



EVALUATIONS CONNECTIONS

Robert Picciotto Editor. Ole Winckler Andersen Associate Editor. Eva Petrová Editorial Assistant.

EES

EUROPEAN
EVALUATION
SOCIETY

JUNE 2015

ARTICLES – NEWS – EVENTS

CONTENT

New Evaluation Approaches and Tools for the Post 2015 Era: An Editorial	1
Rigour (-mortis) in Evaluation	2
Economic Methods and Social Policy	4
Uncertainty and Adaptation: New Trends in Policy Analysis and Implications for Evaluation	6
Emerging Opportunities: How New ICT Technologies Can Help Promote Equity-Focused Evaluation Practice	7
Using 'Big Data' for Equity-Focused Evaluation – Understanding and Utilizing the Dynamics of Data Ecosystems	9
Big Data, Bigger Problems: How Big Data Can Help Development Evaluation Methodology in Connected Markets	10
Billionaire to the Rescue: Does the Gates Foundation Have a Sound Evaluation Policy?	12
Engaging Parliamentarians Towards a Better Enabling Environment for Evaluation	13
The Authors	15

NEW EVALUATION APPROACHES AND TOOLS FOR THE POST 2015 ERA: AN EDITORIAL

Robert Picciotto

Under the aegis of EvalYear evaluators are once again debating the future of their discipline. Many challenges face the evaluation community in a world characterized by social turmoil and technological change: the advent of chronic volatility in society calls for transformation in evaluation approaches and tools.

Organizations that do not adapt perish. The same goes for disciplines. In the leading article of this issue Martin Reynolds evokes the tensions between evaluation commissioners and evaluators caused by the instability of the operating environment. In a vain search for certainty some commissioners insist on rigid compliance with evaluation approaches and methods that are no longer fit for purpose. By contrast, successful leaders seek innovative evaluation approaches and tools that help their organizations adjust.

In order to demonstrate their utility evaluators will have to take a leap towards a new generation of methods and tools. For example in a climate of austerity state of the art evaluation can help decision makers obtain "more for less" as Rory Tierney and Jacqueline Mallender argue in the second article of this Connections issue. Specifically in complex interventions cost benefit analysis in combination with decision tree mapping

and probabilistic simulation can assist in the choice among diverse intervention logics so as to secure the desired social benefits at minimum cost.

In the same vein Leon Hermans explores the implications for evaluation of novel planning approaches being tested in the Netherlands water sector. Under conditions of uncertainty assigning probabilities to events that may affect policy is not feasible. In such cases state of the art Dutch planning focuses on choices among alternative adaptation pathways that resemble the schematics of subway-station-maps. Equally evaluators will increasingly have to shed the linearity assumptions of theory based approaches and think of "out of the box" in order to meet the needs of adaptive management.

The new information technologies present additional challenges. The next three articles of this Newsletter address the promises and limitations of Big Data for the evaluation discipline. First, Michael Bamberger and Linda Raftree outline the potential of Big Data towards overcoming the information asymmetries that have long hampered equity oriented evaluations. They visualize the revival of participatory evaluation methods through social networking and crowd-sourcing. They

also note that fascination with the new instruments should not displace basic principles of sound evaluation design.

Second, Kim Forss and Jonas Norén warn that tapping the full potential of Big Data calls for explicit support by evaluation commissioners, capacity development and training as well as a commitment to transparency by the companies generating and storing data. Third, Kara Chiuchiarelli stresses that Big Data is no panacea given the digital divide; selection biases; privacy concerns and the lure of

meaningless correlations among explanatory factors. Evidently Big Data needs small data secured from interviews and surveys.

The two closing articles stress the crucial importance of the policy environment. Both are highly relevant to the on-going EvalYear debate about the democratic future of evaluation. Linette Lim's contribution focuses on the rising influence of private foundations in public policy characteristic of the post 2015 era. She recommends major upgrading of the Gates Foundation's evaluation policies

and practices. Finally Liisa Horelli's article sketches a bracing vision of evaluation transparency, accountability and independence at the national level in response to greater parliamentary involvement.

No one can foresee all the disturbances and shocks that will affect evaluation practice in the years ahead. But one thing is clear: for evaluation to continue functioning as a valuable instrument of adaptive management in society it will have to be managed adaptively as well. More than ever evaluation needs evaluation.

RIGOUR (-MORTIS) IN EVALUATION¹

Martin Reynolds

Evaluation-in-practice can be regarded as a confluence of interactions between three broad idealized sets of stakeholders – the evaluand, evaluators and commissioners of evaluations. Elsewhere I have suggested two contrasting manifestations in which these interactions might be expressed; one as an 'evaluation-industrial complex' (similar in form to the 'military-industrial complex' originally used by Dwight Eisenhower in 1961) and another as a more benign 'evaluation-adaptive complex' (Reynolds, 2015).

Building on the idea of an iron triangle that empowers the military-industrial complex, I represented the relationships of evaluation-in-practice as a triadic interplay involving six activities that influence the evaluation process (Fig. 1). Here I focus on only one of the six activities – commissioning – and I summarize what it might look like for an evaluation-adaptive complex.

One of the key influences illustrated in Figure 1 is the relationship between 'commissioners'

and 'evaluators' and the associated requirements for assurance of trustworthiness. Collectively such assurances aim to guarantee rigour. In an evaluation-industrial complex scenario, assurances of rigour are frequently experienced as stifling and rigid leading to an unhelpful and malign confluence that I call *rigour-mortis* resulting in an inability to support radical and transformative interventions. Conversely I contrast this option with more benign forms of rigour, i.e. a set of co-guarantor attributes – or CoGs. In so doing I draw on traditions of American pragmatism (cf. William James and John Dewey), critical social theory (particularly Jürgen Habermas), and critical systems thinking (particularly C. West Churchman and Werner Ulrich).

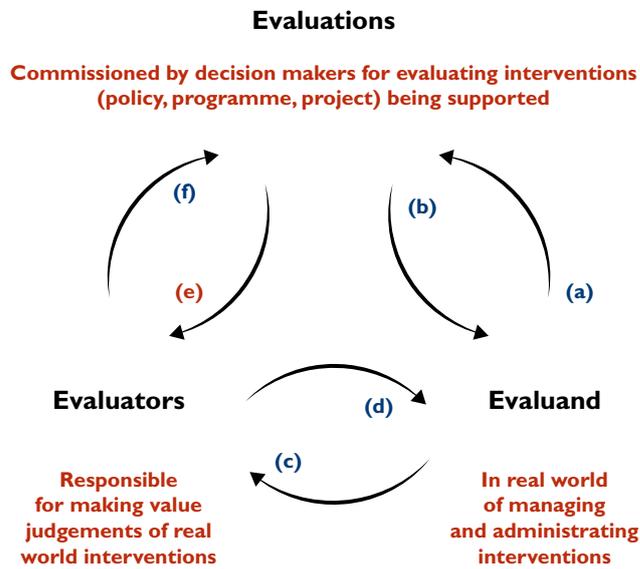
Each of these traditions is based on Immanuel Kant's fundamental idea that the existence of an absolute guarantee of certainty is a fundamentally flawed notion. Rather rigour lies in the assurance of expert support given to decision-makers (e.g. commissioners/funders) in evaluating interventions. Specifically evalu-

ators act as guarantors of successful implementation of plans/interventions (projects, programmes, policy). However this can turn into a source of deception where the guarantee is worthless or false given the inherent limits associated with evaluation inquiry. In any intervention there is always a built-in risk about the value of the evaluative guarantee.

The search for a more robust albeit provisional set of co-guarantors springs from Habermas' distinction between three knowledge constitutive interests (Habermas, 1972; 2014): (i) *technical* interest in prediction and control of natural and social affairs; (ii) *practical* interest in fostering mutual understanding; and (iii) an *emancipatory* interest in being free from coercion. Table 1 reconfigures these interests in terms of three sets of CoGs for rigour in evaluation – objectivity in making appropriate representation of the evaluand, complementarity in resonating with alternative representations amongst evaluators, and responsibility in making transparent whose purpose is being served in the evaluation.

¹ Adapted from 2014 EES conference paper entitled "From rigour-mortis to systemic triangulation: ethics and competences in evaluating complexity".

Figure 1: Evaluation-in-practice involving six key activities
(adapted from Reynolds, 2015 p. 75).



