

EVALUATION CONNECTIONS

President's message

María Bustelo, EES President

Dear members and colleagues,
This issue of Connections reflects the overarching Conference theme: “*Evaluation in the networked society: new concepts, new challenges, new solutions*”. It mirrors the breadth, wealth and quality of your contributions to the 10th EES Biennial Conference in Helsinki. The Conference attracted 634 delegates and featured 275 paper presentations, 52 panels/round tables and 21 posters. In a uniquely hospitable and distinctly Finnish venue, it provided a convenient platform for sharing ideas and experience. We were able to successfully connect evaluators and organisations who share a passion for our fledgling profession – within Europe and beyond.

More than ever, the EES Conference helped to forge links across borders and disciplinary boundaries. Thus it offered an opportunity for all European evaluation societies (including EES) to link up through the annual meeting of the Network of Evaluation Societies of Europe (NESE). The fulsome debates that the Conference triggered ranged widely across the five Conference strands in such diverse issue areas as evaluation theory, methods, practices, ethics and capabilities as well as in diverse regional and sector contexts. Thanks to the generous support of Finland and other donor countries 75 bursaries were offered to developing countries' evaluators.

Similarly the launch of Thematic Working Groups (TWGs) in the course of the Conference is emblematic of your Society's determination to reach out and involve all members. The ‘*Gender and Evaluation*’ TWG sponsored several Conference panels and the set up of three other TWGs was announced – ‘*Inter-*

national Engagement in Fragile States’, ‘*Evaluation Professionalization*’, ‘*Evaluating Sustainable Development*’. Since then, as highlighted below a fifth TWG on ‘*Private Sector Evaluation*’ has been approved by the EES board. We are looking forward to the creation of more TWGs and encourage you to come forward and propose a TWG within an evaluation domain that you wish to promote.

As other evaluation associations EES is seeking to promote evaluation in the civil society. This is a privileged goal of the ‘EvalPartners’ initiative. I therefore represented EES at the International EvalPartners Forum on Civil Society's Evaluation Capacities that took place in Chiang Mai (Thailand) on 3–6 December 2012. The event aimed to contribute to the enhancement of Civil Society Organizations (CSOs) evaluation capacities through Voluntary Organizations of Professional Evaluators (VOPEs). Its ultimate objective was to find ways to amplify the voice of citizens with a view to influence policy makers, other key stakeholders and public opinion so that policies and programmes are based on evidence, and incorporate considerations of effectiveness, social inclusion, human rights, gender equality and environmental sustainability.

Following the e-learning programme that started last September and was joined by more than 4,500 people, the EvalPartners Forum in Chiang Mai represents an important milestone in a process that will enable Voluntary Organizations for Professional Evaluation (VOPEs) to better promote evaluation. For the first time VOPEs, national, regional and global evaluation associations and networks as well as CSOs and international institutions

(UNICEF, UNWOMEN, UNDP, OECD and the World Bank) as well as bilateral donors (the Finnish Ministry of Foreign Affairs, the Swiss Agency for Development and Cooperation, USAID and AusAID) joined together to develop a three year strategy (2013–2015) towards declaring 2015 the *Year of Evaluation*.

A Chiang Mai Declaration on ‘*Civil Society Working in Partnership for Better Evaluation*’ was signed. Five Task Forces were set up. The EES board will consider how best to contribute to this international movement. If you wish to volunteer and take part in this initiative so as to strengthen our links with Evalpartners and the civil society, please do not hesitate to contact me through the EES Secretariat. Finally, let me wish you all the best for this new coming year during which I plan to work together with all of you for evaluation and a better world!

María Bustelo, EES President

December 2012



CONTENT

President's message	1
Editorial	2
TWG – Private Sector Evaluation	2
The Evaluation Society	3
Evaluation theory and practice	5
Complexity, systems thinking and evaluation	7
Judging Value	9
Development evaluation and “Big Data”	11
Evaluation credentialing	13
The authors	14

Editorial

Robert Picciotto and Claudine Voyadzis



Robert Picciotto



Claudine Voyadzis

Dear colleagues,

Many Helsinki Conference participants have expressed frustration at not being able to be in two (or three) places at once in order to attend panels, round tables or paper sessions that took place at the same time in parallel sessions. A special Evaluation publication (featuring a selection of papers submitted to the Conference organizers) as well as this and forthcoming Connections newsletters (drawing on Conference panel sessions) partially fill this ubiquity gap.

Dissemination of evaluation knowledge and promotion of principled debate about the principles, methods and practices of our discipline are core objectives of the European Evaluation Society. We are therefore delighted that participants in six Conference panels agreed to share the highlights of their Helsinki Conference contributions with Connections readers. With the benefit of further reflection they were able to cap-

ture many of the salient concerns evoked by the Conference theme – Evaluation in the Networked Society.

For example, Ken Chomitz views the advent of bigger, faster, cheaper and more customized data made possible by the new information and social networking technologies as a revolutionary trend that will increase transparency and accountability for results in society since “anyone with an internet connection and a computer will be able to evaluate anything – and they will”.

For Peter Dahler-Larsen, the evaluation process itself is inherently a democratic undertaking so that evaluation is “societing” while, for George Julnes, supporting valuation in the public interest (the overarching mandate of our discipline) is not a technical problem to be solved but a human one to be managed within a particular social context.

Similarly in a session chaired by Elliot Stern panelists contributed novel notions and far reaching perspectives about the potential of the emerging complexity sciences for the future of evaluation. They identified the network as the unit of analysis and argued that new systems thinking will support timely, diverse, purposeful and ethical innovations to evaluation practice.

The role of evolving evaluation theory in shaping our discipline is echoed by Stew-

art Donaldson. He considers that evaluation theory and practice are inextricably linked. Furthermore, theory gives us the language we use to communicate and it makes us different from other professions. Equally, the issue of professional identity is central to Martha McGuire’s clear sighted article about Canada’s promising credentialing initiative.

This newsletter also unveils the fifth Thematic Working Group (TWG) of the European Evaluation Society. TWGs connect evaluation commissioners, managers, practitioners and users committed to evaluation excellence and utilization within a particular issue area or sector. They are emblematic of the Society’s determination to deepen and broaden its reach; to demonstrate its practical relevance to members; and to promote collaboration and exchange of evaluation experience.

All EES board members are delighted that Fredrik Korfker will lead the work of the Society in the private sector as head of this latest addition to the TWG family. His exposure to diverse facets of private sector operations in a wide range of country contexts combined with his grasp of banking issues and his broad international perspective make him ideally placed to promote evaluation in the private sector at a time of unprecedented turmoil in the international economy.

TWG – PRIVATE SECTOR EVALUATION

Fredrik Korfker

Private sector entities play a central role in the economy and the society. They create wealth, provide employment and generate innovation. In partnership with government, they are increasingly active in infrastructure development and the delivery of public goods and services. Yet, with notable exceptions (aid funded private sector projects in developing countries¹; European structural funds; social and environmental impact assessments and regulatory framework reviews) private sector interventions are not routinely subjected to systematic evaluative scrutiny.

