1980) 224
- evaluation capture 3, 7, 12, 18, 163–164, 207, 212
- Evaluation Management* (Catsambas & Davidson, 2024) *see* "Commissioning Great Evaluations" (Catsambas & Davidson, 2024)
- Evaluation Management Services (EMS): cited elsewhere 14, 210; structure and consortium of 132–135; operation and governance of 136–140; assessment and knowledge management 141–145
- Evaluation Officer Council (US) 103, 114

- evaluation policy 81, 96, 208–210
 Evaluation Quality Standards (OECD/
 DAC, 2010) 29–30
 evaluation society (concept) 7
 evaluation systems: cited elsewhere 6–7,
 12–13, 204, 207–208, 214; institutional
 context of (Chapter 6) 81, 84, 96,
 99–100; in Norway 184–185; in Poland
 190–191, 199
 evaluation use and influence 32–33, 35,
 225, 228
 evaluators: as facilitators of learning 67, 97,
 99; resilience of 164; role of, contrasted
 with commissioners 161, 163, 168–171,
 213–214; selection of 117, 119, 127
 Evans et al. (2020) framework 41–42
 Evidence Act (US, 2018) 84, 103, 111,
 114–115
 evidence-based approaches to evaluation 6–7
 evidence-based policymaking 84, 102–103,
 106, 111, 113–115
 evidence ecosystem 49, 58
 ex ante evaluation services 119
 Experiential Learning Model (Kolb) 196
- feminist theory (of agency) 164
 Finland: cited elsewhere 14, 98; Ministry
 for Foreign Affairs evaluation case study
 132–145
 fiscal year budget cycles 41
 Fishkin, James 49, 209, **215**
 five questions for commissioners (Ling &
 Millard framework) 51–57
 flexibility (in contracts and evaluation
 design) 1–2, 5, 9–10, 14, 28–29
 Food and Agriculture Organization (FAO)
 92, 237
 Ford Foundation 85
 Fossetøl, Knut cited in chapter 1 9, 13, 16,
 19; co-author of Chapter 13 (Training for
 Critical Evaluative Thinking) 177; cited
 in chapter 15 217–218, 223, 228
 fossil fuel subsidies 38
 Foundations for Evidence-Based
 Policymaking Act *see* Evidence Act (US,
 2018)
 four-stage model of commissioning *see*
 commissioning (evaluation)
 fragile and conflict-affected contexts 86
 France 232
 Freedom House 83
Freedom in the World Report 83
 Frierson, Henry T. 64
- gender-based analysis 41, 170, 175
 gender equality 170, 175
 “Getting the Best from Consultants” (Perrin
 & Boschen, 1991) 7, 26, 28
 good commissioner *see* competencies
 frameworks (evaluators and
 commissioners)
 Gorzelak, Grzegorz 202
 governance of evaluation process 38–39
 GRADE (Grading of Recommendations,
 Assessment, Development and
 Evaluation) 44
 Grant, Adam 235
 Greene, Jennifer 1–2, 6
 Guijt, Irene 233
- Hanberger’s typology of democracy-
 oriented evaluation 38, 209, **215**
 hard-to-reach populations 53, 58–60
 Hart, Nicholas 103, 106
 Hassnain, Hur 11, 19
 Herfindahl–Hirschman Index (HHI) 125
 heterogeneity, attention to 10, 37, 42–46,
 146
 Hill, Michael 5
 Hirschman, Albert O. 5
 Hopson, Rodney K. 64
 House, Ernest 6, 37–38, 224
 Housing justice 15, 64, 67–72
 Howe, Kenneth 37–38
 Hungary 85
 Hupe, Peter 5
 Høydal, Øyunn Syrstad cited in chapter
 1 9, 13, 16, 19; co-author of Chapter
 13 (Training for Critical Evaluative
 Thinking) 177–182; cited in chapter 15
 217–218, 223, 228
- IBM Center for the Business of
 Government 106, 114
 ICAI (Independent Commission for Aid
 Impact, UK) 235–236
 implementation (of evaluated programs) 5
 implications for evaluators and
 commissioners (Housing Justice case)
 72–73
 inception report 152, 156
 inclusion, values of 10–11, 41–42
 independence of evaluators 152, 156, 158
 “Independence: When, Why and How
 It Matters for Evaluation Credibility”
 (Lindkvist & Kennedy-Chouane, 2024)
 156, 158