Six activities:

- (a) auditing
- (b) planning
- (c) evaluating (summative)
- (d) evaluating (formative)
- (e) Commissioning ... need for "rigour"
- (f) learning

False guarantors exist where either the criteria of rigour are not appropriately fulfilled or where one set of CoGs is privileged over the other two. So for example, a false guarantor of objectivity may be manifest when, say, a randomized control trial (RCT) might be inappropriately used according to the *scientific* disciplinary guidelines of use, such as in circumstances where control experiments are not feasible and/or ethical. A false guarantor might also be apparent when an RCT is used as sole guarantor of an evaluation with associated claims of abiding by 'best practice', through being 'scientifically' objective and neutral. Such arguments and critiques against the dominant use of evidence-based evaluations such as RCTs for evaluating social interventions have an increasingly impressive and effective tradition in the field of evaluation (cf. Patton, 2010; Pawson et al., 2011; Rogers, 2008).

Arising from this critique of *best practice* emerges an alternative notion of *best fit* based on the contingency approach to evaluation. An expression of contingency is the demarcation between simple, complicated, and complex interventions (Glouberman and Zimmerman, 2002); an approach recently critiqued by Mowles (2014) and Reynolds (2015). The simple-complicated-complex idea is a very helpful heuristic for understanding systemic failure of interventions where inevitable complex situations of an evaluand are misconceived as either complicated or simple. The idea is less useful as a heuristic for rigour in prescribing evaluation 'tools' for predefined situations.

Table 1: Co-guarantor attributes of rigour
(adapted from Reynolds, 2001 and 2003).

Co-Guarantor attributes (CoGs) of Rigour	Features of each set of CoGs
CoG1 Objectivity (technical interest)	<ol style="list-style-type: none"> 1. intra/multidisciplinary 2. based on criteria of reliability 3. inviting disciplinary responsibility in representing the 'real world'
CoG2 Complementarity (practical interest)	<ol style="list-style-type: none"> 1. interdisciplinary 2. based on criteria of resonance 3. inviting general academic critique in valuing different representations
CoG3 (social and ecological) Responsibility (emancipatory interest)	<ol style="list-style-type: none"> 1. transdisciplinary 2. based on criteria of relevance 3. inviting social and ecological critique in making transparent the wider purpose of support

Table 2 sketches a few expressions of CoGs in relation to the two archetypal forms of evaluation-in-practice. Some features of the contingency approach are used here to illustrate some features of rigour in the evaluation-industrial complex. They are intended to invite conversation on other features of rigour that may also inhibit a shift to an evaluation-adaptive complex.

The shift towards a more radical type of rigour involves (i) humility about the built-in fallibility of any evaluation tools, (ii) empathy and openness with alternative forms of evaluation, and (iii) innovative political practice in regarding evaluation as purposeful systems design.

Table 2: Some co-guarantor attributes (CoGs) of rigour compared.

CoGs	Evaluation-industrial complex (for example, false guarantors associated with contingency approach)	Evaluation-adaptive complex (towards critical systems thinking approach towards developing CoGs)
1. Objectivity (and reliability)	Reality can be regarded objectively as simple, complicated, or complex.	Reality regarded as ‘unknowable’; an integral mixture of complicated, complex, and conflictual... need for systemic inquiry.
2. Complementarity (and resonance)	Pursuit of an ever-increasing ‘toolbox’ driven by multiple-method ethos, where discrete tools are deemed ‘fit’ for discrete situations.	Cultivation and adaptation of existing methods and approaches through social learning using others’ expertise and experiences... need for pragmatism.
3. Responsibility (and relevance)	Legitimacy given by terms of reference for evaluation and perceived fixed bias of different tools regarded as being relevant for different purposes.	Legitimacy given by evaluators’ ‘political’ role in sustaining or challenging purposes of evaluation... need for radical constructivism.

References

- Glouberman, S. and B. Zimmerman (2002). *Complicated and complex systems: what would successful reform of Medicare look like?*, Commission on the Future of Health Care in Canada Toronto.
- Habermas, J. (1972). *Knowledge and Human Interests*. London, Heinemann.
- Habermas, J. (2014). *Truth and justification*. London. John Wiley & Sons.
- Mowles, C. (2014). “Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship.” *Evaluation* 20 (2): 160–175.
- Patton, M. Q. (2010). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, Guilford Press.
- Pawson, R., et al. (2011). “Known knowns, known unknowns, unknown unknowns: the predicament of evidence-based policy.” *American Journal of Evaluation* 32 (4): 518–546.
- Reynolds, M. (2001). *Co-Guarantor Attributes: A Systemic Approach to Evaluating Expert Support*. Eighth European Conference on Information Technology Evaluation: Conference Proceedings, Oriel College, Oxford.
- Reynolds, M. (2003). “Towards Systemic Evaluation: A Framework of Co-guarantor Attributes”. *Evaluating Regional Sustainable Development* submission to Workshop of the EU Thematic Network project REGIONET Evaluation methods and tools for regional sustainable development, University of Manchester (UK). 11 – 13 June 2003.
- Reynolds, M. (2015). (Breaking) The Iron Triangle of Evaluation. *IDS Bulletin*, 46 (1), 71–86. <http://oro.open.ac.uk/41879/>.
- Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, 14(1), 29–48.

ECONOMIC METHODS AND SOCIAL POLICY

Rory Tierney and Jacqueline Mallender

Policymakers have to make choices as to which goods and services to pay for with limited budgets. Any time a question is asked along the lines of “what do I get for the money?” a form of economic analysis is being conducted to answer it.

Cost-effectiveness analysis (CEA) is a useful technique employed by economists to understand ‘value per Euro spent’. ‘Cost’ is simply the cost of implementing a policy, but it goes beyond financial, “cashable” costs. Opportunity cost is included as well (i.e. what could

have been done with the resources allocated to a programme – people’s time, building facilities, etc.).

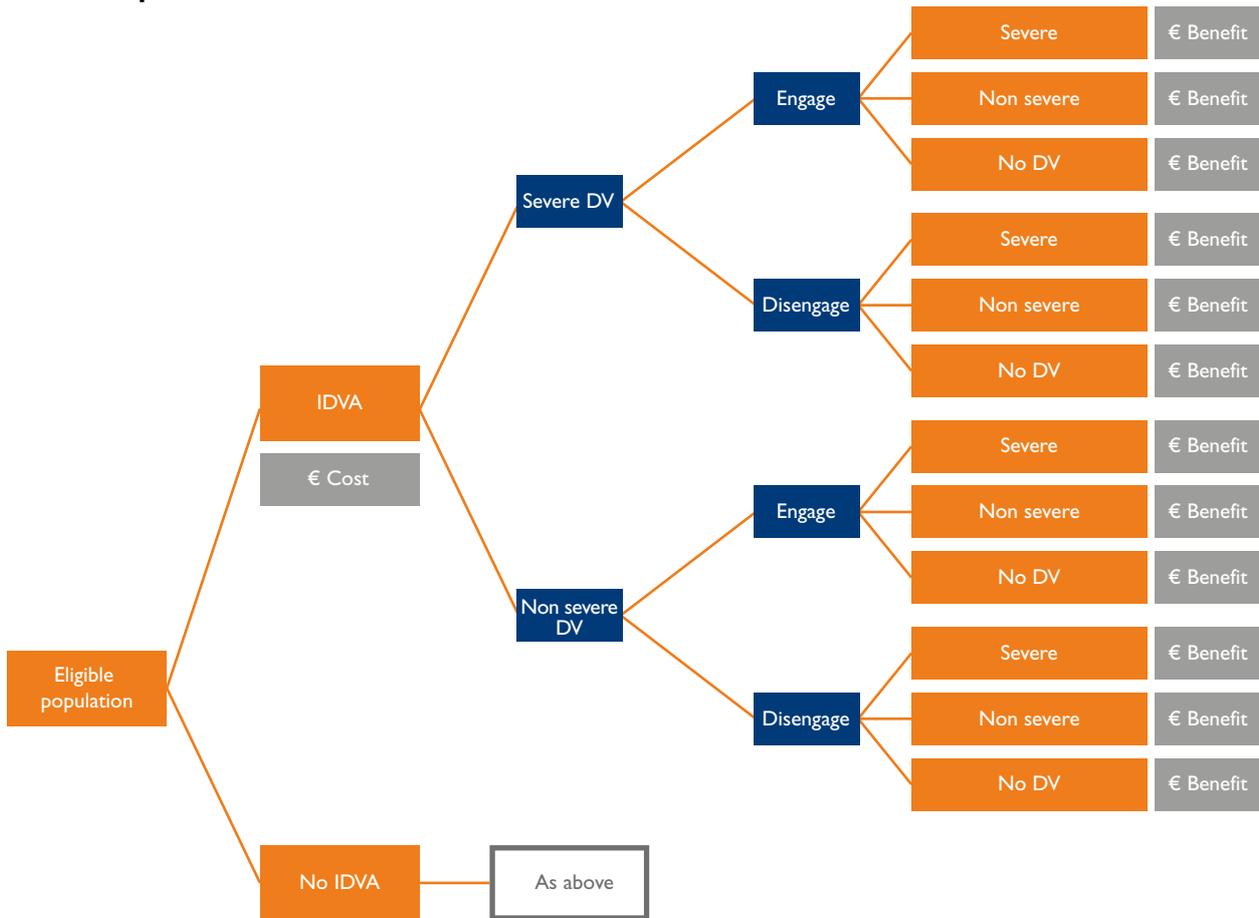
‘Effectiveness’ refers to how well a policy achieves its goals. For CEA, it is measured in natural units – for instance, the number of offences prevented or kilometres of road built. Assigning monetary valuations to effectiveness measures (including intangible costs such as pain and frustration) is known as cost-benefit analysis (CBA) and while more complex, with more data requirements, pro-

vides a single benefit metric and thus greater comparison with other policies.