Since 2008 the global financial crisis has induced national audit offices, central banks and public policy think tanks and academics to carry out studies of banking sector governance and practices. The crisis has also elicited interest in learning from experience about what works and does not work in schemes designed to promote private sector infrastructure investment, youth employment, entrepreneurship and innovation. But these studies have not always involved all relevant disciplines. Nor have they always been scrutinized with sufficient evaluative rigor.

Thus evaluation of private sector initiatives are currently limited and scattered. They have yet to capture the evaluation mainstream. The TWG on Private Sector Evaluation would unite evaluators working in this field within and beyond Europe and promote interest in evaluation within the private sector, the auditing profession and academic think tanks. It is intended to serve as a learning platform for private sector methodologies, practices and processes. The new working group will examine what makes private sector evaluation different from evaluating

public sector activities, e.g. with respect to the evaluative implications of corporate ethics and business integrity principles; the interface between private sector auditing and evaluation; the soundness of corporate social responsibility reviews and social and environmental assessments, etc.

In short, the thematic working group will deal with the evaluation of private sector

activities from a public interest perspective. But it will also pay attention to the internal accountability and learning mechanisms in place within private corporations including their use of the new information technologies. Accordingly the TWG would help bring together evaluators, systems analysts and academics interested in assessing the merit, worth and value of private sector operations from a variety of perspectives.

¹ Both multilateral and bilateral aid organizations have acquired substantial experience in evaluation of development finance institutions, micro-credit initiatives, small and medium enterprise development, trade facilitation operations, etc. The Evaluation Cooperation Group (ECG) in which the heads of the evaluation departments in MDBs collaborate have developed good practice standards with respect to private sector evaluation.

Fredrik Korfker

(Dutch nationality) gained experience in private sector evaluation during 17 years at the European Bank for Reconstruction and Development (EBRD) in London where he was Chief Evaluator. Since his retirement early 2011 he did short term consultant assignments for the UNDP, the World Bank and the Belgian Government. Fredrik's background is in project finance/banking having worked for, the Dutch development finance company FMO, a Dutch subsidiary of the former Chase Manhattan Bank, a Dutch mortgage bank and the Inter-American Development Bank in Washington DC. He studied business administration at Nijenrode Business School in the Netherlands. After obtaining a masters degree in economics in 1972 from Erasmus University (former NEH) in Rotterdam, majoring in development planning, he worked five years for United Nations (FAO) stationed in several developing countries and at FAO headquarters in Rome.



THE EVALUATION SOCIETY

Peter Dahler-Larsen



Research on evaluation helps to enhance our practice. But there are many ways of researching evaluation. One of them is to explain common observations and experiences in our field through the theoretical lenses of other bodies of thought. Since evaluation is no longer an individual choice, but rather an organizational, institutional, social and political phenomenon, it is only fair to ask such questions as:

- Why do we have an evaluation “wave”?
- Why do organizations ask for evaluations?
- Why do organizations ask questions that cannot be answered by evaluators?
- Is “evaluation machines” a useful metaphor?
- Do evaluations have constitutive effects?
- What broad social changes shape evaluation?

I pose these questions because I view evaluation as a construction. It is constructed by means of expectations, activities, institutions, norms, interests etc., all of which

Thomas Schwandt calls an “*evaluation imaginary*.” At the same time, evaluation is a construction of practices, policies, “*improvements*”, reforms, etc. Finally, evaluation is a conceptual construct. Its meaning and content are not constant over time. In fact, evaluation has demonstrated an amazing capability to change over time and to adapt to both internal and external pressures, expectations, and critique.

To analyze more specifically how evaluation is both constructed and a construction, I see it as embedded in two great sources of order in modernity, called “organization” and “society”. So, I subscribe to organization theory and sociological theory in an attempt to answer the questions above. I let each of these theories unfold in a tripartite structure.

In organization theory, I look at rational organization, the learning organization, and the institutionalized organization. From sociology, I borrow theories about modernity, reflexive modernity, and the audit society. I describe and explain the various forms of evaluation that characterize each of the forms of organization and society, respectively. I also show that forms of evaluation which are useful, appropriate and meaningful in a particular era may well turn out to have severe limitations in another. For example, reflexive forms of evaluation that invite a variety of stakeholders to participate

may later be found to have severe limitations when judged by the standards of a subsequent era, such as the more management-oriented one we call “the audit society.” In fact, if we understand transitions from modernity to reflexive modernity to the audit society, then we understand a large part of the trends in the history of evaluation.

If on the other hand we understand evaluation – at any given time – as a reflection of broader social norms, values and assumptions that characterize a particular era, we are in a position to answer the questions above. For example, some of the demands for evaluation, for evaluation criteria, or for evaluation data, are highly institutionalized and not tightly connected to the agents who actually use evaluation. New actions in organizations are not a result of previous learning. At least this is an observation we would make if we apply the idea of the “institutionalized organization” to our analysis. Using that model, we better understand why learning from evaluation is sometimes limited even though organizational learning demands are high.

Another example: I propose that in contemporary audit society, we observe a rise in the “evaluation machine” phenomenon. Evaluation machines are mandatory, routine-based, non-subjective, and repetitive forms of evaluation that cover large grounds in terms of time, space and activity. They seem not

to depend very much on the judgment of individuals. Instead, they are akin to bureaucratic procedures based on manuals, indicators, guidelines, handbooks, and statistical paraphernalia. “Evaluation machines” help define social realities as well as describe them.

The advent of “evaluation machines” gives rise to important democratic questions. The political and the democratic concern is an ever-present aspect of evaluation. It is not a “dark side of life” which should be “expelled” from the “rational, well-planned and well-intended” noble art of evaluation. The democratic ambition is full of tensions and ambiguities. It is necessary to clarify them and develop them in different ways in different moments in the history of democratic nations and in different situations.

Exactly the same is true for key terms in evaluation such as “means”, “ends”, “effectiveness”, “stakeholders”, “indicators”, “quality”, “values”, “evaluands”, and “use of evaluation”, not to mention “evaluation” itself (!) which are all ambiguous structuring principles with major democratic relevance. However, the meaning of each of these terms is itself a democratic riddle rather than a fixed technological or methodological given. What evaluators do as they work with these terms in practice is therefore also inherently a democratic undertaking. Evaluation is “societing”.



EVALUATION THEORY AND PRACTICE

Stewart I. Donaldson

One of the topics I was determined to discuss with my European colleagues at the Helsinki Conference was the use of theory in evaluation practice. I was thrilled to stumble upon several thought provoking discussions regarding the use of theory in evaluation. A common theme was the belief that teaching evaluation practitioners about theory was critical to a better future for the evaluation discipline. But why is theory so important? The session I addressed on this topic also involved Frans Leeuw, Evert Vedung and Gary Henry. I emerged from the session with some new (and old) insights about evaluation theory and evaluation practice. This brief article summarizes what I learnt – and what I said.