- India 92
- Indigenous evaluation 18, 39–42, 47
- innovation systems 54–55
- Institute for Public Policy Evaluation (Spain) 172
- institutionalization of evaluation 3–4, 116, 204
- intellectual property (in evaluation contracts) 68–69
- International Financial Institutions (IFIs) 84–85
- International Institute for Environment and Development (IIED) 90, 97, 100
- International Organization for Cooperation in Evaluation (IOCE) 97
- INTEVAL: as network informing this volume's methodology 4, 8, 18–19; meetings of (Rome, The Hague, Washington) 215–216; cited in chapter 15 216
- invitation to tender (ITT) 56
- IRS (evaluation contractor) 124
- Ismeri 124
- ITAD: collaboration with Norad 11, 146–150; role across evaluation stages 151–158
- Italy: Cohesion Policy evaluation market in 116–120; regional variation in evaluation expenditure 121, 123–126; Southern Italy, democratic deficit in 122, 127–128; cited in chapter 15 221
- Joint Committee on Standards for Educational Evaluation 19
- Kabell, Dorte 234
- Kantola, Johanna 164
- Kennedy-Chouane, Megan 156, 158
- knowledge brokers: commissioners and evaluation units as 104, 146, 150; Academy of Evaluation curriculum on 189, 194–195, 197, 199–200; cited in chapter 15 219
- Kolb, David A. 196
- Koleros, Drew: cited in chapter 1 9, 12, 15; co-author of Chapter 5 (equitable evaluation platform) 210, 213, **214**, 218, 219, 221, 223–225; cited in chapter 15 218–219, 223–225
- Kozak, Marek 189, 202
- Krogh, Andreas 233
- Kupiec, Tomasz cited in chapter 1 13, 16, 21; co-author of Chapter 8 (Learning Agendas) 102; 189, 202–203; cited in chapter 15 210, 218–219, 223
- Laaksonen, Sari: cited in chapter 1 3, 13–14, 16; co-author of Chapter 10 (Finland EMS) 132–140; cited in chapter 15 210, 218, 219, 221, 222, 223, 225
- Lahey, Robert: cited in chapter 1 6, 10–11, 15; co-author of Chapter 7 (Democracy, SDGs and commissioners) 89–100; cited in chapter 15 218, 220, 223–224
- Lattanzio (KIBS SpA) 118, 124
- Learning Agenda Design Sprint (LADS) 106, 108
- Learning Agendas: cited in chapter 1 13, 16; as evaluative tool, theory of change for (Figure 8.1) 102–105; design of, options and steps (Table 8.1) 106–115; cited elsewhere 210
- Learning Briefs (Norad) 155
- Leeuw, Frans 7, 12, 18–19
- Leigh, Andrew 10, 43
- Ligero, Juan Andrés: cited in chapter 1 9; co-author of Chapter 12 161–172; cited in chapter 15 216–228
- Lindkvist, Ida: cited in chapter 1 9, 11–12, 16; co-author of Chapter 11 (art of commissioning) 146–158; cited in chapter 15 218–219, 220–221
- Ling, Tom: cited in chapter 1 9–11, 15; co-author of Chapter 4 (commissioning with clarity) 49–60; cited in chapter 15 209, 213, **214**, 215, 218–225
- Liston, Monique: cited in chapter 1 9, 12, 15; co-author of Chapter 5 (equitable evaluation platform) 210, 213–214, **214**, 218, 219, 221, 223; cited in chapter 15 218–219, 223–225
- Lloyd, Rob: cited in chapter 1 9, 11–12, 16; co-author of Chapter 11 (art of commissioning) 146–158; cited in chapter 15 218–221, 222–225, 228
- localization strategy (in foreign aid) 86
- Lombardo, Emanuela 164
- Lucks, Dorothy cited in chapter 1 6, 10–11, 13, 15; co-author of Chapter 7 (Democracy, SDGs and Commissioners) 89, 96; cited in chapter 15 218, 220, 223–225, 228
- MacDonald, Barry 6, 37, 208–209, 211, 214–215, 228

- MacDonald's typology of evaluation 37, 207–208, 211, 214–215, 228; autocratic evaluation 37, 208, 215, 228, **215**; bureaucratic evaluation 37, 207–208, 209, 212–213, **214**, **215**, 221; democratic evaluation, as third type 37, 210–211; adaptation to commissioning (Table 15.2) **215**