Both cost and effectiveness must be measured in comparison to a ‘counterfactual’, i.e. what would have happened without the programme (or what is likely to happen if it isn’t introduced), so that only the impact caused by the programme is measured.

When conducting a CEA or CBA, the first step is to review the available evidence. Systematic reviews of primary studies can pro-

Figure 1: Example of a decision model.⁴



vide a synthesis of data on the effectiveness of a given programme. However, this data often have their limitations: in many fields, reliable evidence is scarce. Equally context matters – a programme that works well in one geography or for one population may not achieve the same results in another.

Overcoming this data gap requires building a framework with which one can draw on multiple data sources to conduct a rigorous analysis applicable to the population of interest. To this end, decision modeling is growing¹. Decision modeling combines data from multiple sources, as parameters, into a mathematical relationship between inputs, outputs and outcomes. A conceptual example is shown below (figure 1: Example of a decision model).

The diagram below shows a simple decision tree used for an analysis of IDVAs (Independent Domestic Violence Advisors), who “represent a service for victims of domestic

violence who are at high risk of homicide or serious harm.” The support they provide “ranges from help with social services, the criminal justice system and immigration issues to gaining access to counseling and GP [General Practitioner] services.”² Calculations are made for both the ‘intervention’ arm (IDVA) and the ‘no intervention’ (counterfactual) arm, and the differences examined to understand the incremental impact of the intervention.

Each of the lines represent probabilities: the probability of an individual being classified as experiencing severe domestic violence, the probability of an individual engaging with support services available to them, and then the outcome probabilities (does the individual experience severe, non-severe or no domestic violence at the end of the process). By primary data on the intervention, and combining this with data from other sources on, for example, the likelihood of engaging with services without the presence

of IDVAs, the model can be populated and the outcomes assessed. Calculating the cost of the intervention, and subtracting the cost of the counterfactual (in this case, ‘no intervention’, so the cost is zero) gives the incremental cost of the intervention and allows an assessment of cost-effectiveness. Assigning monetary values to the outcomes (i.e. the damage caused by domestic violence: psychological damage, health problems, criminal justice system costs etc.), from sources such as the British Crime Survey, allows a full cost-benefit analysis to be conducted.

Collected data can be defined broadly as ‘scientific evidence’ and ‘colloquial evidence’.³ Scientific evidence “is explicit (codified and propositional), systemic (uses transparent and explicit methods for codifying) and replicable (using the same methods with the same samples will lead to the same results).” It can be “context-free” (i.e. applying universally) or “context-sensitive” (i.e. related to “specific real-life circumstances”). Colloquial

evidence, on the other hand, is derived from expert testimony and stakeholder opinion and as such is subjective and value-driven.

In order to populate a locally relevant model it may be necessary to draw on all of these types of data, using colloquial evidence to make assumptions on the transferability of context-sensitive scientific evidence, for example. A well-constructed decision model should be clear and transparent, showing where parameters were sourced, and which assumptions and calculations were made. Sensitivity analysis – varying uncertain parameters and assessing the impact on the final result – should be undertaken in order to demonstrate the degree of confidence with which the final result can be stated.

A comprehensive economic analysis can be a useful tool in the drive towards evidence-based policymaking, and there are many examples of their use ‘in the real world’. Much

of this comes from the health field, where high quality experimental primary data is frequently generated and economic analysis is becoming more common. For example, in England and Wales the National Institute for Health and Care Excellence (NICE) is mandated by the Department of Health to issue guidance on best practice in public health interventions (among other things), and use economic analysis in order to do so.

In other fields, the use of CEA and CBA is also growing. In criminal justice, for example, a number of organisations in both Europe and the United States conduct analyses of alternatives to custody, looking at the impact on re-offending versus short-term prison sentences. Crucially, while some programmes – such as drug rehabilitation – may be expensive, the use of decision models can demonstrate the enormous savings they can achieve (financial and non-financial) if they are effective (reduce crime).

Of course, it can be difficult to put a logical modeling framework around some of the effects of social policy, especially when these are complex and dynamic. Techniques exist for modeling complex systems, such as Markov chains or discrete event simulation, but data can be a limitation. Where the data gap is significant enough to prevent a full CEA from being conducted, cost-consequence analysis – where costs of the intervention are compared against a ‘balance sheet’ of both quantified and descriptive consequences – can still provide a useful tool.

In all areas of public policy, collecting primary evidence is an important step in understanding what represents value for money. Yet evidence is always incomplete, and context matters. Decision modeling – making implicit decisions and collation of data sources explicit – represents a powerful tool for conducting cost-effectiveness analysis to demonstrate the impact of a spending decision.

References

- 1 Shemilt I, Mugford M, Vale L, Marsh K, Donaldson C. (eds.). Evidence-Based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice. Oxford: Wiley-Blackwell.
- 2 Matrix (2013). Economic analysis of interventions to reduce incidence and harm of domestic violence. Report for the National Institute for Health and Care Excellence. Accessed 05.02.15 from <http://www.nice.org.uk/guidance/ph50/evidence/economic-analysis-domestic-violence-final-report-for-consultation2>.
- 3 National Institute for Health and Care Excellence. (2012). Methods for the development of NICE public health guidance (third edition). Accessed 05.02.15 from <http://www.nice.org.uk/article/pmg4/chapter/determining-the-evidence-for-review-and-consideration>.
- 4 Adapted from Matrix (2013). Economic analysis of interventions to reduce incidence and harm of domestic violence. Report for the National Institute for Health and Care Excellence. Accessed 05.02.15 from <http://www.nice.org.uk/guidance/ph50/evidence/economic-analysis-domestic-violence-final-report-for-consultation2>.

UNCERTAINTY AND ADAPTATION: NEW TRENDS IN POLICY ANALYSIS AND IMPLICATIONS FOR EVALUATION

Leon Hermans

Evaluation and ex-ante policy analysis have much in common. Both inform planners and decision-makers and both seek to combine rigour and relevance in their methods. Within the planning cycle, evaluations and ex-ante policy analyses differ mostly in their focus and positioning: before or after decisions are made. Consequently progress in one discipline may well hold lessons for the other. In particular, emerging approaches to

address uncertainties and adaptation within the planning domain may well have relevance for the evaluation community.

There has been a surge of interest about decision-making under uncertainty and adaptive management among policy analysts and planners. These topics have always been important but they have recently been popularized in best-selling books by Nicholas Taleb

and Daniel Kahneman. Equally in the domain of practice new analytical approaches have been put to work in policy work and programme planning.

So what is new in planning?

In presence of deep uncertainty traditional analytic approaches do not hold. Assigning probabilities to events that may impact on

policy effectiveness is not feasible: in such contexts events cannot be foreseen. All one can do is identify key potential uncertainties that may impact on strategic choices. Such uncertainties may then be linked to key factors or mechanisms within a conceptual theory or model. If these factors develop in unexpected ways the strategy may lose its effectiveness. Hence, monitoring of signpost variables is critically important. When the values of these variables exceed certain threshold values a judicious response should be triggered: a pre-defined adaptation or mitigation action may be needed or a more fundamental rethinking of the original planning logic may be called for.

This core idea informs planning under uncertainty. Specifically it suggests resort to adaptation pathways. If the effectiveness of a plan hinges on potential but unknown developments planners are well advised to take such uncertainties into account. For instance, in water management, one could make large investments in large-scale dike reinforcements to address potential but unknown climate events thus reaping efficiency benefits from economies of scale in construction. On the other hand, it may turn out that the reinforcements are either insufficient or unnecessary. In such cases the investment is wasted. Consequently the focus of planning turns to choices among alternative adaptation pathways that resemble the schematics of a subway-station-map. This is quite different from the conventional linear tables and diagrams of traditional planning documents.

Is this for real?

These new planning concepts are actually being used to improve water policy and delta management in the Netherlands. The Delta Programme established some ten years ago embodies a long-term strategy designed to ensure water security and flood protection. One of its hallmarks is its use of adaptation pathways. Thinking is currently ongoing about how to set up a monitoring and evaluation system for adaptive delta management as an integral part of the Netherlands' National Adaptation Strategy.