Leeuw made it clear there are many inter-related uses of the word theory sprinkled throughout the evaluation literature. In fact, he described this vast diversity of the use of theory as a “jungle.” Imagine a newcomer to the field, or even a seasoned veteran trying to sort through the concepts of program theory, policy theory, systems theory, theories of change, theory-based evaluation, theory-driven evaluation, program theory-driven evaluation science, program process theory, program impact theory, intervening mechanisms theories, program logic, logic models, log frames, theories of policy change, policy process theory, social science theories, evaluation theories, evaluation models, evaluation approaches, evaluation forms, concept mapping, implementation theory, middle range theory, translational theory, theory weaving, theory knitting and the like.

In my presentation I emphasized the need for evaluators to seek clarity about the role that specific theories are playing in the context of a specific evaluation, and defined three of the most popular types used to improve evaluation practice – theories of change, social science theory (and research), and evaluation theory. Donaldson, Lipsey and Mark have provided detailed accounts of how to optimize the use of these three types of theory in contemporary evaluation practice.



The third leg of this stool (evaluation theory) emerged as the main topic of conversation throughout the presentations, the panel discussion and the engagement with the audience. For Marvin Alkin evaluation theories are largely prescriptive and “offer a set of rules, prescriptions, prohibitions, and guiding frameworks that specify what a good or proper evaluation is and how evaluation should be done”. My presentation emphasized the need to better inform practicing evaluators about the latest developments in evaluation theory despite the common misunderstanding that theory is not practical or relevant to the lives of practitioners.

In addition to referencing my own work on this topic I encouraged the audience to contemplate why Shadish vigorously asserted evaluation theory is central to our professional identity and why he urged all evaluators to learn about evaluation theory. He claimed that this is what we talk about more than anything else. For him there is little doubt that evaluation theory gives rise to our most trenchant debates. It gives us the language we use for talking to each other, and perhaps most important, it is what makes us different from other professions.

He claims every profession needs a unique knowledge base. For the discipline and profession of evaluation, evaluation theory is that knowledge base.

The good news for practicing evaluators is there are now useful frameworks and categorizations systems to help guide the development of a sound evaluation theory background (see bibliography below). For example, Shadish, Cook, & Leviton provided one of the first frameworks showing how evaluation theory developed through stages over time. Donaldson & Scriven attempted to update and expand upon this early work by having a diverse group of evaluation theorists articulate their visions for the future of evaluation practice. Alkin published a second volume of his book “Evaluation Roots,” which offers a theory tree metaphor for organizing and understanding the similarities and differences between evaluation theories¹. Finally, Mertens & Wilson have recently offered us a more inclusive evaluation theory tree which adds many more theorists and a new branch (social justice).

Despite the advantages of these frameworks for helping practitioners better understand

the links between theory and high quality evaluation, Henry warned us in our session that most of this work is prescriptive and wanting of an empirical basis. He emphasized the need for a better metaphor for representing evaluation theory, and offered an analysis that suggested evaluation theory would be better represented as “rudderless” instead of rooted. He advocated more research on evaluation theory. I am encouraged by some of the recent work that has been done along these lines to develop criteria for evaluating theories of evaluation practice and the actual systematic evaluations of empowerment evaluation and theory-driven evaluation in practice.

Having been energized by the session I have agreed to contribute to a new article on how to improve evaluation theories with Frans Leeuw and Gary Henry. We aim to capture many of the insights gleaned from our panel presentation and the stimulating questions and comments from the audience during the session, as well as during the hallway conversations at Finlandia Hall and in follow up emails. Stay tuned.

References

- Alkin, M. C. (Ed.) (2013). *Evaluation roots: A wider perspective of theorists' views and influences*. Thousand Oaks: Sage.
- Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schroter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, 32(2), 199–226.
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. New York: Psychology Press.
- Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.) (2008). *What counts as credible evidence in applied research and evaluation practice?* Newbury Park, CA: Sage.
- Donaldson, S. I., & Crano, W. C. (2011). Theory-driven evaluation science and applied social psychology: Exploring the intersection. In M. M. Mark, S. I. Donaldson, & B. Campbell (Eds.), *Social psychology and evaluation* (pp. 141–161). New York: Guilford.
- Donaldson, S. I., & Lipsey, M. W. (2006). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In I. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The Handbook of Evaluation: Policies, Programs, and Practices* (pp. 56–75). London: Sage.
- Donaldson, S. I. & Scriven, M. (Eds.) (2003). *Evaluating social programs and problems: Visions for the new millennium*. Mahwah, NJ: Erlbaum.
- Mark, M. M., Donaldson, S. I., & Campbell, B. (Eds.) (2011a). *Social psychology and evaluation*. New York: Guilford.
- Mark, M. M., Donaldson, S. I., & Campbell, B. (2011b). The past, the present, and possible futures for social psychology and evaluation. In M. M. Mark, S. I. Donaldson, & B. Campbell (Eds.), *Social psychology and evaluation* (pp. 4–28). New York: Guilford.
- Mark, M. M., Donaldson, S. I., & Campbell, B. (2011c). Building a better future. In M. M. Mark, S. I. Donaldson, & B. Campbell (Eds.), *Social psychology and evaluation* (373–385). New York: Guilford.
- Mertens, D. M., & Wilson, A. T. (2012). *Program evaluation theory and practice: A comprehensive guide*. New York: Guilford.
- Miller, R. L. (2010). Developing standards for empirical examination of evaluation theory. *American Journal of Evaluation*, 31(3), 390–399.
- Miller, R. L., & Campbell, R. (2006). Taking stock of empowerment evaluation: An empirical review. *American Journal of Evaluation*, 27(3), 296–319.
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1), 1–19.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
- Stame, N. (2013). A European evaluation theory tree. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influences* (pp. 355–370). Thousand Oaks: Sage.

¹ One notable new contribution in the second volume is a chapter by Nicolette Stame offering a European Evaluation Theory Tree.



COMPLEXITY, SYSTEMS THINKING AND EVALUATION

Martin Reynolds, Kim Forss, Richard Hummelbrunner, Mita Marra, and Burt Perrin

“How is it that ideas that variously sail under the flag of ‘complexity’ – or ‘holism’, ‘non-linearity’, ‘systems theory’ and ‘cybernetics’ – have come to the fore? Implicit in this question is the acknowledgement that in the wider world of ideas, complexity as a focus is not that new”.

Elliot Stern

Starting point

‘Complexity’, ‘systems thinking’, ‘evaluation’ are all good words with positive connotations. At the 2012 EES Helsinki Conference a panel discussion explored the relationship between these concepts. This article sums up the discussion, and outlines an agenda for further research and debate.

A recent publication proposed a framework for evaluating complex policies and complex situations¹ grounded in the interaction of four elements – simplicity, inventiveness, flexibility, and specificity. In a foreword, Elliott Stern commented on the legacy of a systems approach based on the work of influential thinkers (Ludwig von Bertalanffy, James Miller, Fred Emery, Eric Trist, Stafford Beer, Russ Ackoff and Peter Checkland) who challenged prior reductionist, linear approaches some four decades ago.

