THE EES NEWSLETTER

EVALUATION
CONNECTIONS

EES EUROPEAN
EVALUATION
SOCIETY

JUNE 2013

ARTICLES – NEWS – EVENTS

CONTENT

A Message from the President and the Vice President

The EES 2014 Biennial Conference will take place in Dublin from 1 to 3 October 2014 in the Convention Centre Dublin, Ireland. Its theme ("Evaluation for an Equitable Society") encompasses key evaluation principles – Independence, Partnership and Participation. Given their popularity in past years, pre-conference training workshops will be held on 29 and 30 September 2014.

The international financial crises that have swept through the global economy over the past decades reflect fundamental changes in the economic and financial system leading to wider gaps in wealth among countries and within countries.

Inequalities, exclusion, vulnerability, discrimination and exploitation plague society despite international efforts to reduce poverty and foster equitable growth. In parallel, the demand for tailor-made and formative evaluations that focus on real-life problems and are grounded in democratic values has considerably increased. Evaluation is being called upon to make a difference.

But how should the evaluation discipline respond?

Is it equipped to take on the diverse, unprecedented and inter-connected challenges of economic disparities, social tensions and poor governance? Are contemporary evalu-ation models, practices and priorities fit for purpose?

How should the enabling environment for evaluation evolve to meet the challenges of social injustice and environmental destruction?

The programme of the Dublin Conference will be structured to provide a forum for exchanges of views, experiences and empirical evidence from different perspectives to help the evaluation community combat the scourge of economic disparities in this new era.

As in previous biennial conferences, the programme will comprise strands that cover all facets of the evaluation discipline, i.e., governance, methods and practices, ethics and capabilities, public, private and voluntary sector concerns, etc.

Cross-cutting all strands, the major principles of sound evaluation practice – independence, partnership and participation – will be debated and analysed so as to find ways of better promoting the values and priorities of democratic evaluation.

We look forward to welcoming you in Dublin.

María Bustelo,
President, European Evaluation Society
Claudine Voyadzis,
Vice President, European Evaluation Society

EDITORIAL: FROM EVALUATION THEORY TO EVALUATION PRACTICE

Robert Picciotto

"Theory is splendid but until you put it into practice it is valueless"
J.C. Penney

Linking up evaluators from all parts of the world is the mandate of all evaluation associations. The bewildering expansion of diverse evaluation associations and networks is described below by Jim Rugh while Murray Saunders explores the theory of change that animates *Evalpartners*, a global movement designed to promote evaluation excellence and equity-oriented evaluations. Your Newsletter has a similar remit – to encourage cross-fertilization of evaluation ideas across borders. Thus this issue of Connections brings together evaluation thinking and lessons of experience from various European countries as well as New Zealand, Canada, India and Brazil.

The intersection between evaluation theory and practice – a theme evoked in Stewart I. Donaldson's article in the December 2012 edition – is explored in the articles that follow. Theory provides practitioners with ideas and tools that evaluators put to work in order to grasp and tackle a problematic situation. Conversely practicing evaluators generate data and knowledge that challenge existing mental models and call for new interpretations, concepts and approaches.

Four articles in this issue explore the frontiers of the evaluation discipline. Bob Williams leads the way with a lucid exposition of three core concepts that the systems field has contributed to evaluation practice in complicated and complex situations: interrelationships, perspectives and boundaries.

Next, Richard Hummelbrunner and Martin Reynolds apply these tools to organizational and social learning processes. They demonstrate that learning to adapt, learning to change and learning to learn involve distinct mechanisms that bring values to the centre stage of stakeholders' interrelationships, illuminate their perspectives and delineate the boundaries of evaluation practice within given social contexts.

Dealing with change and uncertainty in impact evaluation practice is the subject of John Mayne's piece: in complex contexts the identification of valid causal links between an intervention and its observed impact calls for judicious use of theories of change combined with contribution analysis. In turn, Rick Davies investigates the problems that arise when many theories of change are deployed to probe complex situations. He shows that decision tree models can help solve such problems in user friendly ways through systematic fitness tests and effective use of data. Software packages facilitate the use of decision trees in a wide range of evaluation assignments.

The new information technologies also enhance real world, real time evaluation practice. Thus, Kai Matturi's article demonstrates conclusively that digital data gathering allows instantaneous access to program implementation data, improved data reliability, reduced data loss, rapid data analysis and centralized data management. On the other hand, I.C. Awasthi shows that web-based management information systems can be dysfunctional when success indicators are poorly selected, programme data are not authenticated and/or accessible or when they are delivered in an untimely way to guide decision making.

The potential of systems theory allied with the new communications technologies to resolve the tensions between intended policy and programme goals and real world constraints is illustrated by both articles. Similarly, Paula Saikkonen implicitly engages in systems thinking when she highlights the unexpected influence that monitoring protocols can trigger at the social worker-citizens interface and when she points to the need for explicit learning loop mechanisms to help adapt national welfare policies and control systems to diverse local needs.

In the same vein, Giorgio Garau et. al. focus on the potential of decision tree mapping for the design of monitoring systems and the judicious definition of the distinctive accountabilities and reciprocal obligations embedded in public-private partnerships. Finally, Elizabeth Moreira dos Santos et. al. confirm that participatory monitoring systems can be powerful instruments of popular mobilization, capacity building and social innovation.

In a nutshell the core message of this issue is that social learning is more likely to take place when sound evaluation theory is put into practice. In the words of Lewin (1952): *"there is nothing more practical than a good theory"*.

Reference

Lewin, K. (1952). Field theory in social science: Selected theoretical papers by Kurt Lewin. London: Tavistock.

EVALUATION ASSOCIATIONS

THE GROWTH AND EVOLVING CAPACITIES OF VOLUNTARY ORGANIZATIONS FOR PROFESSIONAL EVALUATION[1]

Jim Rugh

In recent decades, civil society has increasingly played a central and active role in promoting greater accountability for public action, through the use of evaluation. National and regional Voluntary Organizations for Professional Evaluation (VOPEs) grew from 15 in the 1990s to more than 140 by early 2013. Acknowledging the enhanced role of civil society in general, and VOPEs in particular, UNICEF and IOCE launched the EvalPartners Initiative in early 2012. This is a global initiative that promotes coordinated efforts among development organizations, governments and civil society, with the aim of strengthening civil society evaluation capacity, in order to fortify the voice of civil society in policy-making and in promoting equity-focused and gender-responsive evaluations. EvalPartners was met with a surge of enthusiasm evidenced in the joining of 27 members, including major international development agencies and all regional VOPEs, within a few months of its launch.

The goal of the EvalPartners Initiative[2] is to contribute to the enhancement of the capacity of Civil Society Organizations (CSOs) – notably, VOPEs – to influence policy-makers, other key stakeholders and public opinion, so that public policies are evidence-informed and support equity in development processes and results. There are three main expected outcomes of EvalPartners:

1. VOPEs are stronger. Their institutional and organizational capacities are enhanced.

2. VOPEs are more influential. They are better able to play strategic roles in strengthening the enabling environment for evaluation within their countries. In so doing, they help to improve national evaluation systems and to promote the use of evaluation evidence in the development of policies geared towards effective, equity-focused and gender-responsive development results.

3. VOPEs develop sustainable strategies to enhance the evaluation skills, knowledge and capacities of their members, and of evaluators more widely, to manage and conduct credible and useful evaluations.

One of the initial activities of the EvalPartners Initiative was a mapping survey to update the database of VOPEs maintained by IOCE in order to learn more about them.

The first phase of the exercise was to update contact information for all the VOPEs on the IOCE database, and continue the search for others. The second phase was to solicit current information about each VOPE. We previously had basic profile information on only 54 VOPEs. Ninety four VOPEs responded to the survey and provided basic information about their purposes, memberships and organizational capacities. Based on the responses, the third phase of this mapping exercise focused on the VOPEs with significant experience in advocating enhanced evaluation policies and systems. These organizations were invited to provide descriptions of their experiences in the form of case studies. Such case studies were received from 38 national and regional VOPEs[3]. Figure 1 below gives the 'birth' years of some of the VOPEs, highlighting the Big and Regional VOPEs. It is extracted from a longer list that includes the formation years reported by 83 VOPEs.[4] Of the currently existing evaluation societies or associations, the first to be formed was the Canadian Evaluation Society (CES) in 1981. Though the American Evaluation Association (AEA) was not formed until 1986, it was actually a merger of two existing as-

Year founded	Country	Acronym	Cumulative total
1981	**Canada**	**CES**	1
1986	**USA**	**AEA**	2
1987	**Australasia**	**AES**	3
1988	Canada/Quebec	SQEP	4
1992	UK	UKES	5
1994	France	F3E	6
1994	**Europe Regional**	**EES**	7
1995	Malaysia	MES	8
1996	Peru	Red EvalPerú	9
1997	Germany + Austria	DeGEval	10
1999	Niger	RéNSE	11
1999	Sri Lanka	SLEvA	12
1999	**Africa Regional**	**AfrEA**	13
2000	**Russia & CIS Regional**	**IPEN**	17
2004	**Latin America & Caribbean Regional**	**ReLAC**	28
2008	**Europe Regional**	**NESE**	56
2011	**MENA Regional**	**EvalMENA**	75
2012	Uruguay		83

Figure 1. Years VOPEs were formed, emphasizing big and regional VOPEs.

sociations, one of which was formed in 1979 (so it could claim to be older than CES!).

While recognizing that many evaluators are members of more than one VOPE, i.e. that there is substantial (but unknown) overlap between the membership numbers reported by VOPEs,[5] the total aggregate membership numbers is indicative of the size of the evaluation community. It adds up to about 33,000. This points to the growing number of people who have an interest in or responsibilities related to evaluation, as suppliers or commissioners or academics.

Based on the survey responses, 39% of these VOPEs are informal networks, 13% say that they have adopted a charter and bylaws but are not yet officially recognized, and 47% report that they are legally recognized by their governments (presumably after adopting a charter and bylaws).

The remarkable growth in VOPEs is illustrated in Figure 2 while Figures 3 and 4 provide the sources of the information and Figure 5 summarizes the membership data
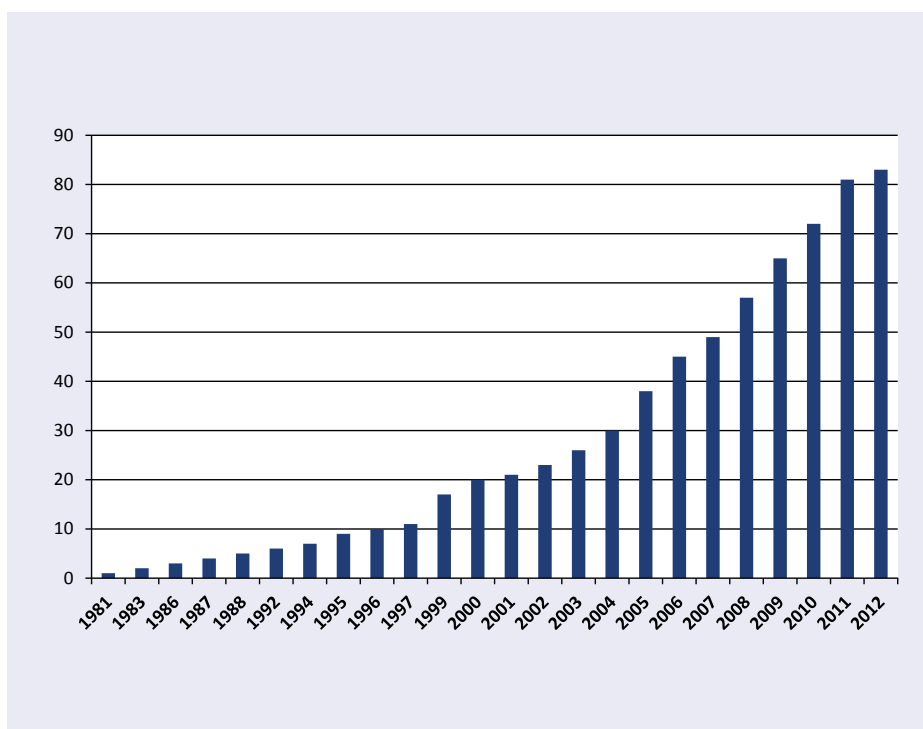

Figure 2. Cumulative number of VOPEs in existence, by year.

Figure 3. EvalPartners VOPE survey by the numbers[6].

Countries in which we have some VOPE contact information	112
Number of national VOPEs identified	141
Number of national VOPEs *verified* (via survey responses or active websites)	96
Number of countries with one or more *verified* VOPEs	86
Regional VOPEs	12
International VOPEs	11

Figure 4. Survey responses and case studies received.