- March, James 238
- marginalized populations and voices:
engagement of, as democratic principle 10, 18, 31, 37, 80; equitable evaluation and 64–65; Neo-Weberian State model and 240, 241
- market-based evaluation *see* commissioning (evaluation)
- Marra, Mita: cited in chapter 1 4, 9, 13–14, 16, 18; co-author of Chapter 9 (Italy's Cohesion Policy) 116–118; cited in chapter 15 218–222, 225
- Marshal's Office (Poland) 193
- Mathematica 67–69
- Mayne, John 100, 150
- Mazzucato, Mariana 128
- Mbava, Nombeko P. 83
- McGhee, Ray: cited in chapter 1 9, 12, 15; co-author of Chapter 5 (equitable evaluation platform) 210, 213–214, **214**, 218; cited in chapter 15 218–219, 223–225
- McTaggart, Robin 40
- meta-evaluations 146
- methodological plurality 127–128
- middle democracy (Dzur) 211
- Millard, Ananda: cited in chapter 1 9–11, 15; co-author of Chapter 4 (commissioning with clarity) 49–60; cited in chapter 15 209, 213, **214**, 218, 219, 221–225, 228
- mind maps (pedagogical tool) 192, 194, 196, 198
- Ministry for Foreign Affairs of Finland (MFA): cited elsewhere 14, 210; Development Evaluation Unit of 132–140; knowledge management within 141–145
- Ministry of Finance (Norway) 178–179, 187
- Ministry of Foreign Affairs (Norway) 147, 156–157
- Ministry of Regional Development (Poland) 190, 193
- Monitoring and Evaluation Act 2024 (Nepal) 91
- Monitoring, Evaluation and Learning (MEL) 148, 155
- Montague, Steve: cited in chapter 1 9–11, 15; co-author of Chapter 3 (evaluation commissioning and democratic processes) 37–40, 51; cited in chapter 15 209, 213, **214**, **215**, 218–221, 224
- moral rights (in evaluation contracts) 40
- Morgenbladet* (Norwegian newspaper) 146, 157–158
- National Evaluation Plan (South Africa) 38–39
- Neo-Weberian State (NWS) model: cited in chapter 1 and chapter 12, 16, 223; as Appendix framework (Pollard), overview of 231–241; citizen needs, orientation toward 233–236; professionalization of leadership under 235, 238–239; reflexive rationality and 238, 239; cited elsewhere 240–241
- Nepal 91
- Newcomer, Kathryn 103, 106
- New Public Governance (NPG) 233
- New Public Management (NPM) 233
- Nigeria 98
- “Ninety-five Theses” (Cronbach, 1980) 2–3
- NIRAS Finland Oy 133, 143–144
- Norad (Norwegian Agency for Development Cooperation): cited in chapter 1 11; evaluation series of (2014–2024) 146–148, 150; commissioner-evaluator collaboration model 149, 151–158
- Norway: Norad evaluation series in 146–158; critical evaluative thinking course in 177–188
- “Nothing as Practical as Good Theory” (Weiss, 1995) 181
- OECD (Organisation for Economic Co-operation and Development): cited elsewhere 19, 238–239, 241; evaluation quality standards of 29–30, 152; peer reviews and country contexts 82, 88, 96, 146, 148, 158; cited in Chapter 13 (Norway training course) 177–181, 188
- Office of Management and Budget (US OMB) 13, 103, 106, 108, 111–112, 114
- Olejniczak, Karol: cited in chapter 1 13, 16; co-author of Chapter 8 (Learning Agendas) 217–18; co-author of Chapter

- 14 (Academy of Evaluation) 189–200;
cited in chapter 15 217–219, 223, 228
- Olson, Johan P. 238
- operationalization of results (right-holder
engagement) 59–61
- outsourcing (of evaluation functions) *see*
commissioning (evaluation)
- Palenberg, Markus: cited in chapter 1 3,
13–14, 16; co-author of Chapter 10
(Finland EMS) 132–140; cited in chapter
15 210, 218, 219, 221, 222, 225
- Park, Yaerin cited in chapter 1 13, 16,
21; co-author of Chapter 8 (learning
Agendas) 102; cited in chapter 15 210,
218–219, 223