Why despite their cogent challenges do such approaches still pervade contemporary management science and evaluation practices? What other expressions of systems thinking may assist methodological development in evaluating complex interventions?

Definitions

We normatively claim that *evaluation* involves *systemically exploring* the worth, the merit and the value of an object or intervention, acknowledging that such judgements are subject to change during the intervention and the evaluation.

Complexity as commonly understood is shorthand for everything that is messy, difficult to grasp, not easily understood, etc. Complexity deals with complicatedness of interconnections amongst entities. ‘Complexity science’ deals with phenomena where inter connections between elements

are non-linear and the emergent properties from interconnections are subject to uncertainty. Complexity can also be as a function of different perspectives on the complicatedness of situations.

Systems thinking originates from three traditions – the philosophic pursuit of ‘getting the bigger picture’ (holistic thinking), the practical pursuit of engaging with multiple perspectives each circumscribed with bounded judgements (pluralistic thinking and participatory practice), and the purposeful pursuit of improving situations (operational research and action research).

How do these relate to each other? The influence diagram below seeks to show how these different paths of thinking relate to each other (thickness of lines signal relative perceived strengths of existing influences)

Paths of convergence

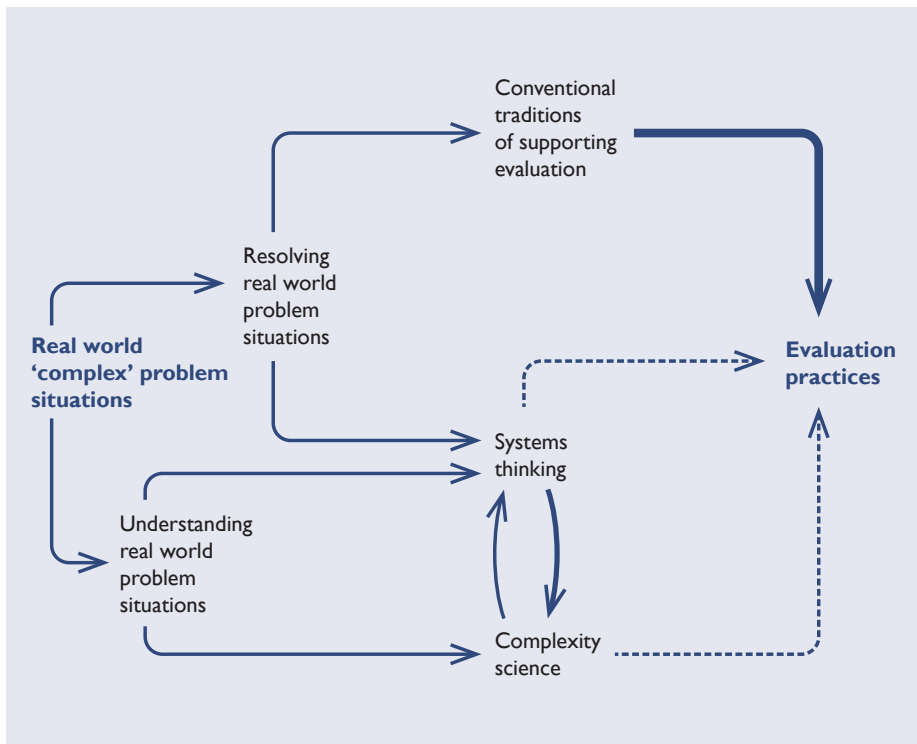
“Not all models are good and useful, but there is nothing as concrete and practical as a good model when evaluating complex policies”
(Forss and Schwartz)

Complexity science and systems thinking share a mutual concern for:

- Challenging narrow-minded reductionist practices, ‘rational’ models of policy-making, linearly assumed causal relationships and experimental evaluation designs which can often inhibit more appropriate or meaningful evaluation.
- Encouraging a dynamic, more holistic – whilst situated – perspective which influences the ability of evaluators to manage deliberative processes about complex problems in a democratic fashion.
- Promoting pluralism and inclusivity in the sense of (i) incorporating contrasting perspectives echoing fourth-generation

approach and empowerment approaches to evaluation and (ii) maximising the use of multiple methods.

- Highlighting the need to address emergence and systems change, drawing upon generative causality open to improvisation and unpredictable outcomes.
- Stressing co-evolution for evaluation practice and for practices being evaluated thereby identifying key issues in the changing landscape and shifting actors’ attention to new issues as they arise.
- Emphasizing the value question and seeking answers to what constitutes ‘worth and merit’ in the project, programme or policy being studied/evaluated.
- Elevating concern with processes; particularly those facilitating meaningful conversation, as in participatory evaluation.
- Drawing upon the theme of conversation as ‘evaluation with action’ while seeking to create “collective intelligence” and innovative interactions.
- Emphasising constructivist perspectives that recognize that evaluation is shaped through practice and through the interaction between organizational/institutional members and processes.
- Being mindful of interrelationships and purpose in conceiving networks as unit of analysis, thus implying a method of inquiry that involves interpretive reframing geared to serving the common good.
- Balancing the “safety” of exploiting well-accepted meanings associated with consolidated practices with the “dangers” of exploring radical interpretations and unfamiliar perspectives.
- Distributing and generating new information and novel action in contrast to the ‘performance management’ and ‘quality’ movements that are preoccupied with controlling deviance.
- Embracing *praxis* – the continual reflection on theories-in-use as against a singular focus on espoused theory – and the learning



A question orientation encourages the use of multiple approaches, including systems methods – or even elements thereof – alongside ‘traditional’ methods.

- Methods associated with the systems field provide distinctive contributions to evaluation, e.g. models for dynamic inter-relationships (e.g. Causal Loop Diagrams, Stock Flow Diagrams, Agent Based modelling or Social Network Analysis) as well as methods for clarifying perspectives (e.g. Soft Systems Methodology) or those that work constructively with diverging opinions (e.g. Circular Dialogue, Contradiction Analysis) to improve understanding, achieve consensus or create new insights.
- A *critical* systems thinking tradition of concern for boundary judgements invites the notion of designing purposeful engagement but also supports an *explicitly* ethical and political engagement with evaluation. This also involves a critical analysis and reflection on the power issues that are often associated with boundaries, which determine what is relevant, legitimate or ignored.

Future pathways

- Treated as a literacy rather than a bounded discipline (as implied through complexity science), are there opportunities for using systems and its rich and evolving heritage of language as a conversation around evaluation issues? Considered this way, systems thinking is not confined to a particular niche (‘systemic evaluation’). Instead, it is valid for and applicable to many other evaluation approaches.
- Systems thinking can help contribute towards a shift from ‘attribution’ (purposive systems) to ‘contribution’ (purposeful systems design) and ‘beyond’ (critical thinking on the ethics and politics of evaluation).
- Treating systems as heuristic (learning) devices means tapping opportunities for complementarity with other evaluation traditions – in particular, theories of change (including programme theory), developmental evaluation, and (critical) realist evaluation.
- Methods, in this view, are semi-structured, providing semi-coherent guides within which room for change is allowed as policy goals, instruments, and relevant indicators change in relation to the evolving features of the environment.

acquired through the practical application of conceptual tools.

