



EVALUATIONS CONNECTIONS

EES

EUROPEAN
EVALUATION
SOCIETY

MARCH 2013

ARTICLES – NEWS – EVENTS

CONTENT

| | |
|---|----|
| President's message | 1 |
| Editorial: Where is evaluation going? | 2 |
| Evaluation for the 21 st Century | 3 |
| EvalPartners: Working together towards country-led evaluation systems responsive to social equity | 4 |
| Addressing the micro-macro disconnect in climate change resilience? | 5 |
| Innovative approaches to impact evaluation | 6 |
| Contribution Analysis at the Helsinki Conference | 8 |
| Payment by Results: Does it work? | 9 |
| Towards more useful comprehensive evaluations of international organizations | 10 |
| Valuing Evaluation Power and the Power of Evaluation in "Speaking Truth to Power" | 12 |
| Strengthening Evaluation Focused Programmes in European Universities | 14 |
| Obituary: Carol Weiss | 16 |
| EES 'Public Hearing' at the EU Parliament | 17 |
| Proposal for a new Thematic Working Group on Public Private Partnerships | 18 |
| EES Board member elections & Board member co-option 2013 | 18 |
| The authors | 19 |

President's message

María Bustelo, EES President

Dear Members and Colleagues,

Welcome to the first 2013 issue of Connections. Relying once again on the rich and exciting debates of the 10th EES Biennial Conference held in Helsinki last October it confirms the vitality of our discipline and highlights the role that EES plays as a platform for evaluation knowledge exchange and a launching pad for initiatives geared to the professionalization of our craft.

Demand for high quality evaluation keeps on growing. Throughout the past decade our practice has undergone a consolidation and institutionalization throughout Europe. The emergence of new national societies in the European space as well as the growth of university based programmes specifically devoted to evaluation testify to this trend. But we still have a long way to go. Evaluation is still an infant industry that needs protection and support.

The gap between the demand and the supply of high quality evaluation training and education is still wide. Many evaluation practitioners still identify themselves as economists, sociologists, psychologists, etc. rather than as evaluators. The membership of evaluation associations fluctuates as "new evaluators" join our ranks and many existing members return to what they still consider their primary discipline. At a critical time in Euro-

pean history accountability and social learning should be at a premium. Yet, genuinely independent evaluation is still limited not only in government but in the private sector and the civil society as well.

From this perspective, communities of practice housed and promoted by evaluation societies offer an incomparable setting for much needed evaluation capacity development. The complex and insecure world in which we live can be made better through evaluation. In turn, this calls for a democratic approach to our practice grounded in principled deliberation and ethical processes. We need to constantly challenge ourselves and remain suspicious of standard and fixed doctrinal positions regarding how programmes or policies can or should be evaluated. The evaluation profession should be built on solid theoretical foundations but it should also be in constant evolution and it should concentrate on real life issues within specific contexts.

Only then will evaluation help to solve practical issues and only then will our practice remain relevant and influential in making the world a better, more democratic, more sustainable, more equitable place.

In an interconnected world, evaluation capacity development is important everywhere. It is not the exclusive preoccupation

of international development evaluators. The western industrial democracies are in dire need of stronger evaluation functions focused on problems that used to be considered the preserve of developing countries – excessive debt, high youth unemployment, massive inequality and governance dysfunctions. Indeed, developed countries have much to learn from the experience of developing countries in their pursuit of mutual accountability and partnership building.

One way of promoting evaluation capacity building is to foster communities of practice. “Communities of practice are groups

of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly” (Wenger, 2007)¹. A structured (and fun!) way to do this is through regular interaction in evaluation conferences and events and in their aftermath. Actively participating in evaluation societies as a VOPE (Voluntary Organization for Program Evaluation) could do much to foster evaluation in Europe and beyond.

We can all benefit from your active participation in communities of practice through diverse voluntary activities. Please do not

hesitate to contact me or any other EES Board member. We stand ready to help you network with like minded colleagues.

María Bustelo
EES President

¹ Wenger, E. (2007) ‘Communities of practice. A brief introduction’. *Communities of practice*, <http://www.ewenger.com/theory>. See also: Lave, J. and E. Wenger (1991) *Situated Learning: Legitimate Peripheral Participation*, New York: Cambridge University Press.