Dutch policy thinking in this area has been influenced by work carried out by the RAND Corporation in the United States. Similar notions have been used to support decision making about the Thames Estuary in the United Kingdom. The approach is likely to be replicated in other policy domains in the Netherlands and beyond.

Will this change evaluation practice?

Adaptive management concepts are perfectly compatible with emerging evaluation thinking. The contemporary evaluation literature has begun to explore the implications of complexity theory. State of the art evaluation practices increasingly resort to systems thinking. Developmental evaluation methods and state of the art theory based models are currently all the rage in evaluation practice. Just in time evaluation through exploitation use of modern information technologies is being pio-

neered. All this fits in well with the signposts and triggers of adaptive management.

Looking ahead we should not simply revert to our traditional instruments. Can we, as evaluators, think of "out of the box" systems beyond those we have grown accustomed to? Where learning is central and collective impacts are measured by different actors shouldn't new decision logics be adopted? If we look seriously at the implications of uncertainty on evaluation subtle differences may well translate into complicated evaluation arrangements as well as entirely new operational procedures and actions. Adaptive management constitutes a great challenge and a great opportunity for our profession.

References

Haasnoot, M., J. Kwakkel, W. Walker, J. ter Maat (2013) Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change* 23, 485–498.

Walker, W. E., Rahman, S. A., and Cave, J. (2001) Adaptive policies, policy analysis, and policymaking. *European Journal of Operational Research* 128(2), 282–289.

Delta Programme 2015. Working on the Delta. The decisions to keep the Netherlands safe and liveable. Staff of the Delta Programme Commissioner. Ministry of Infrastructure and the Environment and Ministry of Economic Affairs, The Hague, the Netherlands.

EMERGING OPPORTUNITIES: HOW NEW ICT TECHNOLOGIES CAN HELP PROMOTE EQUITY-FOCUSED EVALUATION PRACTICE¹

Michael Bamberger and Linda Raftree

Common challenges for evaluators

Evaluations face a number of common challenges when it comes to data collection and analysis. These include: the high costs of data collection, difficulties to study processes of project implementation and behavioral change, collecting and analyzing qualitative

data and mapping distribution of population groups, resources or problems. It can also be difficult to collect data on all of the economic, political, socio-cultural, ecological and other kinds of contextual factors that influence program implementation and outcomes.

Finally, as development programs become increasingly complex, conventional data collec-

tion methods do not allow for collection and analysis of all of the new kinds of information required. Equity-focused and rights-based evaluations pose additional data challenges, such as: giving voice to vulnerable groups and accessing difficult to reach groups; studying domestic, civilian and military violence; studying power relations and control of resources; and the analysis of intra-household resource allocation.

How new information technologies can help address these challenges

The field of information and communication technology has exploded, bringing with it new applications for program monitoring and evaluation. Mobile phones and other held-held devices such as tablets are widely used for data collection, analysis and dissemination – particularly for program monitoring but increasingly for program evaluation. GPS technology is rapidly evolving and permits the incorporation of mapping and tracking technologies. M&E applications of the Internet are also expanding through, for example, internet-based surveys, on-line participatory consultation and planning, theory of change software, and expert consultation methods such as concept mapping. Finally, evaluators are only just becoming aware of the huge potential of big data for the collection and analysis of kinds of data that even a few years ago most evaluators could not have imagined.

The widespread use of mobile devices, particularly by large NGOs, illustrates the great potential of these new technologies. Mobile phones already provide huge cost and time savings when it comes to data collection and analysis. Other areas where ICTs are being used include real-time participant feedback on programs and incident reporting combined with geo-location (for example to report electoral fraud or highlight dangerous areas in a community). There are also new applications for health monitoring such as estimating anthropometric measurements from a photograph or capturing vital measurements through the phone. Two examples can illustrate the potential of ICTs. The first concerns quality control of surveys. It is possible through the use of GPS to check whether the interviewing is actually conducted at the correct location (and not fabricated in the local café). Internal consistency of responses can be checked to reduce errors, and audio functions allow a supervisor to remotely listen-in on how an interview is being conducted.

A second example illustrates the potential application of big data for monitoring poverty levels in a rural area in Africa. The following are five sources of data, each of which could be collected very economically and each of which could provide an independent indicator of short-term increases in poverty levels: (i) reduced purchase of air-time for mobile phones, (ii) draw-down of on-line savings accounts, (iii) reduced phone and on-line orders for seeds and fertilizers, (iv) increased use of words like “hungry” and “sick” on social media, and (v) satellite photos reveal reduced number of lorries travel to the local market. If all of from these estimates indicate consistent trends this would provide an effective early warning system of increased poverty.

Applications to equity-focused evaluation

As discussed above, equity-focused evaluations face particular challenges in data collection and analysis. The following are some of the ways that ICTs can help. First, big data can be used to monitor references to gender, social exclusion or hostility to certain minorities on social media. Social media can be studied before-and-after an intervention such as an on-line magazine designed to empower teenage girls to identify changes in gender-related dialog. Of course, the validity of this approach only works if there is a high level of access to social media for this population. Second, on-line software can be used for participatory consultations and planning, for example developing theories of change or methods such as most-significant change or outcome mapping. Third, the scale of participatory consultations can be expanded through crowd-sourcing so that thousands of people or communities can be consulted. Fourth, feedback through mobile phones can rapidly identify emergency needs following a natural disaster or civil unrest. Finally, software is being developed for the collection and analysis of qualitative data so that this can be incorporated into a mixed methods design.

Just don't forget to apply basic evaluation principles!

As with any new technology, there is a risk that some evaluators will become so fascinated with the speed and economy of data collection (and all of the exciting new apps!) that they overlook the basic principles of evaluation design. A first issue is the possibility that the fascination with large numbers and ‘big data’ eclipses the issue of selection bias. People with access to mobile phones tend to be better-off, more educated, living in urban areas, often younger and a higher proportion is male. Obtaining information from large numbers of people in this group may not provide reliable information on the situation of poor, less educated, older women living in more remote areas.

A second issue relates to the reliability of information received. In households where a phone is shared or where it is used in the presence of other household members, women, young people or the elderly may not reflect their true opinions or situation when speaking on the phone. Third, it is easier to collect numerical data so there is the risk of narrative data being reduced to simple numbers. Fourth, construct validity can be a problem since there may be a tendency to reduce complex constructs such as vulnerability, gender equality of welfare to a few simple (but easily measurable) indicators.

Fifth and finally, a weakness of many evaluations is their failure to identify unintended outcomes of development programs, and it is important to ensure that mobile technology does not also overlook important unintended outcomes (such as for example, increased domestic violence when men feel threatened by the freedom of communication that women may gain through mobile phones). Evaluators should also be aware that low institutional capacity and resistance to change may reduce the effectiveness of mobile technology in many organizations. Also major privacy and security issues arise through mobile technology, and these are

| This paper summarizes a presentation at the 2014 EES conference that was based on Linda Raftree and Michael Bamberger (2014) *Emerging Opportunities: Monitoring and Evaluation in a Tech-Enabled World* with financial support from the Evaluation Office of the Rockefeller Foundation through a grant to ITAD. The full report is available on-line from the Rockefeller Foundation (<http://www.rockefellerfoundation.org/blog/emerging-opportunities-monitoring>).

not currently being well addressed by many organizations.

Conclusion

Though there are huge benefits and great potential for M&E with the introduction

of some ICTs, these tools, devices and approaches need to be considered through a rigorous lens of evaluation principles and practice. Though newer technologies may resolve current operational challenges, they may also lead to new, lesser-understood challenges. Therefore the implications of in-

roducing ICTs into any M&E process should be carefully considered. Finally, evaluators should take the initiative to engage with app developers to design the kinds of apps they really need rather than being the passive recipients of what developers produce. ■

USING 'BIG DATA' FOR EQUITY-FOCUSED EVALUATION – UNDERSTANDING AND UTILIZING THE DYNAMICS OF DATA ECOSYSTEMS

Kim Forss and Jonas Norén

Big Data: what is it?

In the absence of a universal definition there is an emerging consensus on drivers of the Big Data (BD) phenomenon. Besides the exponential growth of computing power, increases in volume, velocity and variety of data are widely regarded to provide the foundation of the BD paradigm.