- Applying three generic concepts common to all areas of the systems and complexity field: (i) exploring interrelationships, (ii) engaging with multiple perspectives, and (iii) reflecting on boundary judgements.

Paths of divergence

“The complexity resides in the evaluated object and how the evaluation handles that process is a step toward understanding, recognising, and making sense out of the complexity ‘out there’”
Forss and Schwartz

- The complexity perspective sees the world as structured, differentiated, and changing. It is structured in layers, from observable and physical features that can be mapped, to increasingly abstract layers of structures and mechanisms. Public policies and programs can be mapped and observed at a physical level and measured but only explained and understood at levels where hidden social, political, and economic structures are uncovered and analysed.
- In the *hard* systems tradition of thinking about systems, the emphasis is placed on the importance of complexity science in revealing the interconnectedness of entities being evaluated. In contrast,

contemporary soft and critical systems thinking approaches regard complexity more as a function of the observer (2nd order cybernetics) rather than the observed (1st order cybernetics).

- The assumption that interventions (projects, programmes, policies) exist as real world systems with independent purposes emphasises systems as purposive systems. The basic task here is to reveal the interconnected workings of such systems in order to ascertain whether the system “works”. Contemporary systems thinking regards systems not as actual real world entities but rather as conceptual devices in order to learn and transform reality. The shift here is from ‘purposive’ to ‘purposeful’ evaluation.
- Systems thinking evokes a literacy of craft skills rather than a bounded discipline. It encompasses a set of evolving abilities to understand and use symbols through language and diagramming for purposeful development. Ideas from complexity science/theory are only part of a wider set of craft skills associated with systems thinking.
- Systems skills and associated methods can be examined in relation to their relevance to evaluation questions. This is the route that evaluators should take in selecting and/or adapting appropriate methods.

- Methods become mechanisms for coordinating actors and organizations/institutions in complex settings and communities. Evaluation tasks thus become more than things to be done but also opportunities to build constructive relationships for collaborative undertakings.
- Embellishing complexity approaches with a literacy framework associated with systems thinking in practice would build on (i) understanding interrelationships (being concrete and specific), (ii) engaging qualitatively with multiple perspectives (be flexible), and (iii) reflecting on boundary judgements (be inventive).

The use of evaluation relies on the connections across actors, when individuals interact and exchange findings, data, opinions, and suggestions for future action. Complexity science urges evaluators to detect and emphasize these enabling relationships and structures emerging between ‘system’ participants. Evaluation should then concentrate on understanding the capacities for interactions to promote improvement, as also evaluators establish both formal

and informal communication, spaces for communication, and rewards for enhancing interaction. As interactions between participants are mostly voluntary, evaluators can create open architectures, forums, focus groups, or electronic work groups. The information and the knowledge thus generated are dynamic qualities that form the basis for improved future interactions and policy interventions.

These ideas although beginning to feature in evaluation thinking are still only partly understood and operationalised. Evaluators by and large do appreciate that increasingly complex situations associated with major public policy challenges (such as combating poverty or addressing climate change) require approaches that recognize and take into account inherent complexities. Yet this message is for the most part not getting through to policymakers. Linear models (such as the logical framework) still dominate. They pretend that an intervention can work, and its impact can be assessed, without taking into account context and multiple interacting activities and variables. The same

applies to many applications of the so-called “gold standard” randomised controlled trial. There is still much too much naive faith that one can assess the value of a complex initiative – and reward, or punish, performance – with just a small number of quantitative indicators, despite considerable evidence to the contrary.

What is the solution to this dilemma? The panel debate at the EES conference implied that there is no simple answer. But it does seem apparent that evaluators cannot discuss these important concepts just amongst themselves. Evaluators need to make greater efforts to reach out to other constituencies, identifying and working together with influential allies who can help create greater awareness and value of the insights from systems thinking and complexity science. As policies and programmes grapple with increasingly innovative policy instruments as well as increasingly sophisticated citizen demands for participation and accountability, systems thinking and complexity concepts could well provide some of the vital responses that evaluators will need in the future.

¹ Kim Forss, Mita Marra and Robert Schwartz Ed., *Evaluating the complex: attribution, contribution and beyond* (Volume 18; ‘Comparative Policy Evaluation’ series, Transaction Publishers, 2011)

JUDGING VALUE

Martin de Alteriis, Thomas Schwandt, Peter Dahler-Larsen, & George Julnes

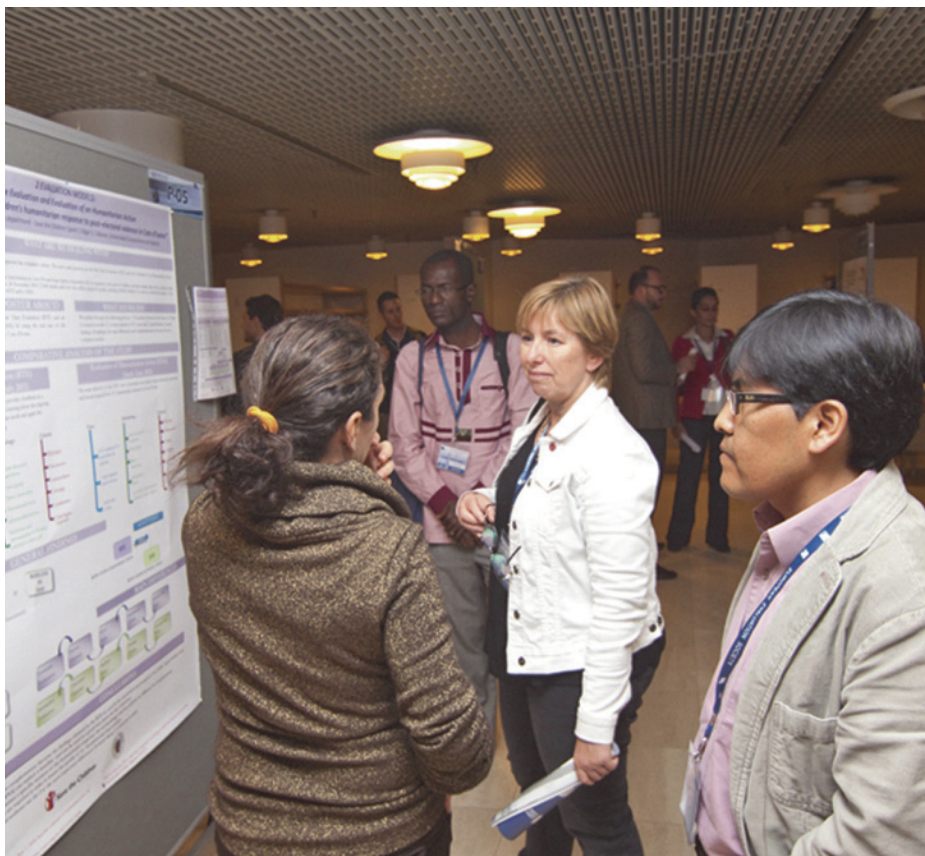
How does evaluation fulfill its central mandate – ascertaining the value of public interventions? A Helsinki Conference panel brought together distinct perspectives on valuing. Martin de Alteriis kicked off the dialogue by summarizing the approach to valuing practiced by the United States’ Government Accountability Office (GAO), an independent agency that performs audits and evaluations of United States federal programs. The GAO issues more than 1,000 products a year that evaluate a broad range of federally-funded activities. Ethical principles such as integrity, objectivity, and the public interest are essential to GAO’s operations. The concept of be-