National / regional VOPEs to which survey questionnaire was sent	140
Survey replies received	94
Case studies received	38

One of the rationales for conducting the EvalPartners survey was to ascertain how engaged VOPEs are in terms of addressing the enabling environment for evaluation, e.g. by advocating for enhanced evaluation-related policies and systems on the part of their governments. Figure 6 indicates that there is a wide range of involvement in policy advocacy.

Name of VOPE	Acronym	Membership
American Evaluation Association	AEA	7,755
Red de Seguimiento, Evaluación y Sistematización en America Latina y el Caribe	ReLAC	3,847
Rede Brasileira de Monitoramento e Avaliação (Brazilian M&E Network)	BMEN	3,519
Red de Monitoreo y Evaluacion de America Latina y Caribe	RedLacMe	2,557
Canadian Evaluation Society / Societe canadienne d'evaluation	CES / SCE	2,016
Australasian Evaluation Society	AES	1,034
Gesellschaft für Evaluation e.V.	DeGEval	722
Société Française de l'Évaluation	SFE	600
International Program Evaluation Network (CIS)	IPEN	556
Indonesian Development Evaluation Community	InDEC	554
Sociedad Española de Evaluación (Spain)	SEE	550
Associação Brasileira de Avaliação Educacional (Brazilian Association of Educational Evaluation)	ABAVE	500
Society for Monitoring and Evaluation, Nigeria	SMEAN	452
European Evaluation Society	EES	411

Figure 5. Membership numbers of largest VOPEs.

Policy advocacy score = 10 (out of 10, i.e. very actively engaged)	19	29%
Policy advocacy score between 5–9 (quite actively engaged)	14	21%
Policy advocacy score between 1–4 (very little engagement)	14	21%
Policy advocacy score = 0 (not involved at all)	19	29%
Total number of survey responses scored	66	100%

Figure 6. Involvement in policy advocacy on the part of VOPEs.

We will now turn from the quantitative growth of the evaluation profession as evidenced by the growing numbers of VOPEs, to some perspectives on what many VOPEs are achieving in addressing the enabling environment for evaluation, including advocating for national monitoring and evaluation policies and systems.

A typical scenario in the early formation of an evaluation network is that a few individuals get over their competitive urges and decide to get together to share experiences in methods applied to evaluation. Following informal exchanges, they might organize workshops, led by some of their own members or outside experts, in order to share theories and experiences. The next phase involves the design of a constitution and of bylaws so as to get officially registered as an association.

Thus the first phase is typically focused on skills development and the second phase on building the VOPE's own institutional capacity. A growing number of VOPEs are going beyond addressing the 'supply side' (capacities of members to conduct evaluation) to tackle the 'demand side' – i.e. the environment that influences requests for evaluation, including the Terms of Reference shaping what clients are asking evaluators to do. Especially in countries where most evaluations have been done to respond to the demands of external donors, VOPEs have begun to realize the need for national governments to appreciate the value of evaluation as a sound governance practice. There are many examples of VOPEs that have achieved significant influence on governments with respect to the design and monitoring of high-level policies at national, ministerial and provincial levels.

---

1 This article is an adaption of a chapter in the book *Voluntary Organizations for Professional Evaluation: Learning from Africa, Americas, Asia, Australasia, Europe and Middle East* to be published by UNICEF with EvalPartners and IOCE. It will be available on www.MyMandE.org/EvalPartners.

2 For more information see the MyMandE.org/EvalPartners website.

3 All of the survey responses and case studies have been uploaded at www.ioce.net/members/national_organizations.shtml.

4 As of mid February, 2013, 96 VOPEs had been verified (via survey responses or currently up-to-date websites) but dates of formation were only reported by 83 VOPEs.

5 AEA, for example, reports that 15% of its membership (over 1,000 people) are "international", i.e. citizens of other countries, therefore presumably also members of their national VOPEs in addition to being members of AEA (and perhaps other big and/or regional VOPEs as well.)

6 As of the end of 2012.

EVALPARTNERS AND ITS THEORY OF CHANGE

Murray Saunders

EvalPartners is a global initiative. Some have called it a movement. Its core purpose is to enhance the capacities of Civil Society Organizations (CSOs) in general, and Voluntary Organizations of 'Professional' Evaluation (VOPEs[1]) in particular, to influence public opinion, policy makers, and other key stakeholders so that public policies are based on evidence, and incorporate considerations of equity, including ethnic and gender equality.

From an international perspective the initiative is timely. The United Nations Secretary General, Ban Ki-moon recently addressed a High Level Meeting on *"UN Results: Are we achieving them? How do we know?"* He told United Nations Ambassadors and the Heads of Evaluation Offices of 43 UN agencies that *"self-evaluation has to be part and parcel of our routine management work. Every good manager must make the effort to look ahead and look behind, and determine the effects, intended or unintended, of our work."* He added: *"All of us share a responsibility to strengthen the evaluation function. We have to tackle the challenge at several levels"*[2].

The theory of change embedded in EvalPartners is predicated on three interconnected propositions. At its heart is the idea of capacity building. A strengthened VOPE should be able to: (1) increase the quality and supply of evaluations by contributing to national evaluation capacities, promote the professionalization of evaluation and attract new evaluators. (2) develop and promote national evaluation standards and ethics in evaluation; and (3) strengthen evaluations at individual and institutional levels so as to promote the demand for evaluation in all contexts. In turn, increased demand for evaluation will require increased VOPEs' contributions to the formulation of national evaluation policies, the design and implementation of country-led evaluation systems, the facilitation of dialogue between evaluation users and the production of sound evaluative evidence, etc.

The expected outcomes of EvalPartners are three-fold:

1. Voluntary Organizations of Professional Evaluation (VOPEs) have a strengthened Institutional capacity

2. VOPEs will be able to play a strategic role within their countries, contributing to country-led evaluation systems and policies, including through better support from regional and international networks/associations (including IOCE and the more developed VOPEs) and institutions (including, inter alia, UNICEF); sharing lessons learned of similar experiences in other countries; and peer to peer mutual support.

3. VOPE members will have stronger evaluation capacities, e.g. by attending live webinars with international keynote speakers; e-learning; mentoring programmes; and training sponsored or organized by local institutions and more developed VOPEs.

Since the 1980s, Civil Society Organizations have been playing increasingly central and active roles in promoting greater accountability for public action through various means, including evaluation. National and regional VOPEs have grown from 15 in the 90s to more than 120 in 2012. There is tremendous scope for exchanges of home-grown and country-driven solutions, ideas and experience to support capacity development in evaluation. In December 2012 EvalPartners convened the International Forum on Civil Society's Evaluation capacities in Chiang Mai, Thailand. The Forum facilitated the sharing of good practice and lessons learned by VOPEs and other stakeholders engaged in Evaluation Capacity Development (ECD).

Thus the *policy strategy* of EvalPartners is to build capacity at individual, institutional and enabling environmental domains. The main *instruments* used by EvalPartners so far have been targeted funding of capacity building activities and partnership building. Such funding has been critical in moving EvalPartners from exhortation to the provision of resources for change. Forum also identified EvalPartners' *mechanisms* which enact the broad intervention priorities:

• Facilitation of peer-to-peer collaborations among VOPEs;
• Development of a toolkit on VOPE institutional capacity;

• Generation of new knowledge on VOPE operation;
• Development and implementation of advocacy strategies to enhance the enabling environment for evaluation;
• Promotion of equity-focused and gender-responsive evaluation.

Participants signed the EvalPartners Chiang Mai Declaration committing themselves to EvalPartners' objectives and principles. They include:
• Equity and social justice as central values;
• Recognizing that the country-led evaluation systems and functions are vital to ensure that development interventions implemented by international donors and governments themselves are effective, efficient and responsive, achieve desirable development outcomes and improve the quality of life of all;
• Recognizing that civil society organizations in general, and VOPEs in particular, must play a key role in influencing and enhancing the demand for evaluation and the use of evaluation results; in developing the capacity of national and local authorities, as well as communities, NGOs, academia and the private sector, to endorse and support evaluations of their own policies and programmes.

EvalPartners' meta-theory of change is based on the assumption that a robust and legitimate set of evaluative practices can play a progressive role in all hemispheres. In this sense, evaluative practice can be understood as a response to the need to build social capital in public sector processes, protocols and procedures. Some policy domains are slippery, ambiguous and unformed, relying on enabling networks, collaborations and partnerships. Increasingly, evaluation is being understood as part of the process by which 'policy learning' or institutional growth and development is encouraged.

This helps to explain the accelerated growth in the last two decades of civil society organizations in general – and Voluntary Organizations for Professional Evaluation (VOPEs) in particular – and it justifies the need to support and develop their capacity. There are of course many definitions of civil society organizations but there is broad based consensus that the term refers to the wide array of non-governmental and not-for-profit organizations that have a presence in public life, express the interests and values of their members (and other stakeholders) based on ethical, cultural, political, scientific, religious or philanthropic considerations. It is to the increased power of these forces that EvalPartners' energy is directed.[3]

---

1 VOPE can be formally constituted (in the form of associations or societies) or operate informally (in the form of networks)

2 See http://mymande.org/evalpartners/UNSG_Speech_Evaluation_EP

3 See **Voluntary Organizations for Professional Evaluation (VOPEs)** Learning from Africa, Americas, Asia, Australasia, Europe and Middle East *Editors* Jim Rugh and Marco Segone (For free download see http://www.mymande.org/selected-books

EVALUATION THEORY

SYSTEMIC THINKING IN EVALUATION: THREE CORE CONCEPTS

Bob Williams

In recent years evaluators have increasingly looked to the systems field to enhance the quality of 'real world' evaluations. While systems methods do not always transfer easily to evaluation, this article suggests that the *principles* that underpin systems methods are fully relevant to evaluation practice in complicated and complex situations. Specifically three core ideas shape the relationship between systemic thinking and evaluation: inter-relationships, perspectives and boundaries. They evoke *twelve evaluative core questions* displayed in the boxes below.

Interrelationships

How things are connected and with what consequence stems from the earliest thinking about systems. The concept is deeply embedded in the popular imagination. When we talk about the education system or the health system, we imagine a set of actors and processes that are interconnected. The popularity of system dynamics and complex adaptive systems in many parts of the world confirms that inter-relationships matter. However, systemic thinking doesn't concern itself with just any inter-relationships. It focuses especially on particular aspects of them:

Dynamics How the interrelationships affect the behaviour of a situation over a period of time

Non-linearity The effect of an interrelationship is typically unrelated to its size. This is frequently caused by *feedback*. Exponential growth patterns are frequent in ecology.

Context sensitivity The same interrelationship may have different results in different contexts. Disease control methods that work in Thailand may not work in the Philippines.

Complexity Interrelationships are often so complicated or complex that they cannot be explained simply in terms of causes and effects.

These are not definitive questions, but good places from which to consider how to make existing practice more systemic.

Perspectives

Just looking at interconnections does not make an inquiry or intervention systemic. People will see and interpret those inter-relationships in different ways depending on their perspectives. A local cafe owner might view issues having to do with preventing the spread of Listeria poisoning quite differently than a health service inspector. What a health inspector is apt to do when he or she "sees" something will be different from what the cafe owner is likely to do. Different perceptions promote different behaviours that affect how the situation unfolds. Indeed, what we see as unintended or unexpected outcomes patterns often results from our inability or unwillingness to understand or appreciate other people's perceptions and behaviours. We use words like "unintended effects" without considering that somebody somewhere may have intended them. Hence to fully comprehend the dynamics of a situation we must identify and understand the range of relevant perspectives that people bring to it. To do so, it is helpful to distinguish between three forms of perspective: stakeholders, stakes and framings.

Stakeholders and stakes Stakeholders are groups of people or things that have a common role in a situation or intervention (e.g. teachers, consumers, writers). In contrast, stakes relate to individual values and motivations (e.g. wealth, honour, fairness, past history, purpose, ideas of professionalism). People belonging to different stakeholder groups may share the same stakes, and any one stakeholder grouping will contain within it several different (perhaps conflicting) stakes.

Framings Deliberating on the impact of different stakeholders and stakes gives us an opportunity to frame issues. Framing

Five basic questions for addressing inter-relationships systemically:

1. What is the structure of the inter-relationships within the situation (eg are the components arranged hierarchically, randomly, sequentially)?
2. What are the processes among components of that structure? (eg simple, complicated, complex, linear)?
3. What is the nature of the inter-relationships (e.g., strong, weak, fast, slow, conflicted, collaborative, direct, indirect)?
4. What patterns emerge from these interrelationships over time, with what consequences and for whom (eg simple, complicated, complex, cyclical)?
5. What are ways in which these complicated and complex dynamics can be identified and managed effectively?

is more than listing stakeholder views. It is trying to work out what the situation is – or could be – about. It identifies how people understand a situation and how they will behave. It is the lens through which you view a situation and the stakeholders involved.