Our full article presents an analytical framework composed of four categories of BD based on how data is entered into databases, how it is stored and how it is utilized:

- *Active driven data*: unstructured data intentionally stored. UN Global Pulse metaphorically describes this as “what people say”. Examples of how evaluations can benefit include for example metadata models and discourses on social media, crowdsourcing/ mobile data collection, and participatory statistics.
- *Passive driven data*: unstructured data unintentionally stored. UN Global Pulse describes this as “what people do”. Examples for evaluations are methods using location data, search queries, web cookies, financial transactions etc.
- *Algorithm driven data (i.e. machine to machine communication)*: unstructured data that is actively and intentionally stored by non-human entities. Data from surveillance footage, cell-tower triangulation, satellite imagery and different types of survey drones are theoretically available and could be of value for evaluations.
- *Public Statistics (or open data)*: usually structured data that is actively entered into

databases by specific actors and individuals (i.e. primary data, public statistics, scientific papers etc.).

Can one get access to it?

The lion's share of BD is privately owned; a valuable commodity and an important driver of the global information economy. As a result, access to data that could be put to use in public services such as evaluations is restricted – as well as regulated by national policies and laws. Attempts to retrieve *active driven data* from Twitter have shown that little is accessible for external stakeholders. *Passive- and algorithm driven data* is deemed equally or even more difficult to access. Both as a cause and effect of the above, there is a broad lack of capacity and competence among traditional evaluators.

What about methodological flaws? Tools to capture BD are often based on metadata and deductive reasoning, which implies potential risks in terms of limited contextual comprehension. Thus, BD needs to be complemented with insights about the issues being scrutinized. There seems to be good reasons to question and critically assess the reliability and validity of the information secured from BD sources.

So how easy is it to critically assess BD? It is not easy at all. One major shortcoming is the lack of transparency around how BD has been generated. This hinders the evaluator's defence of its reliability and validity. Are agencies that own the data likely to disclose

how the data is obtained and stored? Do commercial interests stand in the way of full, timely and transparent disclosure?

The demand for (and use of) BD

We assessed the demand for BD by looking at the terms of reference (ToRs) for a judgmental sample of evaluations. Our review confirmed that ToRs tend to be prescriptive and leave little leeway for evaluation teams to be innovative. First, they set very specific questions. Second, they prescribe the overarching design. Third, they emphasise process aspects and de-emphasise relevance and impact analysis where BD could make a difference.

Thus our review indicates that the demand for BD is weak both as a main source of information and as a supplementary data source. No more than two evaluations out of a sample of twenty three actually used what might be called BD in accordance with the above definitions. In both cases the evaluators used data labelled as public statistics. On the other hand, BD was a significant element in the analysis and the conclusions and recommendations would have been less compelling without this evidence.

Could the evaluations we reviewed have made better use of BD? Would they have been more solid and interesting? Could the evaluation process itself have been more efficient? These questions can only be addressed through guesstimates. On this basis at least 20 of the 23 evaluations could hypotheti-

cally have benefited from one or more of the listed categories of BD.

It would of course be presumptuous to pronounce a definite judgement on these questions. Nor did we assess the quality of the evaluations. The fact that most of them did not use BD does not in any way suggest that they did not serve their purpose or failed to respond to the evaluation questions. They might nevertheless have done so more convincingly and possibly more efficiently by using BD. So why did they not do so?

Incentives to use BD

There are no incentives to use BD. It is rather the other way round: there are incentives **not** to do so.

- As long as ToRs regulate evaluation approaches, methods and data sources in detail – and avoid pointing to utilization of BD – evaluation teams are not likely to avail themselves of BD. Evaluations are put on tender and in a proposal the evaluators show how they respond to the ToRs. To do so with a suggestion to use different data sources would be risky.
- Second, planning to use BD is to engage with uncertainty. At present, the nature

of BD is not transparent and basic information on representation, inclusion, origins, sources of bias, etc. are frequently unknown. In such cases the reliability and validity of BD is often open to question.

- Third, most evaluation commissioners privilege well structured and planned evaluations that adhere to budgets and timelines. Hence the relatively open inquiry that presently seems to be called for when using BD clashes with common practice.
- Fourth, evaluation teams may not be equipped with the requisite competencies to locate and use BD. In the evaluations quoted above the evaluation team members were paid to organise, conduct and use interviews and focus groups, to design, disseminate and collect surveys, to observe conditions on project sites, and to meet with stakeholder groups. That is what most evaluators know how to do. To engage with BD involves a different kind of expertise.

It is thus not surprising that BD is not much used. The evaluation community focuses on other methodological issues, such as quality standards, utilization, experimental versus other approaches, etc. Institutional inertia, lack of capacity and competence on how

BD could be used mitigate against any rapid proliferation of BD use in current evaluation practice.

Still, innovation, new rules of interaction, and digital infrastructure are changes that will affect the way people think about evaluations and force the practice of evaluation to adapt. BD could for instance then contribute to impact assessments with more refined data, as well as to make the assessments more efficient. For example, usage of passive driven data (what people do) may be more relevant than data from surveys and interviews (what people say) in assessment of relevance and impact. A more speedy adaptation of BD would require:

- On the demand side, an expression of interest from the agencies commissioning evaluation reflected in ToRs.
- Capacity development on the use of BD, in targeted programmes as well as in basic training programmes in evaluation and research methodology.
- Institutional development in respect of access to BD, where professional associations may have a role to play.
- A commitment to transparency by agencies generating, storing and making BD available so that BD reliability and validity can be verified.

BIG DATA, BIGGER PROBLEMS: HOW BIG DATA CAN HELP DEVELOPMENT EVALUATION METHODOLOGY IN CONNECTED MARKETS

Kara Chiuchiarelli

'Big Data' has become a buzzword in the development community since the United Nations' 2012 *'Big Data For Development'* report, the World Economic Forum (WEF) 2012 *'Big Data, Big Impact'* publication and the World Bank's 2014 *'Big Data in Action for Development'* review. The United Nations defines 'Big Data' (BD) as content that is 'digitally generated, passively produced, automatically collected, geographically or temporally "trackable" and continuously analysed data' (2012), including but not limited to: call logs, mobile-banking transactions, online user-generated content such as blog

posts and Tweets, online searches, satellite images, etc.

BD is stored in large digital files that require terabytes and petabytes of space. For perspective, a terabyte is the approximate size of a personal external hard drive. Petabytes are 1,000 terabytes. This is the measure the National Security Administration uses to describe the amount of data it processes daily – about 30 petabytes (Letouze 2013). The size of BD, however, is not in and by itself revolutionary or useful. It is the *ability to retrieve and analyse* such data, and the novel

tools that have become available to do so that make BD potentially helpful to professionals.

'Mining' this ever-growing data mountain begs two important questions: How to use BD for development, and how to use BD for evaluation? More research, experimentation and evaluation will in time answer these questions. However, based on what we already know, it is a safe bet that with access to a good information and communications technology (ICT) infrastructure, BD can add to the evaluator's toolkit to produce more relevant, effective, efficient and impactful evaluations.

How can it be used for development?

In developed countries as well as in emerging market countries, many individuals are taking full advantage of the new information and communication technologies (ICT), e.g. mobile phones are widely used for banking or social media applications). Already public and private organisations are utilising the ICT infrastructure to monitor roadway and public transit traffic, collect information from satellite imagery, and monitor online search and call logs. Collecting such data can also add depth and nuance to social research initiatives in a variety of fields.

The WEF publication (2012) suggests that BD can be used in such far-reaching areas as financial services, education, health and agriculture. Data can be mined and analysed to produce 'faster outbreak tracking and response, improved understanding of crisis behaviour change, accurate mapping of service needs, and an ability to predict demand and supply changes' (Vital Wave 2012).

The World Bank review (2014) notes that mobile phone data is already used to analyse population displacement and improve emergency preparedness. Search queries and social media can also help quick identification of epidemic changes. The UN report also concludes that BD can offer improved understanding of human behaviour by probing responses to early warning and by generating real-time feedback to social interventions in diverse cultural contexts (2012).

However as the World Bank notes, 'Big Data is not a panacea'. Concerns over BD use include:

- an increased tendency to *apophenia*, i.e. seeing meaningful patterns and connections where none exist (Cohen-Setton and Letouze 2013);
- the advent of a 'new digital divide' that may widen rather than close existing gaps in income and power worldwide;
- a potential for selection bias based on who uses the digital media or whose data is available in mass sets;
- a neglect of internal and external validity considerations: correlation is not causation
- invasion of digital privacy (Shaw 2014).

Thus the costs and risks of using BD should be weighed against the gains.

How can BD be used for evaluation?

Experience suggests that BD can be useful to evaluate infrastructure use; to probe beneficiaries' attitudes and behaviours; and to gauge the extent of social consensus. However, data sets are complex, time-sensitive and valueless without interpretation. Experts are needed to identify correlations, assign value to items and muddle through complex systems. Even if BD is useful, evaluators are needed to make sense of and assign value to the findings.