ing ‘policy neutral’ is also critical, and is understood to mean that GAO does not make policy, but rather conducts research to help inform policymakers in their deliberations. GAO has developed elaborate processes and procedures to ensure adherence to them.

GAO stresses the need for sufficient and appropriate criteria in order to make valuations. Commonly-used sources of criteria include legislation, agency regulations, expert opinions, good practices, and social science and economic principles. GAO recently developed internal guidance on selecting and using criteria that encouraged its analysts

to consider very carefully the context of their evaluation, and determine which of the available sources were reasonable, relevant and attainable. In addition, staff guidance notes the importance of thoughtfully operationalizing criteria for each evaluation in light of the research questions, and determining the most appropriate metrics to apply.

De Alteriis concluded by noting that GAO also relies on the concept of “professional judgment,” which recognizes that absolute assurance is impossible in valuing, but holds that evaluators acting in good faith and with integrity can apply collective knowledge,



skills and experiences to arrive at findings that serve the purposes of public accountability and the public interest.

Thomas Schwandt rejected the notion that evaluation is simply a matter of selecting tools and matching them to appropriate contexts and questions. Instead, evaluation is a political project, and the value judgments evaluators render have to be considered within the context of political and social relations. Schwandt cited several examples to illustrate how evaluation can serve as a social-political practice by helping to frame social values and the public interest. One was the Cambridge Accountability Project, where evaluation was undertaken to contest a prevailing set of values, namely the 'control model' of school accountability. Another was drawn from Eleanor Chelmsky's observations about the increasing numbers of government programs and policies that posit a simple cause and effect relationship, or single narrative, which she viewed as threatening to direct and control the evaluation process.

Schwandt used these examples to argue that strategies for valuing in the public interest

should engage in value-critiques of social policy. He characterized the current climate as one in which policy making relies almost exclusively on empirical evidence of what policies work at the expense of engaging in debate about value commitments, and urged the evaluation community to speak out on that issue.

Peter Dahler-Larsen pointed to the lack of consensus on valuing in the evaluation field, and suggested that evaluators consider the ways in which values are chosen, formed and shaped. He applied a tripartite distinction introduced by the Danish sociologist, Nils Mortensen, who had proposed that norms and values can be established by reflexivity, facticity, and conventionality. In reflexivity, which is viewed as the most desirable approach, some form of explicit or external social practice, such as debate or an explicit algorithm for decision-making is employed, and norms or values are established after learning has taken place. However, in many instances, norms are based on facticity, which involves no discussion or debate about the reality of value choices; rather they are based on a reality that is already taken for granted. Many norms are also based on con-

ventionality, which is a form of normativity that has traditionally concerned sociologists. Similarly, under conventionality, norms and values are institutionalized or again simply taken for granted.

Dahler-Larsen noted that reflexivity in evaluation occurs when evaluators are explicit in how values are applied in valuing; in his view, evaluators that attempt to establish values through reflectivity deserve respect. However, evaluators often rely on facticity and conventionality. Facticity can occur when the evaluator treats value choices as given facts. Conventionality can occur when evaluation takes place in organizational and managerial contexts that requires some order of control and predictability. In addition, Dahler-Larsen pointed to the paradox of reflexivity, namely the risk that a reflexive procedure could become routine or ritual over time.

Dahler-Larsen ended by noting that evaluators no longer typically practice as individuals, but rather as instruments of large organizations that operate on the basis of facticity and conventionality. Consequently, he urged the audience to be aware of the risks of facticity and conventionality, and the paradox of reflexivity whenever value discussions occur.

George Julnes began the final presentation by stating that there are increasing pressures for evidence-based decision making in general, and that financial interests are promoting economic valuation as the only valid method. Not only does this provide a narrow view of the "public interest," it also is vulnerable to mechanical processes of valuing that ignore context. Yet, supporting valuation in the public interest is not a technical problem to be solved, but a human one to be managed. This suggests trying to make sense of our multiple paradigms of valuing, considering both (1) research on our multiple cognitive capacities for what can be called analytic versus holistic valuing and (2) experience on how different approaches to valuing might be more useful in different contexts.

Such considerations imply that we cannot expect to develop an adequate prescriptive model to align valuing methods with contexts, but it seems likely that some patterns of applicability can be salvaged. For example, evaluation can have a variety of purposes,

such as oversight and accountability, program and organizational improvement, and the assessment of merit and worth, and so we might study whether the complexity of valuing increases as we move through this list of purposes. Similarly, there might be contexts in individual valuing (typical in benefit-cost analysis wherein individuals working

alone complete contingent valuation surveys) and collective valuing (people discussing valuation issues together in groups) might be differentially appropriate.

Julnes concluded by noting that effective valuing makes use of multiple capacities, which vary somewhat by context. Different approaches

to valuing (analytic versus holistic; individual versus collective) used in parallel can correct for their individual limitations. Hence we need frameworks that can help us promote this critical corrective process without falling into the trap of believing we understand context enough to prescribe methodologies in any formal and inflexible way.

DEVELOPMENT EVALUATION AND “BIG DATA”

Kenneth M Chomitz

Development evaluation has long been starved of timely, pertinent, and comprehensive data. In large part this is because those who undertake development projects and programs have viewed monitoring as difficult, expensive, and extraneous to their immediate needs – often justifiably.

Today’s situation: data-starved evaluation

Because monitoring is seen as onerous, evaluation is infrequent, with feedback loops that are frustratingly slow for some purposes and premature for others. At the World Bank and the International Finance Corporation, independent evaluation typically takes place five to eight years after project initiation. For many kinds of innovative projects, this comes too late to accelerate the diffusion of good ideas or head off the replication of fad-dish bad ones. But for large infrastructure projects and major institutional reforms, it comes too early to assess impacts.