EDITORIAL: WHERE IS EVALUATION GOING?

Robert Picciotto

This issue of Connections identifies key drivers of change that will shape evaluation practice in the years ahead. First, Robert Kirkpatrick highlights the revolutionary improvements in monitoring and evaluation practice that “big data” harnessed by the new information technologies have already begun to trigger. Vast, cheap, timely and accurate evidence will increasingly become available to track the progress of policy interventions and to adapt social programmes to changing needs.

Complementing the considerable impact that technology will induce on evaluation practice is the vast potential of networking. This is illustrated by Marco Segone’s article: by 2012, the knowledge management system sponsored by an international evaluation partnership initiative had been accessed by 222,000 visitors from over 168 countries. In turn, Derek Poate’s article sketches the emerging role of evaluation partnerships in tackling global challenges and identifies cutting edge methodological issues relevant to climate change adaptation evaluations.

All three articles demonstrate that the future of the evaluation discipline will be influenced by new ways of interacting and networking locally, nationally, regionally and globally. In turn, new evaluation methods will emerge to meet a growing demand for

assessing complex interventions. Thus, innovative impact evaluation approaches adapted to complex circumstances are unveiled in the article authored by Elliot Stern et al. It summarizes the findings of a landmark study commissioned by the United Kingdom Department for International Development which demonstrates that mixed methods tailor made to the context are better fit to purpose than traditional methods when a large array of interrelated evaluation questions are put forward by policy makers.

Heeding this advice will improve the quality of evaluations and help deal with intense pressures on the evaluation discipline to generate valid evidence based prescriptions. Specifically the increased complexity of evaluation contexts, objects and issues is likely to fuel a rising demand for rigorous identification of the causal links that underlie postulated theories of change in evaluations. Specifically in complex operating and authorizing environments new methodological challenges will have to be mastered in order to ascertain the distinctive contributions of various actors and interventions to outcomes and impacts.

In this context, Jacques Toulemonde’s article probes the state of the art of Contribution Analysis and concurs with John Mayne’s as-

sertion that it is coming of age while the assessment of payment by results evaluations confirms authored by Burt Perrin and Peter Wallace suggest that contribution analysis would have greatly improved the relevance and utility of evaluations of the Payment By Results approach to aid delivery.

Another instance of the need to improve the validity and use of evaluations is illustrated by Rob van den Berg’s article on comprehensive evaluations of international development agencies. Here again decision makers have been pressing evaluators to come up with well documented answers to multi-faceted questions that straddle sectors, actors and borders.

While the effective demand for independent evaluation is undoubtedly high and rising it may well be lagging behind actual needs according to Sandiran Premakanthan’s examination of the Canadian experience. Similarly, the article jointly authored by the University Study Programmes in Evaluation (USPE) network representatives points to a similar race between the growing need to enhance European evaluators’ competencies and the supply of high quality university programmes at the diploma level. All in all, one can safely conclude that interesting times lie ahead for the evaluation discipline.

EVALUATION FOR THE 21st CENTURY

Robert Kirkpatrick

After the Internet revolution, the mobile revolution, and the social media revolution, we find ourselves in a new, hyper-connected age. People everywhere access digital services through mobile phones. As they buy and sell goods, perform financial transactions, search for information, and share experiences through social networks they produce vast amounts of “big data.” Private sector companies are already mining this information to discover patterns that reveal changing customer wants and needs or evidence about the effectiveness of their advertising campaigns. Similarly, when people’s health, income, or food security circumstances change, they make use of the same information transmission services and leave recognizable digital footprints behind. Thus “big data” can help steer development programmes in the public sphere just as it guides corporate decision-making in the market place.

The United Nations Global Pulse initiative has demonstrated that the explosion in use of digital technologies by vulnerable populations in the developing world is generating actionable digital information about human well being. This is taking place in real-time as people go about their daily lives. As access to mobile devices and digital services becomes easier and cheaper the digital divide is closing - faster than earlier anticipated. The number of mobile subscribers in Nigeria has grown a thousand-fold in 10 years. Jakarta now produces more tweets per day than any city on earth. Last year, twenty-four percent of the residents of Mogadishu checked into Facebook at least once a week. All of these activities are producing data. Once aggregated and made anonymous to protect privacy the information opens up unprecedented knowledge opportunities. We now have the capacity to understand not only how and where to target social interventions in real time but also to gather reliable evidence about the impacts of development policies and programmes.