As noted by Preskill (2013), BD can offer evaluators: shorter cycles; real-time feedback; innovative data collection and sophisticated data visualisation and graphics. Encouraging everyone to collect and use data as part of on-going practice is the key to securing significant BD benefits. Available research methods that produce accurate and comprehensive impact analyses can either be beyond an evaluator's scope or offer an incomplete picture. A gap remains between traditional methodology and the ability to produce a comprehensive, accurate analysis of a program, policy, project or organisation.

With traditional methods, there is a trade-off between intensive data collection and extensive data collection.

This trade-off is eliminated with BD: one can simultaneously collect in-depth and valid results by cross-referencing data sets using a variety of queries through such open-source software tools as Hadoop which allow evaluators to look to combine new and old data sets and use the results to reach entirely new conclusions.

In a nutshell, BD in evaluation should meet three criteria to generate significant benefits:

- presence of a well-connected population producing multiple sets of BD;
- an evaluation commissioned for a program, policy, project or organisation for which BD is available;
- evaluator's capacity to analyse data.

Not all evaluations will meet such criteria. But when evaluators are trained to analyse BD and the data is available it is a cost-effective way to increase efficiency, impact and relevance of findings. Analysing indepen-

dent BD for societal trends also increases evaluation validity by allowing evaluators to compare outcomes in different contexts.

In sum Big Data is not just a buzzword. It is a tool evaluators should use to conduct relevant, effective, efficient, impactful and sustainable evaluations. However embracing 'Web 3.0' in the evaluation domain will require new attitudes, processes and skills in order to tap the full benefits of the data revolution.

References

Cohen-Setton, J., and E Letouze. (26 March 2013) "Blogs review: Big data, aggregates and individuals". Brugel: Improving economic policy.

Letouze, E. (2012) *Big Data for Development: Challenges and Opportunities*. New York: United Nations Global Pulse.

Letouze, E. (2013) "Big Data for Development: Facts and figures". SciDev.Net's Big Data for Development.

Mayer-Schönberger, V. and K. Cukier. (2013) *Big data: a revolution that will transform how we live, work and think*. London: John Murray.

Morra-Imas, L. G. and R.C. Rist. (2009) *The road to results: designing and conducting effective development evaluations*. Washington, D.C.: World Bank.

Preskill, Hallie (2013) "Introducing Next Generation Evaluation". (November 2013). "Next Generation Evaluation: Embracing Complexity, Connectivity, and Change". Stanford Social Innovation Review. Stanford, CA, USA: Stanford Center on Philanthropy and Civil Society.

Shaw, J. (March-April 2014) "Why Big Data is a Big Deal". Harvard Magazine. Cambridge, MA, USA: Harvard Magazine.

Vital Wave Consulting. (2012) *Big Data, Big Impact: New Possibilities For International Development*. Geneva: World Economic Forum.

World Bank and Second Muse. (2014) *Big Data In Action For Development*. Washington D.C.: World Bank Group.

BILLIONAIRE TO THE RESCUE: DOES THE GATES FOUNDATION HAVE A SOUND EVALUATION POLICY?

Linette Lim

Introduction

Interventions by the Gates Foundation have multiplied since the 2008 global financial crisis. This has helped to fill the void that budget austerity has induced in bilateral aid programmes. But there are risks to democratic accountability associated with the involvement of a private donor in control of resources that dwarf the World Health Organization (Elbe, 2010), as well as the GDPs of some low-income countries. Writing for *Southern Voice*, Hulme (2013, p.4) tells of an encounter with the Ugandan health minister: “If you want to know what is happening in health policy in Uganda, ask Bill Gates and the Global Fund. I am only the Minister of Health.”

The role of evaluation

This is where evaluation has a role to play since it asks questions not just about program effectiveness but also about unintended consequences. Equally evaluation helps to establish a feedback loop aimed at improving programme design and implementation. The degree to which evaluation can meet these goals depends on the evaluation policy framework.

In the absence of generally accepted standards for evaluation governance and practices in private foundations this article addresses three questions. First, how does the Gates Foundation’s evaluation policy compare with that of its peers? Second, to what extent does actual evaluation practice conform to the professed policy? Third, does the policy live up to the evaluation principles adopted by the broader development community?

Peer comparison

Trochim (2009)’s evaluation policy taxonomy consists of Goals, Participation, Capacity Building, Management, Roles, Process and Methods, Utilization, and Meta-evaluation. Viewed through this lens the Gates Foundation does not lag behind its peers in the

philanthropic sector, e.g. the Hewlett Foundation and the Kellogg Foundation, ranked in the top ten in terms of assets (Foundation Center, 2014).

In terms of depth and policy content the scope of the Gates Foundation evaluation function compares well with that of both foundations. There was broad convergence in all areas except Participation. Compared to the Hewlett Foundation, the Gates Foundation prescribes more grantee involvement in evaluation designs, while it had a narrower conceptualization of participation compared to the Kellogg Foundation. The Gates policy explicitly states its commitment to “involv[ing] partners in joint decision making”, but does not extend this involvement to other stakeholders. By contrast, the wording of the Kellogg policy (“The best evaluations... involve a representation of people who care about the project”) is more inclusive since it encompasses the civil society and the general public.

Looking out for consistency in practice

In order to test the extent to which evaluation practice conforms to the stated policy I examined the evaluations of two programmes funded by the Gates Foundation. The first concerns the Foundation’s High School Grants Initiative (2006). The second is the Worldwide Information and Communications Technology for Maternal, Newborn and Child Health Pilot Project (2013).

In both instances the conformity to policy was low. Out of eight evaluation policy areas alignment between policy and practice was observed in only three – Goals, Participation, and Process and Methods. In particular, Capacity Building and Utilization were not addressed in both evaluation reports. The lack of emphasis on capacity building reflects a predilection for the achievement of goals set by the Foundation rather than a concern with the development of indigenous capabilities. Lack of clarity regarding the translation

of findings into recommendations was also a feature of both reports.

Benchmarking against international principles

Lastly, the Gates policy needs to be evaluated against the guiding principles of the broader international development community. Taking the OECD’s Development Assistance Committee (DAC) Principles for the Evaluation of Development Assistance (1991), and the United Nations Development Program (UNDP) Evaluation Policy (2011) as benchmarks, critical limitations of the Gates policy emerge particularly in terms of (i) accountability, (ii) independence (of the evaluation function), (iii) transparency, (iv) attribution and (v) professional standards

- (i) DAC (OECD, 1991, p.4) emphasizes accountability to “political authorities and the general public”, while UNDP (2011, p. 3) brings up the need for evaluation to be “guided by national priorities and concerns”. Apart from a vague one-liner on investing in national evaluation capacity “whenever possible”, the Gates policy evades the public accountability issue.
- (ii) Both DAC and UNDP cite independence of the evaluation process and function as a key principle. However, the Gates Foundation does not spell out how much independence from management the in-house evaluation team enjoys.
- (iii) For DAC (OECD, 1991, p.4), “The evaluation process must be as open as possible with the results made widely available”; similarly, UNDP states that all its evaluation reports are made public. On the contrary, at the Gates Foundation, information is shared at the discretion of program officers.
- (iv) On attribution, the Gates policy maintains that evaluation is “... for learning and decision making rather than for proof that foundation resources are re-

sponsible for the outcomes of our joint efforts with partners.” This falls short of DAC and UNDP principles that emphasize causality and accountability for development results.

- (v) The Gates Foundation’s “current practice in evaluation is characterized by variation” and ample delegation to program teams and program officers. With standards deliberately vague conflict of interest and transparency concerns arise.

Conclusion

The weakness of the Gates Foundation’s evaluation policy with respect to such issues as transparency, independence, and capacity building is worrisome given the political clout and the prominent role it plays in the international development domain. The above assessment corroborates broader criticisms (Bennett, Rihouay, and Camara, 2014) that point to the Foundation’s tendency to disregard national policies – such as the strengthening of health systems – and its relentless focus on Bill Gates’ own agenda. Evidently the Gates Foundation’s evaluation policy needs to be upgraded – especially with respect to the exercise of checks and balances on the Foundation’s own power.