When data finally arrives, it is typically thin and inadequate. The World Bank, to its credit, has been bolstering projects’ use of quantitative results indicators and trying to ensure that baseline data are gathered at inception. Still, it remains unusual for projects to monitor comparison or reference groups, hobbling attempts to assess project impacts against a counterfactual. Indicators are often input- or output- based rather than outcome- or results- oriented. And it has been difficult to garner qualitative feedback from the tens or hundreds of thousands of people who are the project’s intended beneficiaries. Finally, monitoring and evalua-



tion has sometimes been viewed by project proponents as a burden – something that serves the evaluators but doesn’t help with implementation.

The advent of bigger, faster, and more customized data

But change is upon us. The cost of gathering and interpreting data continues to decline exponentially, triggering an explosion of information. The triple revolution of remote sensing, GPS, and easy-to-use GIS (geographic information system) software makes it possible to track people’s influence on environment (and vice versa) in real time, and to control for factors such as climate, road

access, and soil quality in assessing project and program impacts. The widespread adoption of cellphones, even by very poor and remote people, gives voice to the formerly voiceless. Digital service records of health systems, utilities, and other agencies allow assessments of program impacts on organizational effectiveness and household or individual outcomes – analyses that would have been slow, costly, or error prone in the days when records were kept on paper, or not at all. Finally, growing access to the internet means that people’s “data exhaust” – for instance, the commuting routes revealed by their cell phone movements, or their views on the economy expressed via Twitter – become grist for evaluation.

Early adopters of new approaches

Encouragingly, development practitioners are beginning to experiment with these new data streams, priming a pipeline that will flow eventually to evaluation. Some project leaders recognize that real-time feedback can be an integral part of project management. A noteworthy example is the Sujala project in Karnataka, India. This \$100 million watershed management project included a sophisticated monitoring and evaluation system, incorporating case/control household surveys and remote sensing. Feedback during implementation was instrumental in refocusing project benefits towards women and the landless. Project evaluation documented a 24% increase in average household income, together with environmental gains, spurring scale-up and replication.

At the cutting edge, task managers at the Bank are experimenting with the use of statistically representative mobile-phone based surveys in order to rapidly collect comprehensive information that would be prohibitively expensive to obtain with traditional methods. Projects in Africa are soliciting public input into regular project supervision reports. And the Bank's Research Group is pioneering in the development and deployment of computer-aided survey techniques, to allow statistical agencies to gather and process survey data faster and with fewer errors.

These approaches are beginning to find their way into evaluation practice. IEG has

used a facebook presence to solicit input on its forestry evaluation. It has used global remote sensing data to assess, in a statistically controlled way, the impact of protected areas on tropical deforestation. And it is making available its deep archives of ratings of World Bank projects.

Immediate challenges

As big, fast data becomes more prevalent, evaluators will face two big challenges. First, even with improved software, handling and analyzing the data will require evaluators to invest in special skills and capabilities. For example, the deforestation analysis mentioned above required processing a trillion-point database. Second, the flood of high-volume but non-representative data will require careful handling. There are many valid uses of this data which do not require that it is statistically representative. However, there will be strong temptations to treat it as if it were. However, sheer volume of data cannot compensate for the bias introduced by the difference between tweeters and non-tweeters, or mobile phone users vs. nonusers. There is a great need to develop bias-correction methodologies, and to identify robust data uses which do not rely on statistical representativeness.

Long-run challenges for evaluation in the Big Data era

Three trends – increased transparency, the results agenda, and Big Data – will lead

to the democratization of evaluation. As governments and development institutions become increasingly transparent and results-oriented, evaluable data streams will swell. Anyone with an internet connection and a computer will be able to evaluate anything – and they will. This will spur both a profusion of high quality creative analyses (as for instance the Netflix Prize and Kaggle have elicited improved modeling of problems) and dubious work (especially that based on misuse of non-statistically representative data). Building a system of peer review would be one way of steering the emerging system towards reliable evaluation.

A corollary is that official evaluation units, such as the multilateral development banks' independent evaluation units, will need to evolve. They have enjoyed a privileged status in part due to monopoly access to data, but that monopoly is eroding rapidly. Their future may involve ensuring the integrity of the monitoring and evaluation ecosystem. This could involve evaluating or auditing the systems that generate monitoring data, a task that will become increasingly important if development assistance shifts towards output-based aid. Another role could be undertaking authoritative meta-analyses of evaluations done by others.

The forthcoming flood of data promises to transform the world of evaluation. With the right incentives in place, faster, more detailed and more comprehensive evaluation could help to accelerate the development process.

EVALUATION CREDENTIALING

Martha McGuire



After two and a half years in existence, the Canadian Evaluation Society's (CES) Professional Designation Programme (PDP) has awarded the Credentialed Evaluator certification to almost 190 evaluators. Despite continued resistance in many quarters about the value and inclusiveness of credentialing, there is growing world-wide interest in the Canadian experiment. Some of the questions being posed regarding CES' PDP are briefly discussed in this article.

What is PDP?

The CES Credentialed Evaluator (CE) designation is designed to define, recognize and promote the practice of ethical, high quality and competent evaluation in Canada through a program for professional designations.

What is needed to qualify?

Applicants must provide evidence of graduate level degree or certificate. Prior learning assessment is also available. They must also demonstrate at least two years (full-time equivalent) of evaluation-related work ex-

perience within the last ten years as well as education and/or experience related to 70% of the competencies in each of the five agreed domains of competencies associated with Canadian Evaluation Practice: reflective, technical, situational, management, and interpersonal.

Two members of a credentialing board are tasked with reviewing each application to assess whether the applicant meets the requirements. If a person is deemed not to meet the requirements, she/he is advised as to what further learning is required. An individual has three years from the time of registration to provide additional information.

What is needed to maintain the CE Designation?

Each Credentialed Evaluator is required to complete a minimum of forty hours of professional development over a three year period. Evidence of this is submitted through an online process. The emphasis of the Professional Designation Programme is on learning and professional development. Establishing

the PDP has placed greater pressure on CES to provide training and professional development opportunities that can be used towards achieving or maintaining the credential.

What are some of the benefits?

There are benefits at three levels: for CES, for those engaging evaluators and for the evaluators. For CES, it has brought a higher profile and world-wide recognition. The process has also raised consciousness about professionalization among CES members. For commissioners who engage evaluators it provides comfort. It is a way of ascertaining whether an individual has the knowledge, experience, skills and dispositions that support the delivery of high quality professional services. For evaluators, credentialing provides a diagnostic of professional strengths and weaknesses and a clear articulation of the competencies needed to aspire to evaluation excellence. It gives new evaluators something to aspire to and bestows credibility to those who have been able to secure the CE designation.

Can evaluator outside of Canada obtain the designation?

Yes, anyone can apply so long as they are a member of CES. It should be noted that CES has many members who are from outside of Canada.

What is involved in establishing a PDP?

This process does not emerge overnight. In Canada, it started over 10 years ago and involved extensive research and consultation. The steps towards implementation included reaching broad based agreement on relevant core competencies, gaining public support for the program, developing a process for assessing whether an applicant has the necessary competencies including an on-line application process, setting up a credentialing board, an appeals process, and a system for maintaining the credential.