Let's assume your interest in a rock concert situation concerns the play list. Each of the framings in the text box suggests different ways of constructing the play list and thus different songs. A "fun day out" framing (lots of dance oriented music) comes up with a very different playlist than an "income generation" framing (the latest iTunes releases) or a nostalgia framing (greatest hits playlist).

Breaking down the situation into different framings allows you to construct a set list that satisfies most attendees. The ageing population of 1970s and 1980s rock stars are very skilled at working within multiple framings of their performance.

Some possible framings of a Rolling Stones concert:

- a fun evening out
- income generation
- cultural expression
- marketing product
- nostalgia

Four basic questions for exploring perspectives systemically:

6. Who or what are the key stakeholders within the situation (e.g. beneficiaries, victims, human, environmental)?
7. What are the key stakes (eg motives, values, purposes, financial, lifestyle, professional)?
8. What are the different ways in which the situation can be understood or framed?
9. How are these different framings going to affect the way in which stakeholders act or have expectations of each other and thus need to be considered?

Framings are not necessarily "right" or "wrong". Sometimes seeing things through a different framing helps solve a tricky problem. There's a well known story[1] about Russ Ackoff – a key figure in the systems field. A big machine tool manufacturer experienced considerable fluctuations in demand for his products. This led to low morale, poor productivity and bad industrial relations. Russ was called in to sort out this "production smoothing" problem. After some failed attempts to model various production scenarios, he encouraged the company to reframe the situation as a "demand smoothing" issue rather than a "production smoothing" issue. To smooth demand you needed to manufacture a product that was counter-cyclical to the demand

of the existing product line. Road-building equipment was found to be counter-cyclical to that for machine tools and required much of the same technology and marketing and distribution skills. Fluctuation in demand was reduced to a minimum, which in turn resulted in stable employment and good morale.

Boundaries

Every endeavour has to set boundaries. That's because a boundary differentiates between what is "in" and what is "out," what is deemed relevant and what is considered irrelevant, what is important and what is unimportant, what is worthwhile and what is not, what suits the one in a position of power and what doesn't, who benefits and who is disadvantaged. Boundaries are where values are exposed and disagreements are highlighted. A lot of power issues get wrapped up in boundaries—just as the person with the magic marker controls what goes on the whiteboard, the person who decides the boundaries exercises powerful influence on any situation.

Scale is a big boundary issue. Take food production. The boundary can be set at different scales geographically (village, country, region, global), sectorially (apple growing, horticulture, agriculture, food production), professionally (research microbiologist, microbiologist, biologist, natural scientist, scientist). Decisions about scale are extremely important boundary decisions because something that might be seen "valuable" at one scale may not be valuable at another scale (e.g. the use of certain pesticides may benefit particular crops but devastate biological diversity and the overall food production economy).

Setting boundaries is not optional. One cannot do everything, consider everything, see everything or record everything. Treating boundaries systemically means that you set boundaries consciously and consider the implications.

From an ethical point of view, you hold certain values and those values reflect your ethical stance. If you believe that women have an essential role in preventing Dengue Fever, then you will want your intervention to ensure that their voices are heard and acknowledged.

From a pragmatic point of view, those who are marginalized are not likely to take things lying down. A policy maker may not like to consider the interests of loan sharks in addressing housing foreclosures, but by not doing he or she incurs a risk that they will oppose the intervention and hinder its execution. The last question of course raises the further question of what kind of harm to whom. Hence the iterative nature of boundary questions; they raise the possibility that you may need to reassess your initial judgments on interrelationships and boundaries.

Three basic questions for surfacing boundary decisions systemically:

10. Which interrelationships are privileged and which are marginalized? With what effect on whom?
11. Which perspectives are privileged and which are marginalised? With what effect on whom?
12. How can the ethical, political, and practical consequences of these decisions be managed, especially those that cause harm or have the potential to cause harm because they exclude an interrelationship or perspective?

Summarising: Systems methods are often hard to learn and difficult to apply in evaluations. However, evaluators can apply a systems approach using systems tools and/or adapting their existing tools of evaluation by addressing three important factors within a situation: (1) the inter-relationships between aspects of a situation; (2) the perspectives through which that situation can be understood; and (3) the boundaries that are necessary to allow us to address a situation. Evaluators can systemically gauge the pertinence of these factors by asking the *twelve core questions*.

1 Recounted in Jackson (2002) *System approaches to management*, p. 235.

SYSTEMS THINKING, LEARNING AND VALUES IN EVALUATION

Richard Hummelbrunner and Martin Reynolds

Applying the three core systems concepts – interrelationships, perspectives and boundaries (see Bob Williams' article in this issue) – for the evaluation of a "situation" has implications for the type of learning that it generates. It also helps to make explicit its value base. This article proposes a conceptual framework that connects the three systems concepts with learning and values.[1]

First we examine the implications for learning. To do so we use a model that has been widely used in the Organizational Development literature. It is based on the work of Gregory Bateson (1972) as well as Chris Argyris and Donald Schön (1978) and addresses the purpose and extent of learning. It distinguishes between three types of learning:

- *Single loop learning (Learning to adapt):* results in a change of strategy or tactics without questioning the underlying goals or assumptions. It helps to control individual behaviour within existing decision making protocols; provides short-term solutions to implementation problems and deals with symptoms more than root causes. The core question is *'Are we doing things right?'*
- *Double loop learning (Learning to change):* by reflecting on goals and assumptions, one probes the generative mechanisms of problems, their underlying causes and their consequences. This leads to adjustments in strategy and to better mid- and long-term course corrections in response to contextual changes. The core question is *'Are we doing the right things?'*
- *Triple loop learning (Learning to learn):* by reflecting on the learning mechanisms, existing rules are challenged and possibly changed in ways that affect knowledge acquisition and behaviour, i.e. by identifying different patterns of recognising and handling problems or coping more appropriately with contextual changes. The core question is *'What makes this the right thing to do?'*

Although each of these levels addresses different questions, the progression from single

to double and triple loop learning can be expected to lead to deeper and more sustainable learning.

Based on this model, we associate each of the three systems concepts with a specific loop of learning. Figure 1 below illustrates these connections for the generic case of evaluating the effects of an intervention:

- *Single loop learning:* The focus is on *inter-relationships*, primarily between the intervention and its effects, but also within them (e.g. between the actions of an intervention or the various effects produced). In case of divergence from original plans, adaptive recommendations are made; for example, modifying a strategy or activities in order to better achieve stated aims and objectives. Significantly, the purpose of the intervention is not questioned.
- *Double loop learning:* Assumptions underpinning an intervention can only be reflected if multiple *perspectives* are taken into account. When acknowledging that a situation can be framed in different ways, this also questions the purpose and goals of an intervention.
- *Triple loop learning:* Here the focus is on the *boundaries* inevitably made with any intervention and its evaluation. Reflecting on boundary judgements is very helpful (and needed) for critically reflecting on the rules and relations of power that affect behaviour and cognition patterns (Flood and Romm, 1996). This notably involves looking at the power relations that determine the boundaries of an intervention and its evaluation, including the role of evaluation commissioners and evaluators themselves.

The key role of the evaluator is in assigning value. Each of the learning loops can be associated with a different set of values:
- *Single loop learning* is based on *instrumental values* embedded in an intervention. These underpin the intervention logic and can be derived from the respective documents, either explicitly or (probably more often) only implicitly. Utility is perhaps the best

example of such value. Instrumental values inform evaluative measures regarding issues of 'efficacy' (does it work?) and 'efficiency' (how well does it work using available resources?).
- *Double loop learning* is based on the *intrinsic value* underpinning the various framings of an intervention and/or the wider situation being evaluated. They can include personal, organizational or social values. Intrinsic values inform evaluative measures regarding issues of 'relevance' (why is it important that the intervention works and works well?) and 'effectiveness' (are the right things getting done?)[2]
- *Triple loop learning* is based on *critical value*; that is, value in reflecting on the rules and customs that govern dominant behaviour and cognition patterns in a particular context. Critical values inform evaluative measures regarding issues of equity and emancipation (what and who determines the importance of some measures of success over others?)

Figure 2 is an attempt to integrate all of these associations into a single framework.

The framework can be used to interrogate the coherence among the various components of an evaluation assignment. For example, is the type of learning envisaged in line with the evaluation's value base? Can the value-base be modified or expanded if a deeper level of learning is envisaged or needed? Which of the systems concepts might be more appropriately applied in making value judgments in an evaluation?

The progression from single to triple loop learning is expected to lead to more sustainable learning. Similarly, the progression in focusing from *Interrelationships* to *Perspectives* and *Boundaries* indicates the extent and depth of systemic practice. Both sustainable learning and systemic practice in evaluation can be enhanced by applying progressively wider measures of value. This does not imply that the 'upper' level should and can always be reached. Often only one specific level might be feasible or can be appropriately attained
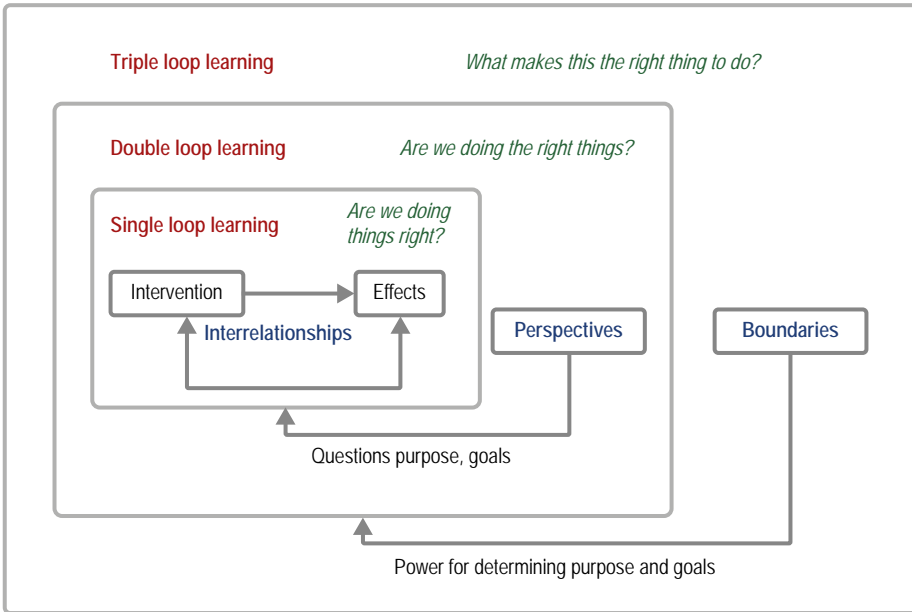
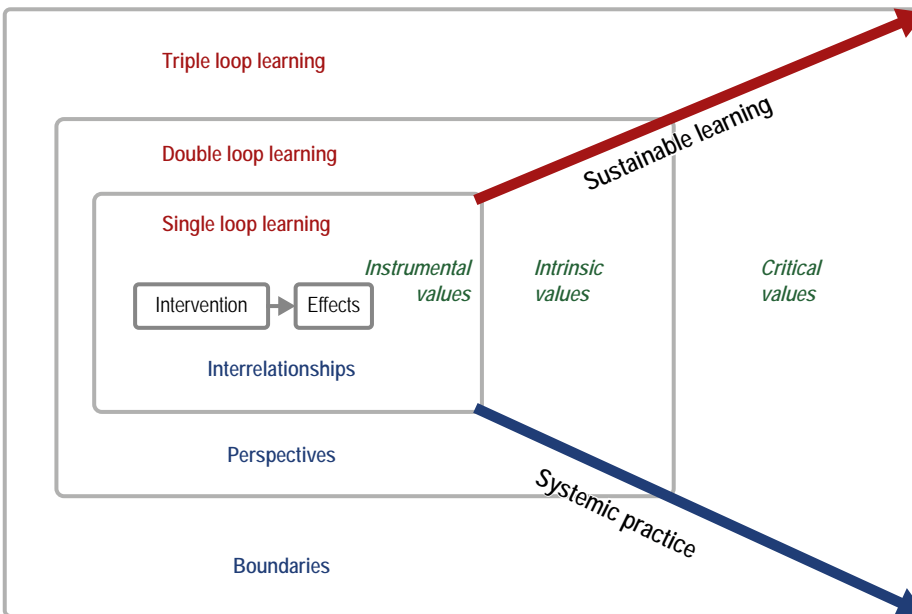*Fig. 1: Types of Learning and Systems Concepts.*

Fig. 2: A systems – based framework for rigor in evaluation.

ing type) can enhance the relevance, validity and credibility of evaluations.

We are aware that this framework is still tentative and provisional, and that some connections and their implications invite further exploration. Additional associations are possible and could be integrated in this framework. For instance, the OECD's Development Assistance Committee (DAC) evaluation criteria could be grouped and aligned along the same three levels (i.e. efficacy/efficiency; relevance; and sustainability/impact). The framework may also be applied to inform assessments of governance and/or performance by distinguishing between different levels of authority (power over, power with, and power to) and agency (personal, organizational, social).