References

- American Institutes for Research & SRI International. (2006). Evaluation of the Bill & Melinda Gates Foundation’s High School Grants Initiative: 2001 – 2005 Final Report. [Online] Available from: <https://docs.gatesfoundation.org/Documents/Year4EvaluationAIRSRI.pdf> [Assessed: 3 November 2014].
- Bennett, S., Rihouay, F., and Camara, O. (2014). Bill and Melinda Gates Are Spending Their Money the Wrong Way, Critics Say. Bloomberg. 22nd December. Available from: <http://mobile.bloomberg.com/news/2014-12-22/gates-left-ebola-door-open-with-aids-focus-critics-say.html> [Assessed: 3 Jan 2015].
- Bill & Melinda Gates Foundation. (2014). How We Work: Evaluation Policy. [Online] 2014. Available from: <http://www.gatesfoundation.org/How-We-Work/General-Information/Evaluation-Policy> [Assessed: 30 October 2014].
- Elbe, S. (2010). Security and Global Health: Towards the Medicalization of Insecurity. Cambridge: Polity Press.
- Foundation Center. (2014). Key Facts on U.S. Foundations: 2014 Edition. [Online] Available from: http://foundationcenter.org/gainknowledge/research/keyfacts2014/pdfs/Key_Facts_on_US_Foundations_2014.pdf [Assessed: 20 November 2014].
- Hulme, D. (2013) The Post-2015 Development Agenda: Learning from the MDGs. Southern Voice on Post-MDG International Development Goals, Occasional Paper 2. Dhaka: Centre for Policy Dialogue.
- Invest in Knowledge Initiative. (2013). Evaluation of the Information and Communications Technology for Maternal, Newborn and Child Health Project. [Online] 2013. Available from: http://innovationsformnch.org/uploads/resources/pdfs/ICT_for_MNCH_Report_131211md_FINAL,_AY_to_Gates_language.pdf [Assessed: 3 November 2014].
- OECD. (1991). DAC Principles for Evaluation of Development Assistance. [Online] 1991. Available from: <http://www.oecd.org/development/evaluation/2755284.pdf> [Assessed: 29 March 2015].
- Trochim, W. (2009). Evaluation policy and evaluation practice. In: Trochim, W., Mark, M., & Cooksy, L. (eds). Evaluation Policy and Evaluation Practice: Where Do We Go From Here? New Directions for Evaluation. 123. p. 13–32.
- UNDP. (2011). The evaluation policy of UNDP. [Online] 2011. Available from: <http://web.undp.org/evaluation/policy.shtml> [Assessed: 29 March 2015].
- W.K. Kellogg Foundation. (2004). W.K. Kellogg Foundation Evaluation Handbook. [Online] 2004. Available from: <https://www.wkkf.org/resource-directory/resource/2010/w-k-kellogg-foundation-evaluation-handbook> [Assessed: 2 Jan 2015].
- William and Flora Hewlett Foundation. (2012). Evaluation Principles and Practices: An Internal Working Paper. [Online] 2012. Available from: <http://www.hewlett.org/uploads/documents/EvaluationPrinciples-FINAL.pdf> [Assessed: 2 Jan 2015].

ENGAGING PARLIAMENTARIANS TOWARDS A BETTER ENABLING ENVIRONMENT FOR EVALUATION

Liisa Horelli

The International EvalYear 2015 is progressing well with a host of initiatives from all over the world¹. The European Evaluation Society (EES) has organized special events and participated in such projects as the Innovation Challenge Competition. One of the competition winners was a project sponsored by EES

“*Engaging Parliamentarians for an Evaluation Culture*”². Its core idea was to enhance the supply and demand side of evaluation by mobilizing parliamentarians, policy makers and practitioners to enhance the enabling environment for evidence-based policy making and institution building (Horelli, 2015).

The aim of this article is to identify basic evaluation concepts relevant to the parliamentary context and to disclose the results of an explorative study comprising a literature review and a survey. Although the survey sample was small (19 parliamentarians) it captured revealing data from dif-

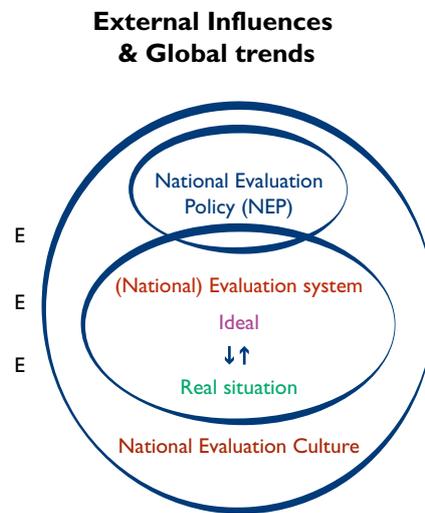
ferent continents and reflected the diverse maturity of evaluation cultures. The research questions were: (i) How to define the basic concepts around evaluation, especially the enabling environment for evaluation (EEE)? (ii) Why should parliamentarians get involved and what is the role of the Parliament in the creation of the EEE? and (iii) how can the citizen voice be amplified by or through the EEE? The analysis of the results was guided by critical systems heuristics (Reynolds, 2005).

EES is conditioned by the evaluation policy, system and culture

Both the literature review and the survey indicate that an EEE is a polymorphic setting, which requires a great deal of resources and different types of knowledge. It can be characterized by features that are *structural* (national evaluation policy, a legal framework, a well-functioning evaluation system, e-units in the parliament), *functional* (demand and use of evaluations as a routine in legislation, in debates over policy issues and in oversight), *temporal* (more time for debates, slower decision-making) and *methodological* (more ex-post evaluations, new methods to tap future evaluations, simple language).

The developmental situation of the country as well as global trends and external influences have an impact on the demand for a national evaluation policy (NEP). Developing countries regard the NEP as a nation

Figure 1: The Enabling environment for evaluation (EEE).



building instrument. This is not the case in mature evaluation cultures, where evaluations are considered routine. Whereas NEP provides the normative framework for evaluations the national evaluation system (NES) is the operational mechanism that enables implementation of the NEP policy.

Both NEP and NES are embedded in and interacting with the National Evaluation Culture the institutionalization of which seems

to be a decisive marker of cultural maturity (Rosenstein, 2013; Jacob et al., 2015). In sum, the Enabling Environment for Evaluation (EEE) can be defined as the complex setting for the demand, supply and use of evaluations which in turn is conditioned by the policy, system and culture (see Figure 1).

The role of Parliament reflects the maturity of the evaluation culture

The reasons why parliamentarians are engaged with an EEE (or should do so) are: (i) evidence-based policy, (ii) accountability for the public good, (iii) mechanism for transparency, (iv) the oversight role of Parliament and (v) lack of NEPs. The diverse prerogatives and missions of national parliaments were reflected in the answers which, in turn, mirrored the maturity of the evaluation culture.

Developing countries respondents stressed the significance of National Evaluation Policies (NEPs) and the oversight function of the Parliament, whereas the representatives of the more mature evaluation cultures viewed the role of Parliament as one way to enhance the role of evaluation as part of a *managerial model* featuring standardized procedures and interaction with executive agencies (Toulemonde, 2001).

Some representatives from developed countries aspired for the Parliament to adopt a *democratic model* in which evaluation is at the core of political debates (Table I). It was based on flexible institutionalization, meaning that the organization of evaluation structures or the EEE would depend on the changing context and challenges which can be transformed, according to democratic deliberations (Picciotto, 2015).

Strengthening the citizen voice through transparency and access

Globalization has brought forth wicked problems that require cross border solutions with major implications for evaluation. Through the new information technologies a wider range of stakeholders and ordinary citizens can now engage in evaluation processes. In this context, survey respondents underlined openness, transparency and access to information flows as basic principles

Table I: Democracy in evaluation and Evaluation in democracy.

Maturity of evaluation culture	Democracy in evaluation	Evaluation in democracy	Examples of EES characteristics
Low	NEP, NES	Standardized procedures; access to information flows	Parliamentarian Forum Involvement of NGOs
Medium	Managerial model of evaluation arrangements Centralized agencies Support to SAI	Standardized procedures Access to information flows	Equity and gender sensitive evaluations (Segone, 2014)
High	Democratic model of evaluation Flexible institutionalization	Radical transparency; new real time methods & procedures (Stern, 2013)	Accessing the prime minister; new stakeholders (citizens) Future policies as the object of evaluations

to be observed in order to amplify citizens' voices and influence on decision making.

Diverse historical periods require distinctive solutions. No one size fits all. National Evaluation Policies (NEPs), which are a suitable instrument for countries with an emerging evaluation cultures might be considered redundant in more mature culture. However, Parliaments in developed countries may opt to become more active in evaluation through adoption of a NEP that stresses democratic decision making and enhances the enabling environment for evaluation and the role of evidence in policy making. In an increasingly complex and interconnected policy world flexible epistemologies and mixed methods (rather than simply experimental methods) should be favored.

Democratic evaluation in a globalized world is particularly demanding. Policy makers and politicians will have to ask probing questions and commission evaluations that link the local, regional, national and the global. They will also have to adopt novel roles as

evaluation infrastructure builders, evaluation agenda setters and promoters of hubs that link fragmented knowledge sources (Stern, 2013). This is the vision that should emerge from the Global Parliamentarian Forum in November 2015 at the final EvalYear event.

References

Horelli, L. (2015) Engaging Parliamentarians for an Enabling Environment of Evaluation. An Interim report of the action research on the mobilization of parliamentarians to enhance evaluation cultures. Unpublished manuscript.