There also needs to be a mechanism for collecting the fees associated with the application and maintenance process. CES set fee to cover the actual cost. The current fees are Cnd\$485 for the initial application and Cnd\$50 per year to maintain the credential.

In carrying out the consultation process needed to achieve legitimacy, it was important to be inclusive and demonstrate openness to divergent views. Much of the public discussion led to improvements in the planning and implementation of the program. Some of the challenges that CES experienced were

reaching a national consensus and overcoming technical glitches in the early stages. There continue to be questions about the cost relative to the value of credentialing. On the other hand, there has been tremendous support from those who have been involved in the process—only one person actually withdrew from CES as an expression of disapproval. The CES membership continues to challenge us to make the program better.

There were numerous people who contributed to the process - too many to be named. However, Keiko Kuji-Shikatani

merits particular attention. She is currently the Vice-President of the Professional Designation Programme, with responsibility for implementation. She was a major driver throughout the process.

A rich literature about professional designation is available. But in order to make the process work for evaluators it must be stressed that the gestation process requires extensive volunteer time, inclusive consultation that is not rushed and a willingness to ensure that the tested competencies are meaningful and adapted to the context.

THE AUTHORS

Martin de Alteriis

Martin de Alteriis, Assistant Director in the Applied Research and Methods Team at the United States Government Accountability Office (GAO), has conducted research on a wide range of U.S. federal policies and programs during his 20 years with GAO. He currently focuses on international affairs and trade issues. He holds a Ph.D. from the State University of New York, and a M.Sc. (Econ) from the London School of Economics.

Kenneth Chomitz

Kenneth Chomitz, Senior Advisor in the Independent Evaluation Group of the World Bank Group (IEG) holds a degree in mathematics from MIT and a PhD in Economics from the University of California, Irvine. Previously he was Lead Economist with the Bank's Development Research Group. He is the author of *At Loggerheads? Agricultural Expansion, Poverty Reduction and Environment in the Tropical Forests*.

Peter Dahler-Larsen

Peter Dahler-Larsen, PhD, professor of evaluation at the Department of Political Science and Public Management, University of Southern Denmark, coordinates its Master Program in Evaluation. He was President of European Evaluation Society in 2006-07. His

most recent publication is *"The Evaluation Society"* (Stanford University Press, 2012).

Stewart Donaldson

Stewart Donaldson, Dean, Professor, & Director of the Claremont Evaluation Centre at Claremont Graduate University (USA) leads one of largest university-based evaluation training programs in the world. He serves on the Boards of the *American Evaluation Association*, *American Journal of Evaluation*, *New Directions for Evaluation*, and *Evaluation and Program Planning*. His most recent book is *Emerging Practices in International Development Evaluation* (2013).

Kim Forss

Kim Forss works as an independent evaluation consultant based in Sweden. He has co-edited the recently published book *'Evaluating the Complex'*. He has been President of the Swedish Evaluation Society and is a Board Member of the European Evaluation Society.

Richard Hummelbrunner

Richard Hummelbrunner, senior associate of ÖAR Regionalberatung Graz, Austria, has more than 30 years of experience as a consultant/evaluator in the fields of regional and international development. He has been active in promoting the use of systems thinking

in evaluation as a practitioner, trainer and author.

George Julnes

George Julnes is Professor of Public and International Affairs at the University of Baltimore (USA). He is currently a Board Member of the American Evaluation Association. His doctoral training was in clinical psychology at the University of Hawaii, with subsequent graduate degrees in public policy and business at the University of Michigan.

Mita Marra

Mita Marra is Assistant Professor of Public Policy and Public Administration at University of Salerno, Italy. She has consulted with the World Bank and the United Nations for the evaluation of public sector reform programs and policies for gender equality.

Martha McGuire

Martha McGuire is a Credentialed Evaluator with almost 25 years of experience in evaluation. She established an evaluation consulting firm in 2001. She was actively involved in the development of the Professional Designation Program with the Canadian Evaluation Society and continues her involvement as Past President of CES and as a member of the IOCE board.

Burt Perrin

Burt Perrin is an independent consultant based in France. He has assisted governments and other organisations internationally with quality assurance, evaluation management and planning, and related services in evaluation and strategic planning.

Martin Reynolds

Martin Reynolds, a senior lecturer at the Open University (UK) teaches and has

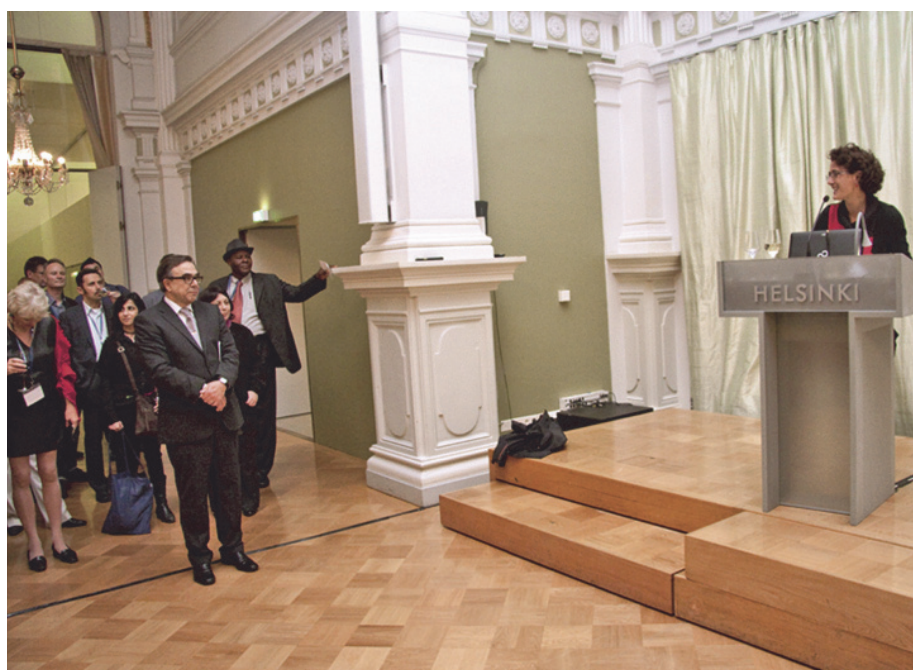
published widely in fields ranging from environmental management to international development. He provides professional development workshop facilitation for critical systems thinking.

Thomas Schwandt

Thomas Schwandt is Professor of Educational Psychology at the University of Illinois at Urbana-Champaign (USA) and Editor of the *American Journal of Evaluation*.

Elliot Stern

Elliot Stern, an evaluation practitioner and researcher based in the United Kingdom, edits the journal *Evaluation*. He is visiting Professor at Bristol University and Professor Emeritus at Lancaster University. He is a past President of the European Evaluation Society.





The photographs by Josef Rabara were taken at the 10th EES Biennial Conference held in Helsinki (October 3–5, 2012)