References

Argyris, C. and D. Schön (1978). *Organizational Learning.* London, Addison-Wesley.

Bateson, G. (1972). "The logical categories of learning and communication." *Steps to an Ecology of Mind*: 279-308.

Flood, R. L. and N. Romm (1996). *Diversity Management: Triple Loop Learning.* Chichester, John Wiley & Sons.

1 The ideas in this article were originally presented by the authors in a panel session at the workshop 'Impact, Learning and Innovation' at the Institute of Development Studies, Brighton UK (March 26–27 2013).
2 The term 'effectiveness' can have different meanings in different contexts of use; sometimes used interchangeably with efficacy. In this article we make a clear distinction between evaluative criteria of efficacy ("getting things done") and efficiency ("getting things done right") – both of which constitute single loop learning – and effectiveness ("getting the right things done"). Effectiveness in this latter sense implies relevance and invokes double loop learning.

given the actual circumstance and conditions of an evaluation. But the framework is helpful for reflecting on the constraints and limitations of an evaluation, as well as pointing at hidden opportunities that might otherwise be missed.

This framework applies systems thinking to the evaluation process by proposing three sets of typologies, their respective boundaries as well as some suggested correspondence between them. We believe that reflections based on such a framework can add rigour to evaluation practice. The value base of an evaluation can be made more explicit and congruent with the evaluation mandate. Addressing the appropriate value base (in coherence with the envisaged learn-

MAKING CAUSAL CLAIMS

John Mayne

Evaluation is often expected to provide credible evidence about the causal links between the actions embedded in an intervention and the intended outcomes and impacts that have been observed— i.e. to make a credible claim that the intervention 'caused' the observed results. But in many real-life situations, interventions are only one of several factors that may have caused the results that were observed: contextual variables may matter and other events may be required for the intervention to 'work' as intended. Yet we still want to identify causal links between the intervention and the observed results. We want to establish if the intervention 'made a difference'.

Experimental designs have traditionally been viewed as the most appropriate for making causal claims. But in many situations, such designs are not feasible or they are impractical. What can then be done? In this article, I argue that interventions need to be thought of as *contributory causes* and that evaluation can provide evidence that an intervention *made a difference* by showing the intervention to be a contributory cause.

Contributory causes are causes that on their own are neither necessary nor sufficient. Smoking and lung cancer is a well-known example: smoking isn't necessary to get lung cancer and not all smokers get lung cancer. Similarly, many interventions are not on their own sufficient to bring about the intended results and the intended results can be realized through means other than the intervention. In what sense can we really say that smoking 'causes' lung cancer?

Consider another example: an intervention aimed at improving the education outcomes for girls in a setting where education for girls is not a priority, through raising the knowledge, skills and awareness of teachers in schools with respect to girls' education. For this intervention to work, one can easily think of several other events and conditions (supporting factors) that are probably needed for the intervention to work as intended, i.e., for education outcomes of girls to improve:

- the willingness of teachers to support the education of girls
- the ability of girls to get to the schools
- the willingness of parents to allow their daughters to attend schools, and to study at home
- the adequacy of the schools to accommodate girls

Say that after several years we indeed see an increase in education outcomes for girls. What kind of a causal claim can we make?

The intervention itself is clearly not sufficient. Rather, the intervention and the other supporting factors are a *causal package* that is expected to be sufficient to bring about better education outcomes for girls. Further, the intervention on its own is not necessary: there might be other kinds of interventions that could bring about better educational outcomes, such as investment in school access, a different teacher-pupil ratio, etc. An evaluation of the intervention would therefore want to assess if the causal package worked, i.e., *in this case* whether education outcomes for girls improved as a result of the causal package, and since it is an evaluation of the intervention, whether the intervention was a needed part of the causal package?

If this can be demonstrated, then it is possible to make the causal claim that the intervention was a *contributory cause* and we can say that the intervention *made a difference*. It made a difference in the sense that is was a necessary component of a causal package that was sufficient to bring about the intended results. A DFID report by Stern, Stame, Mayne, Forss, Davies and Befani (2012) was perhaps the first to introduce the idea of a contributory cause in the context of evaluation, but contributory causes are well known in the literature on causality (Mackie 1974). In this short article, only the deterministic case – a cause either is or is not sufficient – is discussed, but the same ideas hold if one wants to consider the probabilistic case, where the terms likely sufficient and likely necessary are used (Mayne 2012: 276).

But we probably want to say more than the intervention was one of several contributory causes. We can talk about the *role* that the intervention played in the causal package. For example we might be able to show that the intervention was the *trigger* that starting the causal chain leading to the intended results, such as improved education outcomes. Or we may be able to show that the intervention played a more modest role in supporting the changes along the causal chain. In such cases, we could then say that the intervention was a *principal contributory cause*.

All well and good as far as ideas and concepts go, but how would an evaluation set about actually showing that an intervention is a principal contributory cause? Contributory causes and sufficient causal packages are quite closely related to theory-based approaches to evaluation, and theories of change. I have argued that good theories of change are in fact models of an intervention as a causal package (Mayne 2012). Many theories of change are little more than results chains, showing the expected sequence of outputs, outcomes and impacts. A 'good' theory of change (Figure 1) also includes the assumptions and risks associated with each link in the results chain. It spells out what events and conditions have to happen for the results chain to work. The sum of these necessary events and conditions are the supporting factors of the causal package.

With this relationship made, showing the intervention is a contributory cause implies showing that the theory of change 'worked' – that it was sufficient – that results and the assumptions were all realized, and that the intervention was essential for the theory of change to bring about the results. This is exactly what contribution analysis aims to do (Mayne 2008). One can also use the theory of change to explore what role the intervention played, whether it was a trigger or whether it played some other role.

In summary, I have argued that many interventions are best thought of as contributory causes and in that sense they 'made a differ-

ence'. Further, that demonstrating contributory causes can be done using theory-based approaches such as contribution analysis.

References

Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation*, Oxford University Press.

Mayne, J. (2008). *Contribution Analysis: An Approach to Exploring Cause and Effect*, ILAC Brief No. 16: The Institutional Learning and Change Initiative. Available at http://www.cgiar-ilac.org/files/publications/briefs/ILAC_Brief16_Contribution_Analysis.pdf.

Mayne, J. (2012). "Contribution Analysis: Coming of Age?" *Evaluation* 18(3): 270-280.

Stern, E., N. Stame, J. Mayne, K. Forss, R. Davies and B. Befani (2012). *Broadening the Range of Designs and Methods for Impact Evaluations*, DFID Working Paper, 38. London: DFID. Available at http://www.dfid.gov.uk/R4D/Output/189575/Default.aspx.
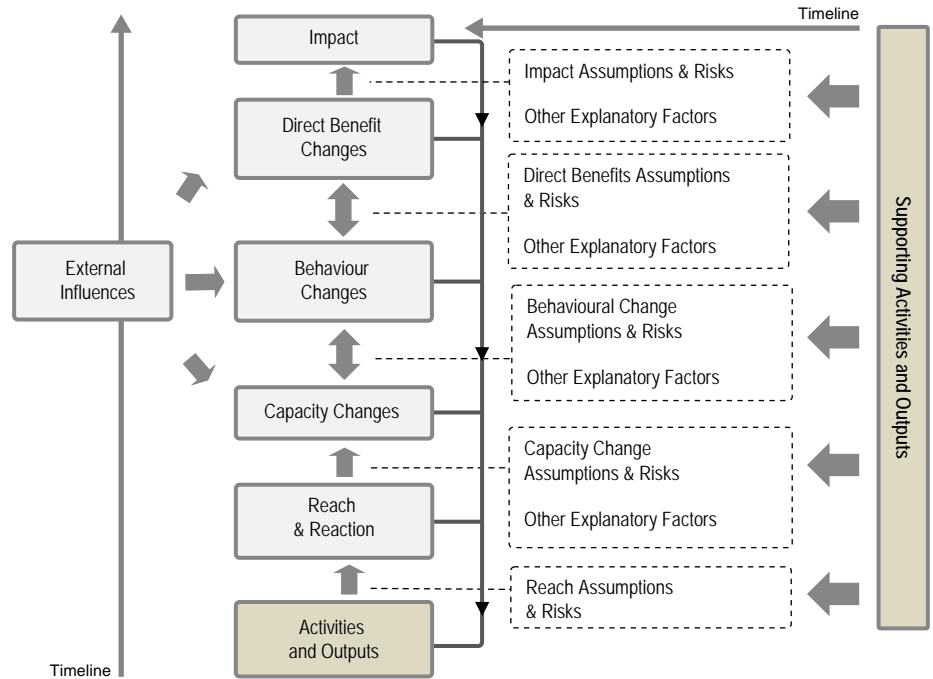
Figure 1 A Basic Theory of Change.

WHERE THERE IS NO SINGLE THEORY OF CHANGE: THE USEFULNESS OF DECISION TREE MODELS[1]

Rick Davies

Theories of Change (ToC) are in the limelight. In 2012 reviews of ToCs were commissioned by DFID, CARE, and Comic Relief and a host of training courses about the use of ToCs were put on offer. A ToC clarifies the programme logic, specifies expectations of outcomes and facilitates evaluation. But the ToC approach suffers from three major limitations. First, few ToCs adequately represent the complexity of reality. Second, multiple stakeholders' perspectives can be hard to reconcile in a single ToC. Third, even for relatively simple interventions the size of the combinatorial space within which a suitable design might be found escalates as the number of relevant design parameters rises. With as few as ten attributes (relevant to context, partners and procedures) there are 2^{10} or 1,024 possible best designs.

Large companies and research enterprises have responded to the scale and complexity of their work by developing an array of data mining tools designed to search for patterns, in the form of clusters and association rules (Siegel, 2013). These include algorithms designed to learn and to generalise from one data set to another. One type of algorithm widely used for such "predictive modelling" is the Decision Tree (DT). It summarises how different combinations of conditions are associated with different outcomes. Applied to an existing set of data it provides a summary classification. When applied to a new but comparable set of data it provides testable predictions about the outcomes associated with given conditions.

DT models have considerable potential as analytic tools in evaluation. First, they can represent *sophisticated forms of causality*, i.e. causal packages involving multiple conditions: a single DT model can contain multiple such packages. This is what Ragin (1989) calls "multiple and conjectural causation". This capacity enables what systems theorists describe as equifinality – multiple routes to the same end. Causal asymmetry can also be represented, recognising that the absence of an outcome is not always simply because of the absence of conditions required for its presence. There may be other inhibiting factors at work.

"Multiple and conjectural causation visibly differentiates sufficient, necessary and INUS causes (insufficient but necessary parts of a condition that is unnecessary but sufficient). Thus DTs provide the means

to differentiate attribution from contribution, in the sense used by Mayne (2012), and to go one step further and actually enumerate the extent to which a given condition is an influential contributory cause, relative to others in the model.

Despite their sophistication, a second argument favouring DT representations emphasises their *user friendliness*. This is a distinct advantage over regression models, expressed in mathematical notation or Qualitative Comparative Analysis (QCA) expressed in Boolean logic language. In DTs the outcomes of interest are represented by the "leaves" and the configuration of conditions associated with a given outcome are represented by the branches and sub-branches leading to that leaf.

Ease of understanding is important for communication to non-expert audiences. An example is shown in Box 1 below.

The third argument emphasizes *testability*. When a DT model is developed, it is standard practice to use only part of the data set available, commonly around 60% (randomly selected). At this stage the focus is on the adequacy of the DT as a descriptive model. Different measures of fitness can be applied, regardless of the specific contents of the model. One is the percentage of cases correctly classified within each leaf. Another is the ratio of cases correctly classified compared to their overall incidence, called "uplift". Still another is the simplicity of the tree.