Jacob, S., Speer, S. & Furubo, J-E. (2015) The Institutionalization of Evaluation Matters: Updating the International Atlas of Evaluation 10 Years Later. *Evaluation*, 21 (6–31). doi:10.1177/1356389014564248.

Picciotto, R. Democratic evaluation for the 21st Century. *Evaluation*, April 2015 21 (150–166), doi:10.1177/1356389015577511.

Reynolds, M. (2005) Evaluation Based on Critical Systems Heuristics. In B. Williams & I. Imam (Eds.) *Systems Approaches to Evaluation: An Expert Anthology* (pp. 101–122) California: American Evaluation Association.

Rosenstein, B. (2013) Mapping the status of national evaluation policies <http://gendereval.org/forum/topics/parliamentarians-forum-for-development-evaluation-publishes>.

Segone, M (Ed.) (2014) National evaluation policies for sustainable and equitable development. How to integrate gender equality and social equity in national evaluation policies and systems. Available <http://mymande.org/selected-books>.

Stern, E. (2013) Evaluation in an evolving democratic setting. *Connections*, Special Issue, June 2013, 5–6.

Toulemonde, J. (2000) Evaluation Culture(s) in Europe: Differences and Convergence between National Practices. *Vierteljahrshefte zur Wirtschaftsforschung*, 69 (3), 350–357.

1 <http://mymande.org/evalyear/evaluationtorch2015>.

2 It was submitted by the Parliamentarians Forum of South Asia jointly with the Pakistan Evaluation Network (Pen), South Asian Community of Evaluators (CoE), and the European Evaluation Society (EES).

THE AUTHORS

Michael Bamberger

Michael Bamberger worked on urban community development in Latin America for over a decade and in the evaluation, gender and training departments of the World Bank. He has consulted and taught on development evaluation with UN, multilateral and bilateral agencies and international NGOs. He also advises the Evaluation Office of the Rockefeller Foundation. Recent publications include: "RealWorld Evaluation: Working under budget, time, data and political constraints" (with Jim Rugh and Linda Mabry), "How to design and manage equity focused evaluations" (with Marco Segone), "Engendering Monitoring and Evaluation" and "Emerging opportunities: Monitoring and Evaluation in a Tech-enabled world" (with Linda Raftree).



Kara Chiuchiarelli

Kara Chiuchiarelli is a masters candidate in Emerging Economies and Inclusive Development King's College London's International Development



ment Institute. Her areas of research lie in communications, data, information, and technology as related to program modeling and evaluation. She is particularly interested in issues surrounding big data use and open data policies. She keeps up-to-date with recent work on her website, www.kchiuch.com. Ms. Chiuchiarelli holds an H.B.A. in Communications from Marquette University in the United States.

Kim Forss

Kim Forss holds a Ph.D. from the Stockholm School of Economics. He is the founder/owner and manager of Andante – tools for thinking AB, and in the company works with research, teaching, and consultancy assignments in evaluation. Over the past 25 years, he has evaluated policy and strategy in the field of development cooperation, public administration, the national system of Parliamentary inquiries, as well as policies for research and development. He has worked extensively with organisational analysis, for example of the Swedish Government Offices, the national museum structure,



music exports, Swedish regional administration and of several UN agencies (governance systems in UNEP, reform processes at WHO and UNDP, organisational learning in UNICEF, and a comprehensive evaluation of UNESCO). His recent research publications include a recent volume on evaluating complex policy, and he contributes to a forthcoming volume on evaluation cultures, as well as evaluation and equity. He is a member of the European Evaluation Society since 1992 and a founding member of Swedish Evaluation Society, where he has served on the Board for six years, in 2010 and 2011 as the President.

Leon Hermans

Leon Hermans is an assistant professor at the Faculty of Technology, Policy and Management of Delft University of Technology in the Netherlands. His research and teaching focuses on methods and approaches for evaluation and policy analysis in multi-actor systems. Application areas are water, environment, and delta management, with projects currently ongoing in the Netherlands, Bangladesh, India and Vietnam.



Liisa Horelli

Liisa Horelli, PhD in environmental psychology, is adjunct professor at the Centre of Urban and Regional Studies, Aalto University, Finland. She has carried out action research (i.e. action and its evaluation) and published articles on participatory urban and regional planning, especially from the perspective of health promotion and the enhancement of human-, and gender-friendly environments with different stakeholders. In addition, she has conducted evaluations since the 1990s on EU-funded programmes and projects. Liisa was a President of the Finnish Evaluation Society (FES).

**Linette Lim**

Linette Lim is pursuing an MSc in Emerging Economies and International Development at King's College London. She is on a career sabbatical from her role in Channel NewsAsia, a pan-Asian television station, where she had been reporting on business and economics for the past four years. In her spare time in London, she is involved in grant writing for Pragya UK, an NGO that develops and delivers education and public health interventions to isolated mountain communities in the Indian Himalayas, Nepal, and Kenya.

**Jacqueline Mallender**

Jacqueline is a Partner with Optimity Advisors and a respected international public policy economist. Over the last 25 years, she has directed many policy appraisal and evaluation studies in health, crime, justice, home affairs, education, employment and social welfare. She is a convenor of the Campbell Collaboration Crime and Justice Coordinating Group and the joint Campbell and Cochrane Economics Methods Group and is an economic advisor to the What Works Centre for Crime Reduction Commissioned Partnership in the UK.

**Jonas Norén**

Jonas Norén holds a master degree (M.Sc.) in political science and economics. During recent years he has been engaged as a consultant within the fields of politics, economics and development assistance. Jonas is experienced in evaluation, policy advice, research, statistics, methodological design and capacity building. From a thematic point of view he has worked extensively with private sector development.

**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.

**Linda Raftree**

Linda Raftree has worked at the intersection of community development, participatory media and Information and Communication Technologies (ICT) since 1994. She is a co-founder of Kurante and advises the Rockefeller Foundation's Evaluation Office on the use of ICTs in Monitoring and Evaluation. Linda has conducted research on adolescent girls and ICTs for UNICEF, the role of ICTs in child/youth migration for the Oak Foundation, and the use of mobile technologies in youth workforce development for the mEducation Alliance. She also convenes the Technology Salons



in New York City, writes 'Wait... What?' a blog about new technology and community development, and tweets at @meowtree.

Martin Reynolds

Martin Reynolds is a Senior Lecturer and Qualifications Director in Systems Thinking in Practice at The Open University, UK. His teaching and research focus on issues of critical systems thinking in relation to international development, environmental management, and business administration. He has published widely in these fields and also provides workshop support and facilitation for professional development in systems practice and critical systems thinking.

**Rory Tierney**

Rory is a Senior Economist at Optimity Advisors, working predominantly in public policy. He has produced cost-benefit and cost-effectiveness analysis for clients such as the European Parliament, EU-OSHA and the National Institute for Health and Care Excellence (NICE, UK), and previously worked in health economics at a major pharmaceutical company.

**Ole Winckler Andersen**

Ole Winckler Andersen is Deputy Permanent Representative at the Danish Delegation to the OECD. He holds a M.Sc. in Economics and a Ph.D. in Public Administration and has worked for the Danish Ministry of Foreign Affairs for the last 20 years, including as Head of Evaluation Department (2007–13). Before that he was assistant and associate professor at Roskilde University in Denmark. He has managed a number of evaluations and has served as team leader for missions to Asian, African and Latin American countries. He has published on evaluation in various international journals and recently co-edited *Evaluation Methodologies for Aid in Conflict* (Routledge 2014).

**GUIDANCE TO CONTRIBUTORS**

Brevity and nimbleness are the hallmarks of Connections: published articles are normally 800–1,200 words long. Even shorter contributions (news items; opinion pieces; book reviews and letters to the editor) are accepted with a view to promote debate and connect evaluators within Europe and beyond. While Connections is not a peer reviewed publication only cogent contributions that add to knowledge about evaluation theory, methods and practices should be submitted.

Articles that highlight European values and evaluation practices are given priority but Connections also reaches out beyond Europe to the international evaluation community and favours articles reflecting a diversity of perspectives. Within the limits of copyrights agreements articles that summarize in a cogent way the substantive content of published (or to be published) studies, papers, book chapters, etc. are welcome (with suitable attribution).

Individuals or organizations wishing to sponsor *special issues* about an evaluation theme or topic of contemporary interest should contact the EES Secretariat (secretariat@europeanevaluation.org). Such special issues usually consist of 6–8 articles in addition to a guest editorial. A Presidential letter is normally included. The guest editor(s) are responsible for the quality of the material and the timeliness of submissions. The regular editorial team ensures that special issues meet 'Connections' standards and takes care of copy editing.