- participatory approaches to evaluation 6–7,
51–52
- Particip GmbH 133, 144
- Patton, Michael Quinn 29, 33
- Pawson, Ray 33
- PEMD (Program Evaluation Methodology
Division, US-GAO) 6
- performance management 116, 179–182,
184–186
- Perrin, Burt: cited in chapter 1 2, 4–5, 7,
9, 11, 19; co-author of Chapter 2 (‘How
to make commissioning matter’) 26, 28,
31, 33; cited in chapter 9 117; cited in
chapter 15 213, **214**, 217–221, 223
- Pew Research Center 61
- philanthropy and democracy 19, 63–64,
67, 70
- Philanthropy in Democratic Societies
(Reich et al., 2016) 19
- Poland: Learning Agendas in 106;
Academy of Evaluation in 189–201
- policymakers, commissioners’ relationship
to 2, 4–5, 10, 12
- Pollard, Alison: cited in chapters 1 and 15
12, 16, 223–224; author of Appendix
(Neo-Weberian State model) 232–239
- Pollitt, Christopher 232–232, 234, 236
- populist vs. public interest evaluation
(Datta) 6, 19
- portfolio management 147–148, 150, 155
- positive thinking (strength-based
evaluation) 33–34
- power asymmetries (race, gender, colonial
legacies) 18, 49
- Power, Michael 7, 19
- Preskill, Hallie 33
- Prime Minister’s Delivery Unit (UK) 100
- Prime Minister’s Office (Poland) 193
- principal investigator (role) 69
- public administration reform 219, 231–234,
241
- public administrations 12–13, 231–234,
236, 238–241
- public good 1, 10, 17, 49, 79–80
- public interest *see* populist vs. public
interest evaluation (Datta)
- Public Management Reform: A
Comparative Analysis* (Pollitt &
Bouckaert) 232–234, 236
- Public Procurement Law (Poland) 196
- Purposeful Program Theory* (Funnell &
Rogers) 129
- quality assurance (of evaluations) 31–32,
136–138
- quasi-market (evaluation services) 125
- Raimondo, Estelle 7, 12, 18–19, 184–185
- randomized controlled trials (RCTs) 10,
43–45, 146, 198, **214**, 220
- realist evaluation and synthesis 33, 44, 46
- reciprocity (Indigenous evaluation
principle) 42
- reflexive rationality 231, 239–240
- Research, Evaluation and Learning (REL)
team (RWJF) 70–71
- Results-Based Management (RBM)
147–148, 150–151, 237–238, 240
- right holders: cited elsewhere 10, 213,
214, 224; in commissioning with clarity
framework (Chapter 4) 50, 53, 56–61
- risk tolerance (of commissioners) 12, 80
- Robert Wood Johnson Foundation (RWJF)
15, 64, 67–72
- Rogers, Patricia: cited in chapter 1 9–11,
15, 34; co-author of Chapter 3 37–40, 51;
cited in chapter 15 209, 213, **214**, **215**,
218–221, 224
- Rome (INTEVAL meeting) 216
- rubric model (commissioning) 11, 46, 52,
58–61
- Sage Handbook of Evaluation* 6, 19
- Sagmo, Tove 146, 158
- Salesforce Foundation 85
- Schön, Donald 66, 73
- Schwandt, Thomas (A.): on civic science
/ democratic purpose of evaluation 1, 6,
163–164; on single, double and triple
loop learning 66–67; cited on evaluation

- and national values 186; cited in chapter 15 208, 210–211, 218–220, 223, 224, 228
- self-evaluating organization (Wildavsky) 5
- self-reflexivity 12, 71
- serial evaluations 29, 33
- Service Coordinator (EMS role) 133, 135, 138–141, 144
- single, double and triple loop learning 66–67, 71, 214
- Śliwowski, Pawel 190, 193, 196
- Soros Foundation (Open Society) 85
- South Africa 38, 40
- Spain: Master's Programme on Evaluation (Complutense University), competencies study in 162–172
- Spanish National Institute of Public Administration 162, 172
- stakeholder engagement: in commissioning process (Chapter 2) 26–32; as democratic evaluation element (Chapter 3) 38–40; in Learning Agendas design 108, 111; in market-based evaluation, Italy 125, 128–129; cited in chapter 15 216–221