Box 1: Decision Tree model based on household poverty data from Ha Tinh province of Vietnam in 2006

The simple Decision Tree shown below was generated by an analysis of a randomly selected 50% of 596 responses to a Basic Necessities Survey that asked about 23 aspects of people's households. Reading the tree from the top, we see that if a household has "a toilet built of stone" and they "eat meat once a week" then there is a 77% probability they will be non-poor. On the other hand, if a household has neither there is a 100% probability they will be poor. When this simple model was tested against the second half of the data set its overall accuracy was 82%.
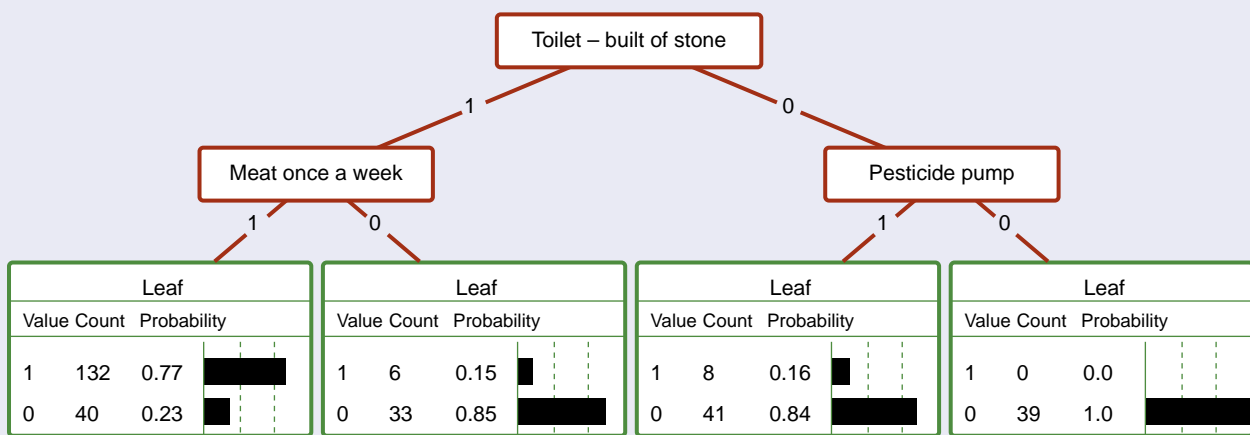
Although accuracy can always be improved by having more and more branches and sub-branches this more complex structure risks "over-fitting" the model and weakening its predictive validity when later tested against the remaining "test" data set. It is the latter measure that really matters.

The fourth argument is about *inter-operability* (the ability of diverse systems to work together). DT models are based on simply structured data sets, where cases are listed row by row and their attributes and associated outcomes of interest are listed column by column. The same kind of data set is used by QCA and is also usable by Social Network Analysis tools. Because QCA and Decision Tree algorithms use different methods of analysis but can be applied to the same data set, their combined use provides an opportunity for triangulation.

Related to the above argument, DT models are *scalable*. They can be used with very small or very large numbers of cases. For small

numbers DT models can be developed using ethnographic methods, as done by Gladwin (1989), and potentially with other more participatory methods. For large numbers there is a range of software packages available, including open source packages.

DT models are pre-eminently *predictive* models. Explanatory models can be seen as a sub-set of such models. They deliver testable and often accurate predictions. But not all predictive models need to be explanatory. A DT may identify poor households based on a predictive model, but it may not provide a good causal explanation of why people are poor. But where there is reasonable care with the selection of case attributes a predictive model can also provide a plausible causal explanations. The validation of the causal content of a model requires attention to the *mechanisms* that might link the associated attributes. These will come from close attention to the workings of individual cases in the data set through process tracing, including the identification of "smoking guns" and "hoop tests", as argued by Mahoney (2012).

Link = 0 = Households did not have the above attribute
Link = 1 = Households did have the above attribute

Leaf value = 0 = Household was poor (low BNS score)
Leaf value = 1 = Household was not poor (high BNS score)

References cited

Gladwin, C. H. 1989. Ethnographic Decision Tree Modelling. Vol. 19. Sage Publications, Incorporated. http://books.google.co.uk/books

Goertz, Gary, and James Mahoney. 2012. A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences. Princeton University Press.

Mahoney, J. 2012. The Logic of Process Tracing Tests in the Social Sciences. 1–28." Sociological Methods & Research XX(X) (March 2): 1–28. doi:10.1177/0049124112437709.

Mayne, John. 2012. "Contribution Analysis: Coming of Age?" Evaluation 18 (3) (July 1): 270–280. doi:10.1177/1356389012451663.Ragin, Charles C. 1989. The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies. University of California Press.

Siegel, Eric. 2013. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. John Wiley & Sons.

1 A more detailed exposition can be found online at http://mande.co.uk/2012/uncategorized/where-there-is-no-single-theory-of-change-the-uses-of-decision-tree-models/

EVALUATION PRACTICE

HOW DIGITAL DATA GATHERING CAN IMPROVE MONITORING AND EVALUATION PRACTICE

Kai Matturi

For development and humanitarian organizations, collection of meaningful data has become essential not just to achieve positive outcomes at the level of individual projects but to compete effectively in an increasingly crowded aid sector. Donors are increasingly demanding value for money and verifiable impact. They are not just insisting on results. They also expect aid projects to be implemented in the most efficient way. This is why Monitoring and Evaluation (M and E), long neglected, is gaining traction within the aid sector. It is now mandatory for programme implementers to generate baseline, mid-line and end-line data.

Collecting meaningful and timely data is easier said than done. Historically, the collection of programmatic data has involved paper-based questionnaires and inputting data into an information management system. More recently, various actors have been looking to the new world of Information and Communication Technology (ICT) to increase the efficiency, speed and accuracy of data collection. Thanks to ICT data can be collected using hand-held devices and seamlessly transferred through Digital Data Gathering (DDG) solutions, to a back-end server for storage and analysis. The term DDG refers to a plethora of electronic handheld devices such as Personal Digital Assistants (PDAs), smartphones or data pens that are used to record data in the field and transfer information back to a server.

In a bid to improve its quantitative survey data Concern Worldwide, hereafter called Concern has over the last two years utilised DDG. Concern is one of Ireland's leading development and humanitarian organisations, reaching more than 20 million people each year in more than 20 countries. The organisation's mission is to help people living in extreme poverty achieve major improvements in their lives which last and spread without on-going support from Concern. Concern has its headquarters in Dublin. The organisation's headquarters supports country offices through the provision of normative guidance and technical assistance, as well as through the activation of response systems.

The use of DDG within Concern involves Personal Digital Assistants to key in answers on forms displayed on the device. Electronic data is then uploaded onto a computer or server, or it can be directly sent to a database via the network. The rationale for using DDG solutions lies in increased data collection and utilization efficiency and improved data quality. The rest of this article explores the contributions that DDG has actually made to Concern's M and E practice.

Instantaneous Access

Digital solutions reduce the time between survey completion and programme managers' access to data. With manual data entry of paper survey results, it takes a minimum of four weeks from the actual survey to the delivery of a full and clean data set to the project team. Having access to accurate data in real-time accelerates evidence based decision making.

Improved Data Reliability

Improvement in error control increases the reliability and quality of data. Requests for monitoring data can be met instantly. Image capture enables visual confirmation of answers. Reduction in error at capture point and removal of the need for manual entry obviates the need for comprehensive data cleaning. The system that Concern uses enables inbuilt controls in electronic forms (e.g. to prevent enumerators skipping questions or control the range of values that can be inputted). This too reduces errors and improves data quality. Experience shows that the removal of the manual data entry phase eliminates the greatest source of error.

Reduction in Data Loss

Automatic uploading of data to a web server using active-sync, as well as storage of data offline until a signal can be found means that data leakage is greatly reduced. For example, the DDG devices are able to auto-sync every

five minutes. Global Positioning System (GPS) functionality militates against deliberate data fabrication by enumerators.

Centralised Data Management

DDG facilitates the development of a centralised data management system. A uniform solution built within a required format ensures conformity in how data is approved/rejected, analysed and formatted. This helps to ensure data integrity. A centralised data management system ensures that evaluators can readily access information and data rather than having to trawl through stacks of paper surveys. There are clear efficiency gains to be made here.

Improved Evaluative Practice

Because of time and budgetary constraints, evaluators often lack good data on which to base their analysis. With the advance of ICT tools, such as DDG it is now possible to collect large data sets in a remarkably short time. This ensures for instance that there is accurate end-line and baseline data which can be used to assess the impact of aid interventions. This type of practice ensures that evaluations are driven by timely and accurate data.

Rapid Data Analysis

The ability of the DDG user to undertake quick descriptive statistical analysis means that Concern's ability to respond to the needs of its target group in real-time is greatly enhanced. Rather than waiting for days or in some cases weeks before responding to a crisis with DDG this can be reduced to hours. In a humanitarian crisis this can literally often mean the difference between life and death.

In conclusion, perhaps the primary motivation of ICT enthusiasts is their appreciation of the potential of ICT innovation to contribute to the improvement of the human condition. But one must not remain blind to the perils of widely held deterministic and utopian expectation that ICT, by virtue of its technical properties, holds the key to development effectiveness. The benefits of DDG, (reduced data loss, centralised data management real-time data, rapid data analysis, and improved evaluation practice) simply make the management of humanitarian and development interventions more efficient and more adaptable to a rapidly changing operational environment.

MONITORING PUBLIC INVESTMENT: LESSONS FROM INDIA

I. C. Awasthi

A well designed Management Information System (MIS) should facilitate the flow of relevant and timely information to decision makers so as to track implementation progress and enhance the development impact of ongoing and future investment programmes. Massive public investments are envisaged by the Government of India (GOI) through different Ministries. Specifically, the equivalent of US$ 32 billion has been allocated to thirteen flagship programmes in 2012–13.

These initiatives aim at improved human well being through a wide range of public investments: rural roads, low cost housing, drinking water, irrigation, rural telephony, rural electrification, rural employment, primary education, child development, school feeding, rural health, urban renewal and sanitation, etc. Out of thirteen flagship programmes under operation, four do not have MIS (irrigation, rural telephony, school feeding and urban renewal) and in some cases one is not permitted to visit MIS website without a valid authorization (primary education and rural health, for instance).

Good MIS practice implies compliance with the following principles: (i) indicators derived from a well constructed log frame; (ii) periodic data reporting, authentication and validation; (iii) public accessibility of data, including a user-friendly website; (iv) data analysis and MIS utilization for decision-making. Actual practice has fallen short of these requirements:

- No explicit results chain (or theory of change) appears to underlie the choice of indicators in flagship programmes. This is despite the fact that GOI has promulgated performance monitoring and evaluation guidelines that task government departments to design and use results frameworks for all major programmes. Only by connecting to such results frameworks would the MIS provide the right signals to decision makers and facilitate the alignment of expenditure frameworks policy with programme goals.

- Data quality appears to be a weakness in all the flagship programmes. No provision for independent verification of data integrity and accuracy appears to have been incorporated in programme designs. This is an especially serious matter where skills gaps, weak administrative structures and poor internet connectivity hinder data collection from beneficiaries (e.g. in child development, irrigation, primary education and school feeding programmes) especially in some of the smaller states of the Union.

- Beneficiaries' involvement in the design of the MIS and the choice of indicators as well as in the generation of feedback about service quality is mostly missing.

- Data analysis, interpretation and timely reporting are frequently lacking. This hinders utilization of the MIS for decision making. Revealingly, evaluation reports do not use the data generated by most of the programmes.

- Unresolved inconsistencies in flagship programme data and official government statistics undermine the credibility of the MIS, especially in the child development scheme and national family health survey -3 (NFHS-2005-06) that give conflicting evidences.

Detailed comments on two major flagship programmes follow. The *Rural Road Programme* was launched in December, 2000 by the Ministry of Rural Development (MoRD). The program aims to provide better access to markets and social services through new roads and upgrading of existing roads so as to ensure connectivity to existing all-weather roads thereby generating increased agricultural incomes, productive employment and poverty reduction.

A web-based online system has been developed to facilitate monitoring, increase transparency and improve decision making geared to on-time delivery, cost management and quality control. Citizens have direct access to the data base.

The MIS suffers from the following short-comings:

- It largely focuses on inspection of contractors' compliance with engineering quality standards and fails to monitor and evaluate development effects.
- Most of the information provided is qualitative
- The judgments offered are not independently verified
- The MIS does not track overall implementation progress relative to original plans.
- No information is provided about decisions made in response to MIS findings

The *Mahatma Gandhi National Rural Employment Guarantee Act* (MGNREGA) aims at enhancing the sustainability of rural livelihoods by guaranteeing hundred days of wage employment annually to rural households whose adult members volunteer for unskilled manual work. Massive information is collected – from registration to implementation from over 0.25 million villages in 35 States & Union territories through a computerized information system.

The MIS portal provides a single-point access to all stakeholders. The aim is to generate transparent and user friendly data. The portal places complete transaction level data in the public domain for job cards, demand for work and attendance-cum-payment sheets for workers. Proper checks are used to validate the data through the biometric data base introduced by the Unique Identification Authority of India. This is expected to bring about more transparency, to curb irregularities and to reduce corruption.

While the system has considerable value, it could be improved as follows:

- The delivery monitoring unit (DMU) treats employment generation, assets and finances as 'outcomes' but does not capture impact (e.g. human well being indicators)
- Inconsistencies between MGNREGA information and data included in the Prime Minister's Office DMU data base need to be resolved.
- There appear to be excessive time-lags in the transmission of information at various administrative levels (village committees, blocks, districts, States and the GOI ministry) particularly, in small states/union territories.
- Lags in transmitting or updating data hinder effective use of MIS findings, e.g. the MGNREGA MIS system did not raise a timely alert regarding major delays in wage payments.
- The number of productive assets created and completed is reported but there is no information regarding their durability and sustainability.
- No analytical reports are being generated.