For example, United Nations research carried out in collaboration with prominent data strategists has shown that social media chatter is an early indicator of spikes in unemployment. The price of food commodities mined from websites closely matches official Consumer Price Index in six Latin America and offers advance warning signs of price volatility. Similarly the volume of Tweets containing keywords is a reliable predictor of official food prices inflation.

It is also likely that exploration of mobile phone calling patterns processed so as to protect privacy will provide policy makers with a powerful lens to detect changes in the collective behavior of communities. At a basic level, such data can be used to generate a real-time awareness of population location and movement. Mobile carriers can, for example, “see” daily commutes as workers travel two or three cell towers away in the morning and return in the evening. Research has already shown that the size and frequency of airtime credit purchases is a good proxy for household income variations: the more one earns, the larger the amount of credit one purchases. Mobile carriers also tell that anonymous records of how people use their phones are valid and accurate pointers of demographic characteristics such as age and gender.

Experts operating in the private sector assert that the true potential of “big data” only becomes apparent when one begins to combine different streams of information. It is likely that this assertion will also hold true in applications to development monitoring and evaluation. For example, a new livelihoods programme just getting underway may be targeting young women and provide them with marketable skills. Soon after the programme is launched “big data” analysis (using models already employed by mobile carriers) would disclose anonymous

increases in airtime purchases as a proxy for increased incomes. Shifts in daily movement patterns from one resembling a job search to one that indicates a regular commute would also be detected. Meanwhile, online social media conversation would reveal young women publicly discussing their participation in the programme and their successes or failures in finding employment and their ongoing experiences at work.

Two distinct opportunities for development evaluation can be identified: one near-term, one a bit further out. The first opportunity derives from the fact that “big data” gets “big” because it is constantly being archived: retrospective analysis of digital signals that a beneficiary population generates can provide a cheap, timely and accurate evidence base for evaluating on-going programmes. The second opportunity is more radical. Built into development programmes from the very outset it would make “real-time evaluation” possible. In combination with contemporary evaluation approaches that seek to ascertain “attribution” and/or “contribution” the new technologies have the potential of revolutionizing evaluation practice by tracking the outcomes and impacts of adaptable programmes within the volatile contexts of developing regions.

Participatory evaluation will also be enriched by the social media technologies. They are not merely data sources but also platforms for engaging directly with beneficiaries, and as such, they represent tools of pro-active social inquiry, evaluative investigation, as well as testing and verification of programme hypotheses. Embracing the new technologies would enable a more agile, iterative, and adaptive approach to development interventions in which real-time feedback enables nimble course correction, greater resilience to disruption by exogenous shocks and accelerated achievement of results¹.

¹ <http://www.unglobalpulse.org/BigDataforDevWhitePaper>.

EVALPARTNERS: WORKING TOGETHER TOWARDS COUNTRY-LED EVALUATION SYSTEMS RESPONSIVE TO SOCIAL EQUITY

Marco Segone

In recent years, and in line with the 2002 Millennium Declaration, multi-lateral and bilateral development partners have been working collectively and individually to promote programmes that are equity-focused, gender-responsive and foster human rights. The Paris Declaration on Aid Effectiveness endorsed in 2005, and the 2008 Accra follow-up meeting, stated that national ownership and leadership are overarching factors for ensuring good development outcomes. The implications for the evaluation function are fundamental.

The principle of ownership means that countries should own and lead their country-led evaluation systems, while donors and international organizations should support sustainable national evaluation capacity development. The recent 2012 Busan High-level forum re-affirmed the above principles, while recognizing that the international development arena has changed significantly. As a result, new modalities such as South/South and triangular cooperation, and new stakeholders such as Civil Society Organizations (CSOs), have been explicitly recognized in the Busan's Declaration on "Partnership for effective development cooperation".