- Stame, Nicoletta: as co-editor, in chapter 1 1–9, 17; co-author of Chapter 2 33–34; cited elsewhere 132, 139, 189; co-author of Chapter 15 207–228
- steering groups 126, 169, 199, 221
- Stepień, Marlenacited in chapter 1 9, 16; co-author of Chapter 14 (Academy of Evaluation) 189–90; cited in chapter 15 217–218, 228
- streams of evidence 102, 113
- structural racism 63
- Strzęboszewski, Piotr cited in chapter 1 9, 16; co-author of Chapter 14 (Academy of Evaluation) 189, 202; cited in chapter 15 217–218, 228
- Sustainable Development Goals (SDGs): cited in chapter 1 1, 6, 10–11, 15–16, 26; complexity of roll-out (Figure 7.1) 91–93; as Chapter 7 theme (Democracy, SDGs, evaluation and commissioners) 82–84, 89–90, 94–100; cited elsewhere 170, 220, 237
- Tanasoca, Ana 43
- technocratic professionalism 210–215, **214**, **215**
- Terms of Reference (ToRs): as commissioning document, defined 3–4, 10–11, 25–26; flexibility and bespoke design of 46, 56, 59; in Evaluation Management Services, Finland 135, 140–141; in Norad evaluation series 150–151, 156; cited elsewhere 169, 210, 213, 216, 220
- The Big Con* (Mazzucato & Collington, 2023) 128
- The Evaluation Enterprise* (Furubo & Stame, 2019) 16
- The Hague (INTEVAL meeting) 216
- theory-based evaluation 57, 123, 129
- theory of change: as evaluative tool, defined 4–5, 10, 57; in Learning Agendas (Figure 8.1) 102, 104–105, 108, 111; in Norad evaluation series 147–148, 150, 153; in competencies frameworks 166, 174; in Norwegian training course (Weiss's definition) 181–182; cited elsewhere 234
- “Think Again: The Power of Knowing What You Don't Know” (Grant, 2021) 238
- token evaluation 38
- Torfin, Jacob 233
- training (of commissioners and evaluators) *see* critical evaluative thinking
- transparency: as democratic evaluation principle 1, 9–12, 15, 37–39, 44–46; in market-based evaluation, Italy 116–118; cited in chapter 15 224, **226–227**; Neo-Weberian State model and 234–239, 241
- triple-loop learning *see* single, double and triple loop learning
- trust (commissioner–evaluator relationship): as democratic value 1, 8–9, 11–12, 19; in Finland and Norad case studies 139, 156; cited in chapter 15 223
- two communities theory 102
- UBUNTU Research & Evaluation, LLC 67–69
- Uganda 86
- UK Evaluation Society (UKES) 8, 19, 161, 163, 166
- UN Country Teams 97
- UNDP Oslo Governance Centre 61
- UNEG (United Nations Evaluation Group) 8, 19, 95, 97
- UNFPA (United Nations Population Fund) 85
- UNICEF 14, 85, 236

- United Kingdom 54, 96, 100, 235–236, 238
- United Nations: cited in chapter 1 11, 13–14; as commissioning sector 81, 84–85; SDGs and 92, 96–97
- United Nations Commission on Human Rights 238–239, 240
- United States 84, 96, 103, 127
- University of Warsaw 189–190, 202
- USAID 103
- US Commission on Evidence-Based Policymaking 103
- US Department of Labor 103
- US-GAO (General Accounting Office) 6
- utilization-focused evaluation 29, 33, 58
- value for money 32, 56–57
- values of inclusion *see* inclusion, values of
- Van de Walle, Steven 233
- vulnerable groups 58–60
- Warsaw 189, 192–193
- Washington, DC (INTEVAL meeting) 215–216
- Weber, Max 231–232, 238
- Wehipeihana's typology of Indigenous evaluation 39–40
- Weiss, Carol H. 181
- “what works” agenda 42–45
- Whynot, Jane 41
- wicked problems 179
- Wig, Tore 146, 156–158
- Wildavsky, Aaron 5, 19
- Winckler Andersen, Ole 179
- Woolcock, Michael 52
- World Bank 4, 92
- World Food Programme 85
- World Health Organization 92
- Żuber, Piotr 202



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>