In conclusion, it emerges that compliance with sound principles is critical in order to justify the expense associated with MIS. A good MIS would deliver enormous benefits in terms of improved decision making and effective implementation. Among the shortcomings observed in the two programmes reviewed above are: (i) irregular reporting; (ii) lack of independent data validation processes; (iii) little analysis and inadequate data on interlinked activities, (iv) limited MIS utilization. This suggests more emphasis on reporting than on action. Since the two programmes reviewed above have the most comprehensive MIS of all it would appear that addressing the lacunae identified in this article would help improve the overall performance of infrastructure and social programmes. Given their size and ambition good MIS practice would make a major contribution to the achievement of India's development objectives.

CHALLENGES IN EVALUATING WELFARE SERVICES: A CASE OF ADULT SOCIAL WORK

Paula Saikkonen

This article ponders the impact of knowledge production and practices in the delivery of welfare services. Assessing the effectiveness of social interventions work brings out connections between welfare services and knowledge management. (Box 1)

The evaluation challenge lies in identifying and understanding multifaceted causal relations in social interventions since it is not unusual for several factors to influence the outcomes of such interventions. From this perspective adult social work is a rather typical example. Examining it holds the potential of illuminating the dynamics that are at work in municipal welfare services. In Finland, social services offered to adults are locally organized so that various models and practices can be observed in the field. Notwithstanding this diversity the core argument of this article is that knowledge production and practices have a significant effect on decision making and that they also impact on the social context within which welfare services are designed and delivered.

This article relies on case study documentation related to a project implemented by the National Institute for Health and Welfare (THL). The project was expected to assess -location-specific measures routinely used in adult social work. An online questionnaire (AVAIN) was tested in real life situations observed in 2011–2012 (Kivipelto & Karjalainen 2012). It quickly became evident however that the use of an online questionnaire had an effect on the quality of conversations between the social worker and the client. It drew attention to specific matters to be discussed and it helped to identify gaps in the quality and scope of the services on offer.

The topics addressed by the questionnaire included the goals, circumstantial factors and methods of social work. In a first phase this

induced social workers and clients to settle on joint goals for their interaction. In a second phase they looked back on the extent to which agreed goals were achieved and how the process went. Thus the questionnaire changed the nature of the relationship and it brought out issues and concerns which otherwise could easily have been left out. Thus, the very fact of evaluating the intervention in real time changed it and that had to acknowledge the interpretation of results.

Local decision making has a strong impact on policy administration, organization and resource allocation for welfare services (see left side in Figure 1). However, the overarching policy goals of social work as well as the control procedures that influence service delivery are set at the national level. Ideally the policy and control framework should be knowledge based but in practice there is no feedback loop that would allow local knowledge to impact on national welfare policies and services. (via a double learning loop).

Help from realist evaluation? Can realist evaluation help resolve the above dilemma? Its main object is to reduce the complexity of the real world by identifying causal mechanisms and recognizing the most influential factors that explain programme or policy outcomes. Realist evaluators ask "What works for whom in what circumstances and in what respects and how?" This approach reduces complexities by disentangling mechanisms, contexts and outcomes (van der Knaap et al. 2008, 50–51).

One useful concept of realist evaluation is generative causality which basically means that the aim of evaluation is to understand why programmes work in some conditions rather than others. The approach emphasizes the variation within programmes, not just between different programmes (Pedersen & Rieper 2008, 271–272, 276). In the case of social work, the realist evaluator scrutinizes contexts and mechanisms in tandem and relates their characteristics to the observed outcomes. These are influenced by client's dispositions (e.g. values, beliefs and attitudes) and resources (e.g. information, skills and material resources).

For instance, if a client is not willing to cooperate with a social worker it is hardly possible to achieve outcomes. This means that welfare outcomes are the joint product of distinct agents acting on the demand and supply side of service delivery. In turn, knowledge practices interact with knowledge production and therefore with outcomes at the level of the individual (Figure 1).

On the other hand, the influence of knowledge production and practices at the aggregate programme level is unclear since no mechanism is in place to provide meta-feedback to decision makers. This means that evaluation use is constrained to the local level and that programme evaluation feedback processes should be introduces and that realist evaluation methods should be put to work for social betterment (Henry & Mark 2003).

References

Henry, T. G. and Mark, M. M. (2003). Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation* 24(3), 293–314.

Kivipelto, M. and Karjalainen, P.(2012). *Adult social work effectiveness evaluation.* Paper presentation at Evidence Informed Practice. April 22nd/25th, Cavan, Ireland.

van der Knaap, L., Leeuw, F. L., Bogaerts, S. and Nijssen, L. T. J. (2008). *Combining Campbell Standards and the Realist Evaluation Approach: The Best of Two Worlds?* American Journal of Evaluation 29(1), pp. 48–57.

Pawson, R. and Tilley, N. (2004). *Realist Evaluation.* In http://www.communitymatters.com.au/RE_chapter.pdf (09/04/2013)

Pawson, R. and Tilley, N. (1997). *Realistic Evaluation.* London: SAGE publication.

Pedersen, L. H. and Rieper, O (2008). Is Realist Evaluation a Realistic Approach for Complex Reforms? *Evaluation* 14(3), pp. 271–293.

Wagenaar, H. and Cook, N.S.D. (2003). *Understanding policy practices: action, dialectic and deliberation.* In M. A. Hajer and H. Wagenaar (eds.) *Deliberative Policy Analysis. Understanding Governance in the Network Society* (pp. 139–172). Cambridge: University Press.
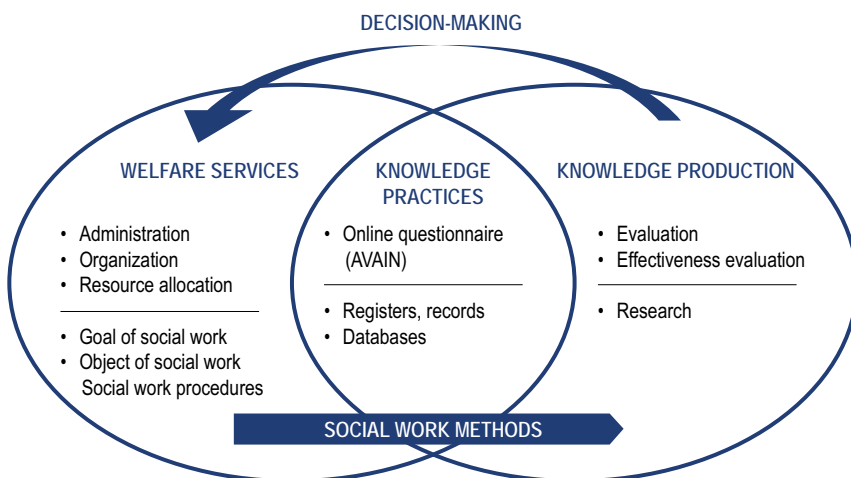
Figure 1: Knowledge production and knowledge practices in welfare services.

MONITORING, EVALUATION AND INFORMATION SYSTEMS

Giorgio Garau, Enrico Garau and Lucia Schirru

Use of "data" to make well-informed decisions and shape judicious policies is at an all time high. This has led to greater attention to the design of monitoring systems as tools for tracking policy implementation and retrospective assessments of the policies themselves. From this perspective, a statistical information system (SIS)[1] can facilitate due attention to policy goals, respect for stakeholders' objectives and the generation of basic information needed for *ex post* evaluation (Garau, Mandras e Schirru 2011). System design is also shaped by the programme logic (Mazzeo 2012).

How is the effectiveness of a policy to be assessed? What kind of data should be collected to allow the conduct of a good ex post evaluation? How is compliance with the applicable legislation assessed? In principle such questions should be tackled when policies and programmes are designed and relevant data collected during implementation. In fact, this does not always happen which leads to a lack of basic information for ex post evaluation. In turn this forces evaluators to use inadequate administrative data and resort to *ad hoc* surveys. The SIS approach helps to overcome this problem.

This article examines the potential of SIS in the context of the Sardinia Employment Plan (EP) which aims to address youth unemployment that, in 2011, attained 42% in the region, compared to a national youth unemployment rate of 29%. The EP's purpose is to improve the employability of young people via "on the job training".

The programme is not intended to lead directly to job creation: there is no expectation that trainees will invariably secure a regular or even a temporary job following the training period. Thus EP is not an instrument of job placement but rather a way of enhancing workers' competencies. It is also an instrument of industrial relations. Each trainee is sponsored by a labour union, guild or association (for blue collar workers) or by a professional order (for white collar workers).

In the particular instance of EP it allowed us to map the data needed for tracking progress and to perform a good summative evaluation of the effectiveness of communication among the main programme actors. The first step in SIS design is the definition of an "*Observer*" so as to delineate appropriate boundaries for the SIS. In this case we adopted an independent evaluator and learning perspective.

The next step is a *Requirements Analysis*, which consists in the collection of relevant documents (laws, rules, etc.) that define the policy to be evaluated. Such analysis facilitated the identification of relevant actors as well as the target population: young EU residents resident in Sardinia, unemployed but skilled. Such youth are eligible to apply to the EP which leads to their enrolment and inclusion in the data base of the REA.

An unemployment assessment is delivered by the Public Job Center (PJC), another actor of the SIS. Further actors are private enterprises, professional cabinets or professional orders and educational establishments that have concluded in advance a special agreement[2] with REA.

Conceptual Modelling

Conceptual modelling helps in policy design as well as in policy evaluation. It requires identification of all the agents involved and a specification of the relationships among them. In the EP case the agents (public and private institutions) are jointly responsible both for the policy and for its achievement. They are also beneficiaries of the policy interventions. If the policy is correctly formulated and the data collected is well specified, the interaction among agents produces sound and effective contractual relations and it creates databases that are suitable for policy evaluation.

On the other hand, if the policy is poorly conceived and badly written the resulting relationships generate inadequate, incomplete or redundant data, i.e. data that is duplicated among multiple stores and maintained by different agents. The task of conceptual modelling is to help policy makers identify the right agents and the correct relations among them.
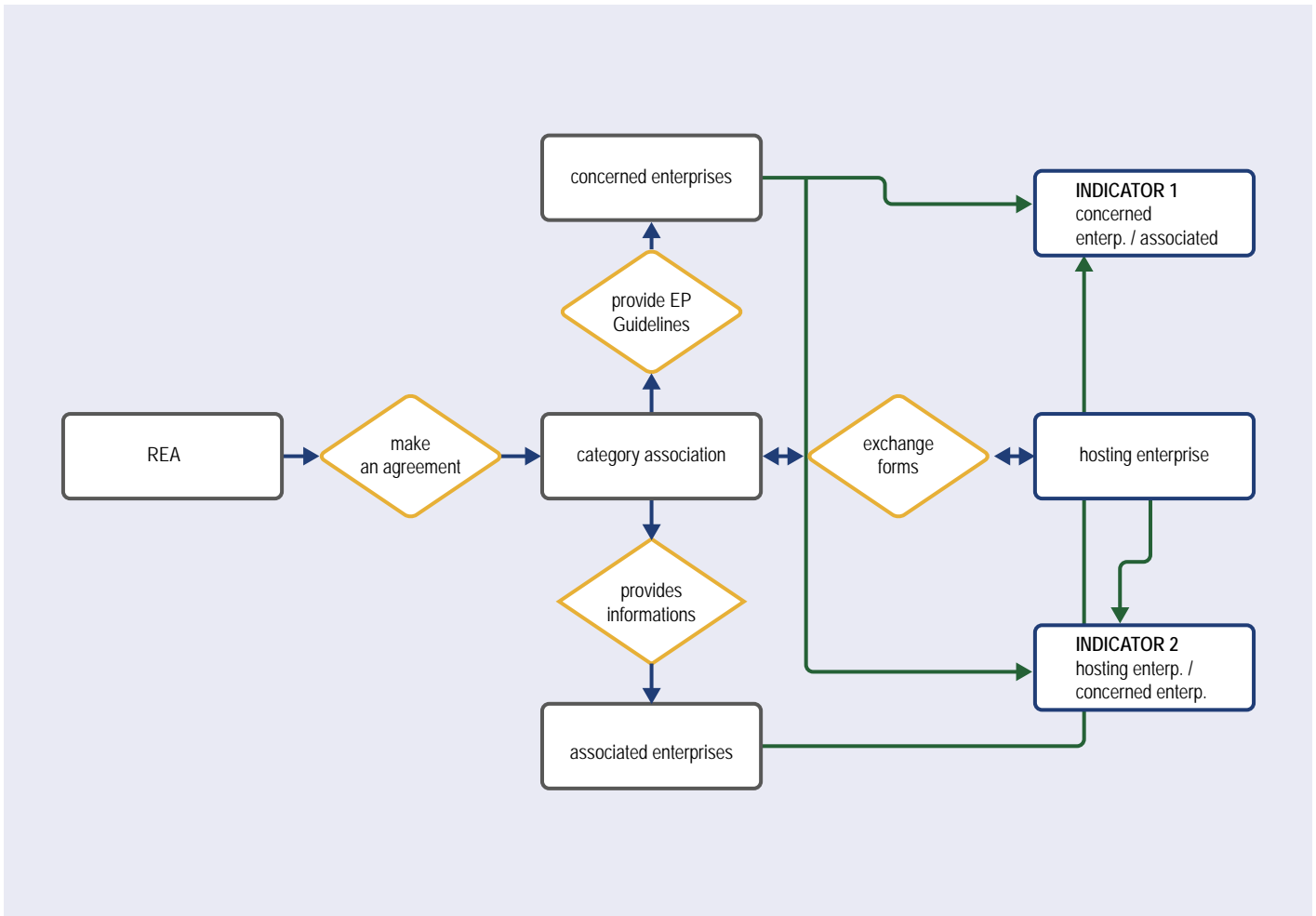
Other benefits of conceptual modelling include placing the policy within the wider context of similar policies (those having the same goal or complementary goals), in order to ensure that their objectives are consistent and that working together that they may be easier to achieve. With such systematic modelling one can appreciate whether the results observed depend on the policy adopted or require other policies operating in concert.

In the conceptual modelling phase below we show in a schematic form the relationship among EP actors previously identified by the documentation analysis. When evaluating the policy following its implementation it was of course necessary to supplement the monitoring information with *ad hoc* interviews of EP actors and agents, e.g. to verify that procedures were performed correctly and to conduct survey designed to ascertain that the monitoring information is accurate and timely.

Next, the graph below models the agreement among REA and *category associations*. At the centre the category association acts as the intermediary among enterprises and REA. The agreement states that associations have a promotional role and that enterprises must have the status of an *associated* enterprise to secure relevant information.

Some enterprises have a stake in the EP but they opt not to be formally *associated* with it. Could they benefit from EP? What promotion actions are expected to be performed by the category associations? Are these activities traced? How many enterprises have evinced an interest in EP?

Such typical questions can be answered using ad hoc interviews. Category associations are the counterparty of Trade Unions in collective bargaining. They are associations of employers, grouping enterprises belonging to a specific

Agreement between REA and Category Association – some indicators.

economic sector (commerce, manufacturing, services). They supply the format of information requests to hosting enterprises regarding participants to the EP program.

Associated enterprises are enterprises belonging to a Category Association.

Concerned enterprises are a subset of associated enterprises that have the intention of hosting EP program participants.

Hosting enterprises are a subset of concerned enterprises made up of enterprises that actually host EP program participants.

The above scheme makes evident critical factors of implementation that contribute to the achievement of the policy goal: the improvement of young skilled employability. Such mapping facilitates the identification of useful performance indicators. For example, the ratios between concerned and associated enterprises (and the ratio of hosting to concerned enterprises) emerge as relevant effectiveness indicators of communication for category associations.

The SIS also pinpoints obstacles to effective implementation. For example, distinct agreements or form designs may yield different communications effectiveness indicators and hence different employment outcomes. It is of course necessary to integrate the SIS data with other qualitative information in carrying out an ex-post evaluation.

1. Garau G., Mandras G e L Schirru: A Statistical Information System Supporting Environmental Policies. *Environmental Engineering and Management Journal*, vol. 12; Sept 2011.
2. Mazzeo Rinaldi F.: *Il monitoraggio per la valutazione*, Franco Angeli, Milano, 2012.

1 A Statistical Information System (SIS) is characterized by the use of aggregated data, where the use of the information produces statistical knowledge. It also devotes special attention to meta-data and meta-information.

2 This special agreement links three actors (REA, an intermediary – e.g. a category association – and hosting enterprises) and tasks category association to inform and oversee hosting enterprises during the training period.

INSTITUTIONALIZATION OF MONITORING AS A REFLEXIVE MANAGERIAL PRACTICE

Elizabeth Moreira dos Santos, Egléubia Andrade de Oliveira, Marly Marques da Cruz, Aline Leal Gonçalves, Aline Duque de Macedo e Carlos Leonardo Figueiredo Cunha

Introduction

This article is a reflection on the monitoring processes embedded in a health action programme – the Programa Ação Saúde (PAS) – which lies at the interface between the primary care services delivered by the unified health system – the Sistema Único de Saúde (SUS) – and local communities in Brazil'surban and rural spaces of Maranhão. The PAS has established itself as a pedagogic project inspired by the "problematization" evaluation model.

The PAS initiative is aligned with Ministry of Health policy priorities and its design characteristics are consistent with the technical skills training and mobilization goals pursued in ENSP's public health forums. Its overarching mother and child theme secured the full support of a network of experts which also endorsed reduction of infant mortality as a key PAS objective using a judicious indicator.

On the other hand, some municipalities in Maranhão required special mobilization and persuasion efforts in order to secure local ownership of programme goals. Accordingly PAS began by sensitizing and mobilizing local communities in support of the integrated management and implementation of PAS maternal and child health interventions at the intersection of health promotion and communication and social mobilization activities. The approach was designed and implemented by a technical partnership involving the Laboratory of Monitoring and Evaluation of Regional Endemic Situations, a unit within the National School of Public Health (LASER/ENSP/FIOCRUZ) and Canal Futura, a television channel committed to social action and public education with the financial help from the private sector notably the Roberto Marinho Foundation and the Vale Foundation, a corporate social responsibility arm of a large mining corporation.

The intervention: a short description

The innovative approach used by ENSP was grounded in Freire's principles of empowerment and autonomy (1996) and implemented through tailor made socio-technical networks as proposed by Latour (2000). This methodology was designed to facilitate connections and negotiation of agendas among the actors and networks involved. Specifically, the programme was geared to "supporting public management towards dialogue and integrated action with the civil society". This presumes pluralistic action, respect for different groups' values and working within the municipalities regardless of party affiliation, ethnic racial or religious background with sensitivity to the local culture and full use of local knowledge.

The challenge was to identify a set of choices that would take full account of the local context. Through reflexive and pedagogical processes the monitoring system encouraged popular participation and collective action to respond flexibly and effectively to health protection needs and promote improved human livelihoods. The methodology implied a focus on the organization of health networks, individual interest in contributing to integrated practice, readiness to participate in a training process and full engagement with the goal of replication and programme expansion. The basic network unit was a locally staffed "health promotion cell" fully dedicated to maternal and child health protection and sharply targeted to local needs.

The monitoring system addressed five major components: 1. Mobilization; 2. Network strengthening; 3. Training; 4. Sustainable governance; and 5. Innovation. It involved 45 indicators. Two thirds of them tracked implementation and delivery processes and one third focused on results. The systematic collection of information was facilitated by data sheets for each indicator designed by

communities of practice. The logistics of data collection included on-site inspection, distance supervision and on-going interaction about the interpretation of monitoring data.

Adaptation characterized the entire approach, The set up of local cells began with a sensitization workshop and adjustments were made continuously throughout the implementation cycle. The process was innovative insofar as a piloting phase did not precede programme expansion. After the first sensitization event and the planning/training workshop replication was launched with support of instructors through a second cycle of sensitization (Figure I). Thus, timely replication constituted the core of the problematization approach. Each cycle ended with a workshop geared to learning from experience and reshaping the implementation process. Thus practice informed the process of adaptation and expansion through real time feedback.

This system differs from traditional approaches focused on accreditation and accountability. Consequently the reflexive and pedagogical characteristics of the monitoring system were not adopted without controversy among experts. Monitoring SMART objectives before replication and mainstreaming contrasts with the problematization approach that uses monitoring as a reflexive and locally situated practice that allows real time social learning, adjustment and with innovations.

Monitoring as a reflexive practice

Interventions can be conceived as socio networks, i.e. in fluid terms (flows, circulations, alliances and movement) rather than as entities. Social actors are agents rather than passive components of a fixed system. They interact and adjust through connections and negotiations (Latour, 2000; Latour, 1996). In democratic societies translating connections between expert networks and
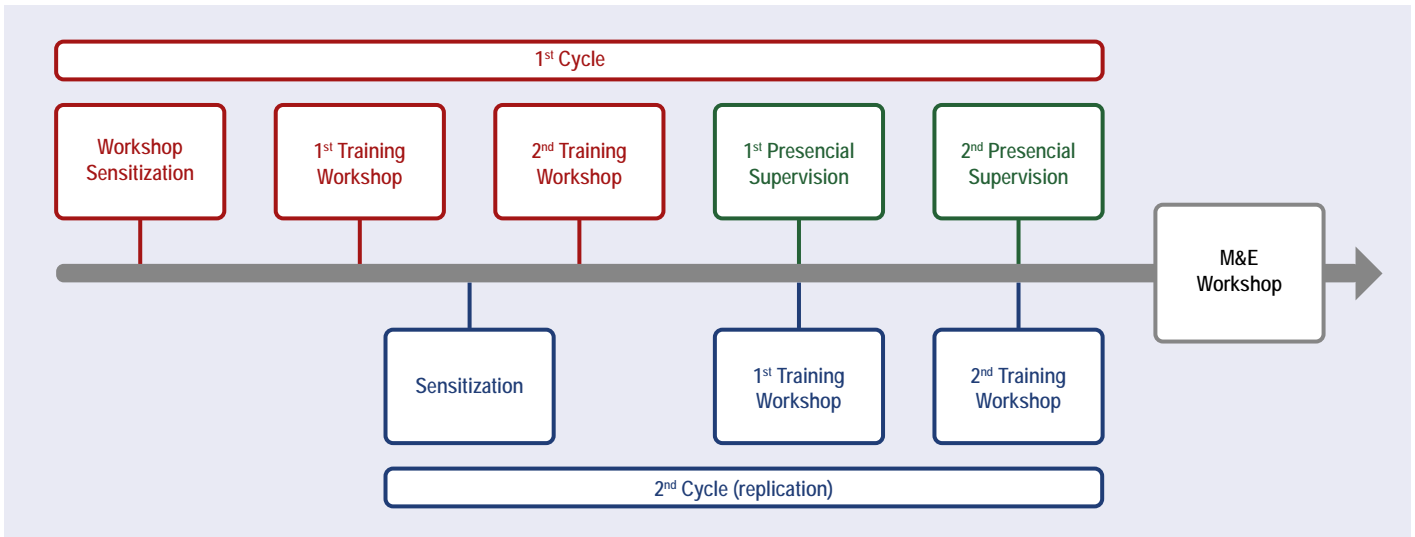
Figure 1: Monitoring training: Replication Cycle.

communities of practice requires mobilization, mediation, conflict resolution, negotiation, appreciation of alternative knowledge insights, adaptive technical solutions and effective use of new technologies.

References

Freire, P. *Pedagogia da autonomia*. Rio de Janeiro: Paz e Terra, 1996.

Hartz Z., Matida Á. H., Denis J.L., Santos, E.M. From knowledge to action: challenges and opportunities for increasing the use of evaluation in health promotion policies and practices (capítulo) IN: Potvin L., Mc-

Queen D.V., editors. Health promotion evaluation practices in the Americas: values and research. Springer Verlag Publishing, 2008.

Latour, B. "Networks, Societies, Spheres: Reflections of an Actor-network Theorist". *Keynote speech for the International Seminar On Network Theory: Network Multidimensionality In The Digital Age, 19th February 2010 Annenberg School for Communication and Journalism, 2010* disponível em http://www.bruno-latour.fr/sites/default/files/121-CASTELLS-GB.pdf, acessado em 12/08/2012.

Latour, B. On actor-network theory: a few clarifications. *Center for Social Theory and*

Technology (CSTT) Keele University, UK. 1998. http://www.nettime.org/Lists-Archives/nettime-1-9801/msg00019.html.

Latour, B. On actor-network theory: a few clarifications plus more than a few complications. *CSI-Paris/Science Studies-San Diego*. 1996. http://www.nettime.org/Lists-Archives/nettime-1-9801/msg00019.html.

Santos, EM, Oliveira EA, Cruz, M, Leal, A, Duque, A, Cunha, L. Monitoramento em Promoção de Saúde: uma experiência de capacitação de adultos em um estado do nordeste brasileiro. Submetido à *EDUCAR Revista*. UFPR, Parana, 2013.

FEEDBACK ON THE 2012 EES BIENNIAL CONFERENCE IN HELSINKI

Claudine Voyadzis[1]

The evaluation of the last biennial conference held in Helsinki in 2012 *"Evaluation in the Networked Society. New Concepts, New Challenges, New Solutions"* brought out the following findings:

- The conference attracted a large number of evaluators, paper presenters and panellists who provided high-level insights on new evaluation trends and on endeavours to adapt the field to and seize the opportunity of advanced information technology

- A total of 634 delegates attended the conference; 456 abstracts had been submitted (for paper presentations, panels and posters) of which some had to be improved on request by abstract reviewers; 5% were rejected
- Thanks to the generous support of Finland and other donor countries, 75 bursaries were offered to developing countries' evaluators.
- EES bestowed awards on 6 professional evaluators and 2 students for the most no-

table contributions to evaluation research, methods and practices.

Largely positive evaluation feedback was received from delegates: 91% of evaluation conference respondents considered that the knowledge and insights gained from the conference was "very valuable" or "valuable" (against "marginally valuable" or "not valuable") while the usefulness of contacts made at the conference was appreciated by 86% of respondents.

The objectives of the programme were considered clear and coherent by 70% of respondents, relevant for work by 82% while 83% of respondents found that their expectations had been reached. The choice of topics was viewed from "excellent" to "very good" to "good" by 87% of respondents while the high quality of keynote speakers was praised by 82% and the quality of panels by 73%.

The *conference administration* – registration, IT support and logistics – were very highly rated. On the other hand, the rating of meals and coffee breaks gave a mixed picture due inter alia to a scarcity of vegetarian options.

Strengths and weaknesses were identified through open-ended questions (229 respondents). The main strengths are the diversity of the participants in terms of country of origin and fields of interest (38 respondents); the great opportunity for networking (31 respondents); the broad range of topics (28 respondents); and the excellent quality of speakers and presentations (24 respondents).

The main weaknesses include concerns about the uneven quality of speakers and presentation content (24 respondents); incoherence of presentations within sessions, (14 respondents); lack of prior information about abstracts (6 respondents); limited coverage of evaluation within a European context and overlapping of topics with conferences addressing international development (5 respondents).

Data on the *pre-conference workshops* indicate that a majority of respondents, 92%, rated the overall impression of the workshops from "excellent" to "good", and only 5% found them "marginal" or "poor"; over 88% of respondents found the goals of the workshops clear and coherent, and 89% found them relevant; 80% of respondents declared that their expectations were met against 17% of respondents whose expectations were not met. The trainers' skills and competencies were very highly rated. Their knowledge and ability to convey clearly their messages was well appreciated by respondents. However a few respondents (less than 17%) thought that the workshop sessions were not sufficiently practical and would not have an impact on their work.

This feedback will be taken into account in planning the Dublin Conference.

---

1 Three students (Marie Gildemyn, Barbara Mineo and, Kettil Nordesjo) contributed to the evaluation.

THE AUTHORS

I.C. Awasthi

I.C. Awasthi is a senior faculty at the Institute of Applied Manpower Research (Planning Commission, Government of India) Delhi. He works in the areas of labour, employment, human resource development and development evaluation. He is a Ph.D. (Economics) from Jawaharlal Nehru University, New Delhi. He has received specialized training on International Programme for Development Evaluation (IPDET) conducted by Independent Evaluation Group of the World Bank and Carleton University. He is a teaching faculty in a three months course on Monitoring & Evaluation organized by the Institute for overseas working professionals under the aegis of the Indian Technical and Economic Cooperation (ITEC), Ministry of External Affairs, Government of India. He has published widely in the areas of employment, livelihoods and labour markets. He is governing council member of a web based initiative (IOCOM), the International Organization for Collaborative Outcome Management (www.iocomsa.org).

María Bustelo

María Bustelo, PhD in Political Science, is Associate Professor of Political Science and Public Administration at the Complutense University of Madrid (UCM), where she teaches on Public Policies and Evaluation. She is the director of the Master on Evaluation of Programmes and Public Policies (UCM) since its commencements in 2002. President of the European Evaluation Society (2012–13), she has also been Member of the Board of Director Committee of the Spanish National Agency for the Evaluation of Public Policies (2007–2011). At the research level she has been the leader at the UCM of several National and European research projects on the quality of gender equality policies, the most recent one GENOVATE (*Transforming organisational culture for gender equality in research and innovation –2013–2016, FP7*) where she and her team act as evaluators of the project. She has a number of publications on evaluation theory and methodology as well as on gender equality policies.

Carlos Leonardo Figueiredo Cunha

Graduated in Nursing at the Federal University of Maranhão (2003). Specialization in Family Health, Training and Educational Planning and a Master of Maternal and Child Health from the Federal University of Maranhão (2010). He is currently PhD candidate in Public Health-IESC / UFRJ. He works as a Research collaborator at the Laboratory of Evaluation (LASER) and has experience in teaching Public Health with emphasis in Health Management and Policy of the Brazilian Primary Care System.

Rick Davies

Dr Rick Davies is an independent Monitoring and Evaluation consultant, based in Cambridge, UK. He has worked extensively in Africa and Asia over the last 25 years with a range of international development aid organisations (multilaterals, bilaterals and INGOs). He also manages the MandE NEWS website and the associated email list, which

has a global membership. He is interested in the development of innovatory methods through the cross-breeding of ideas from different realms of knowledge, including participatory approaches to impact monitoring and poverty measurement, the representation of complex systems in terms of networks, and configurational views of causal processes.

Marly Marques da Cruz
PhD in Public Health (2006) from the National School of Public Health/FIOCRUZ where she is an associated researcher. Currently she is the head of the Department of Endemic Diseases Samuel Pessoa (DENSP/ENSP) and Coordinator of the Laboratory of Evaluation at the National School of Public Health (LASER). She also is the Coordinator of the Master program in Health Evaluation. Her work involves teaching, research and technical cooperation in Brazil, Latin America and African Portuguese speaking countries.

Aline Duque de Macedo
Graduated in Psychology UERJ (2004). He began his research career at the Instituto Fernandes Figueira (IFF / FIOCRUZ) in Mother and Child Health linking stigma and chronic illness in childhood. She has a Diploma Course in Communication and Health at ICICT / FIOCRUZ. She is has a Master's in Public Health in the area of Policy, Planning and Management in Health from the Institute of Social Medicine, State University of Rio de Janeiro . She has experience in Distance learning in monitoring and evaluation.

Egléubia Andrade de Oliveira
PhD in Public Health from the National School of Public Health Sergio Arouca, Oswaldo Cruz Foundation in 2007, Master in Social Work from the School of Social Service, Federal University of Rio de Janeiro (1997) and Graduate in Social Work from the Federal University of Pernambuco (1986). She is a researcher at the Institute of Collective Health Studies – IESC at the Federal University of Rio de Janeiro and a collaborator researcher of the Laboratory Evaluation (LASER/ENSP). She has a strong experience in social and community mobilization.

Elizabeth Moreira dos Santos
MD (1974); MSc (1986) and PhD in Community Health – from University of Illinois-Urbana Urbana Champaign (1995) is retired as a full researcher at the National Scholl of Public Health (ENSP/FIOCRUZ) in Rio de Janeiro, Brazil. She has coordinated and implemented several national evaluations of endemic diseases programs in Brazil and African countries. Her academic experience involves evaluation course development (Brazil, Angola and Ethiopia) including the first Health Evaluation Master Program in Brazil; advising of around 70 thesis (55) and dissertations (16) and publishing 9 book chapters and 20 peer-reviewed journal articles. Currently she is active as independent consultant in the areas of evaluation course development, planning and implementation of M&E systems and evaluations at national and international levels.

Aline Leal Gonçalves
Graduated in Nutrition from the Federal University of the State of Rio de Janeiro, MSc from the National School of Public Health Sergio Arouca (ENSP / FIOCRUZ), acts as a research collaborator at the Laboratory of Evaluation (LASER) in the area of Program evaluation. Her teaching experience includes distance learning in Health Management, at the Federal University of the State of Rio de Janeiro.

Giorgio Garau
Giorgio Garau is full professor of Economic Statistics at the University of Sassari. He obtained his Ph.D at the University of Geneva in 1993 and he was research fellow of the Japan Foundation in 1994.

Enrico Garau
Enrico Garau work at REA, Regional Employment Agency, and is responsible of the technical assistance to the job services regional system.

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.

Kai Matturi
Kai Matturi works as a Knowledge and Learning Adviser with the Irish international non-governmental organisation, Concern Worldwide, where he supports the design, development and implementation of high quality organisational learning and knowledge management functions, processes and systems. He was involved in the initial piloting of DDG in Concern. His work also involves strengthening the use of evaluations as a key driver of learning within Concern. Kai has previously worked with University College Dublin, the Government of Uganda and the United Nations Development Programme. He is a member of the European Evaluation Society, the American Evaluation Association and the Irish Evaluation Network.

John Mayne
Dr. Mayne is an independent advisor on public sector performance. He has been working with a number of government, NGO and international organizations in various jurisdictions, on results management, evaluation and accountability issues. He has authored numerous articles and reports, including several on contribution analysis, and co-edited seven books. In 2006, he was made a Canadian Evaluation Society Fellow.

Robert Picciotto
Robert ('Bob') Picciotto, (UK) Professor, Kings College (London) was Director General of the World Bank's Independent Evaluation Group from 1992 to 2002. He previously served as Vice President, Corporate Planning and Budgeting and Director, Projects in three of the World Bank's Regions. He currently sits on the United Kingdom Evaluation Society Council and the European Evaluation Society's board. He serves as senior evaluation adviser to the International Fund for Agricultural Development and the Global Environment Fund. He is also a member of the International Advisory Committee on Development Impact which reports to the Secretary of State for International Development of the United Kingdom.

Martin Reynolds

Martin Reynolds is a Senior Lecturer in Systems Thinking at The Open University, UK. As well as being active in research and providing consultancy workshop support for critical systems thinking in continual professional development, Martin continues to produce distance learning resources for postgraduate programmes on International Development, Environmental Decision Making, and Systems Thinking in Practice. He has published widely in these fields – many of his publications are available on open access.

Jim Rugh

Jim Rugh has been professionally involved in international development for almost 5 decades; as a specialist in evaluation for 33 years. For 12 of those years he served as head of Evaluation CARE International. From 2008–2011 he served as the AEA Representative to the IOCE. He currently serves as Coordinator of the EvalPartners Initiative that aims to strengthen evaluation capacities of and through Voluntary Organizations for Professional Evaluation (VOPEs) around the world. Jim co-authored the popular and practical *RealWorld Evaluation* book. In recognition of his contributions to the evaluation profession he was awarded the 2010 Alva and Gunnar Myrdal Practice Award by AEA.

Paula Saikkonen

M.Soc.Sc Paula Saikkonen works in FinSoc team at National Institute of Health and Welfare. In her work, she enhances knowledge of evaluation methodology and promotes the use of evaluative information in social services. At the moment, she is finalizing her doctoral thesis which concerns the role of knowledge in the municipal decision making and the local environmental problems.

Murray Saunders

Professor Murray Saunders: Co-Director of the HERE (Higher Education Evaluation and Research) Centre, Department of Educational Research, Lancaster University, UK) and Professor of Evaluation in Education and Work, I have acted as a consultant to, and undertaken a wide range of evaluation projects for, the British Council, DfES [Department for Education and Skills], DFID [Department for International Development], ESRC [The Education and Social Research Council], HEFCE [Higher Education Funding Council], the UN and a variety of regional agencies. I teach on PhD programmes within the Department in the areas of evaluation, connections between learning and working and policy and social change. I have carried out evaluation and research projects in a wide range of cultural contexts, including in Asia: China, Japan, Singapore and India; in Africa: Tanzania, Kenya, Nigeria, Ghana and Ethiopia; in Latin America: Mexico and Chile: In the Middle East: Jordan and Egypt as well as other member states of the EU and Russia. I enjoy promoting and developing evaluation practice as past president and Council member of the UK Evaluation Society, Board member and immediate past president of the European Evaluation Society, chair of the co-ordinating committee for the establishment of the IOCE [International Organisation for Cooperation in Evaluation] and its current vice-president. Most recently, I am on the executive committee of EvalPartners. I am associate editor of the only European international multidisciplinary academic journal in the field – Evaluation.

Lucia Schirru

Lucia Schirru is Data research analyst at the L'Unione Sarda's Research Center. She obtained the Master in applied statistics

in 2004 and she specializes in statistical information system for policy evaluation.

Claudine Voyadzis

Claudine Voyadzis has spent some 20 years working in the evaluation field. Her work consisted mainly of appraisal and monitoring of projects and programmes, impact studies and ex post evaluations in West and East Europe and in Africa. She earned her Ph.D. in the United States in Demography and Sociology and has a dual experience of Anglo-Saxon and European approaches to evaluation. After working in the Independent Evaluation Group of the World Bank in Washington, she joined the Council of Europe Development Bank in Paris as technical adviser for project appraisal and monitoring. Subsequently she was appointed to set up the Ex Post Evaluation Department and remained for six years as Director of the Department. Claudine Voyadzis works now as an independent consultant, and leads evaluations for multilateral organisations and consulting firms.

Bob Williams

Bob Williams has been interested in the application of systems concepts into evaluation for many years. He has edited two books, run many workshops and written a couple of dozen articles on this topic. Originally from the UK, he is now based in New Zealand but most of his work is in Europe, North America and South East Asia.