Meanwhile, civil society has played an increasingly central and active role in promoting greater accountability for public action, through evaluation. National and regional Voluntary Organizations for Professional Evaluation (VOPEs) grew from 15 in the 1990s to more than 135 by 2012¹. Acknowledging the enhanced role of civil society, UNICEF and IOCE launched EvalPartners². EvalPartners was met with a surge of enthusiasm evidenced in the joining of 27 members, notably all regional and major VOPEs (including European Evaluation Society and IDEAS) and several UN agencies, within a few months of its launch.

The goal of the EvalPartners Initiative 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. Based on the shared conceptual framework on National Evaluation Capacity Development³ calling for a system approach covering individual capacities, institutional capacity and an enabling environment, the expected outcome of EvalPartners would be three-fold:

- VOPEs would be stronger. Their institutional and organizational capacities would be enhanced.
- VOPEs would be more influential. They would be better able to play strategic roles in strengthening the enabling environment for evaluation within their countries. In so doing, they would help to improve national evaluation systems and to promote the use of evaluation evidence in the development of policies geared towards effective, equitable and gender responsive development results.
- VOPEs would 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.

Towards these ends, EvalPartners carried out a mapping⁴ of existing VOPEs, documented good practices and organized a major Forum⁵ in Thailand in which, for the first time ever, the presidents/chairs of all Regional VOPEs, as well as 30+ national VOPEs and high-level representatives from the World Bank, UNEG, UNDP, UNICEF, UNWomen, UNV, Red Cross, and other key stakeholders met together to discuss the shared framework on National Evaluation Capacity Development and concrete actions to strengthen VOPE capacities.

EvalPartners also enhanced MyM&E⁶, a user-friendly knowledge management system to strengthen national evaluation capacity. In addition to being a learning resource, MyM&E facilitates the strengthening of a global evaluation community, while identifying good practices and lessons learned. The number of visitors has gone up from about 22,000 in 2010 to almost 43,000 in 2011 and to over 222,000 in 2012 with

visitors from over 168 countries. The downloads increased from about 76,000 pages in 2010 to about 111,000 in 2011 to almost 816,000 in 2012.

In 2012 a new innovative e-learning programme on Development Evaluation⁷ taught by 33 world-level keynote speakers was launched. So far, about 8,000 people from 168 countries registered. At the end of the first course, over 70% declared they were satisfied with their experience of the e-learning, and over 88% stated they are planning to attend future e-learning courses organized by MyM&E.

We stand at a historical moment in evaluation history, in which a critical mass of VOPEs, international organizations and bilateral Governments are ready to work together to enhance the use of evaluation for better policy making, as stated in the Chiang Mai Declaration⁸ endorsed by 80 leaders at the EvalPartners' International Forum in December 2012. It will be up to each of us to sustain this critical momentum, with the aim of declaring 2015 the International Year of Evaluation.

1 Rugh, *The growth and evolving capacities of VOPEs*, forthcoming.

2 Please visit <http://www.mymande.org/evalpartners> for additional information

3 Segone, Heider, Oksanen, de Silva and Sanz, *Towards a shared framework for National Evaluation Capacity Development*, forthcoming

4 Please visit <http://www.mymande.org/evalpartners/international-mapping-of-evaluation> for additional information

5 Please visit <http://www.mymande.org/evalpartners/forum> for additional information on the forum

6 Please visit www.mymande.org for additional information

7 Please visit <http://mymande.org/elearning> for additional information and registration

8 Please visit http://www.mymande.org/sites/default/files/Declaration_evalpartners_English.pdf to download the Declaration

ADDRESSING THE MICRO-MACRO DISCONNECT IN CLIMATE CHANGE RESILIENCE!¹

Derek Poate

Two main responses to climate change are mitigation – an anthropogenic intervention to reduce the sources or enhance the sinks of greenhouse gases – and adaptation – adjustments in natural or human systems in response to actual or expected climatic stimuli or their effects so as to moderate harm or exploit beneficial opportunities.

The two approaches are not necessarily discrete or mutually exclusive. Adaptation may contribute to mitigation (e.g. afforestation) while mitigation may induce adaptation (e.g. demand for renewables may change farming systems). There may also be trade-offs and synergies between the two. The main focus of this article is on adaptation and the capacity to adapt, i.e. resilience

Achieving resilience often requires capacity building to prevent or withstand shocks, rebuild, and respond to change... and surprises. This implies that individuals are equipped to adapt to climate change but in turn this is a function of individuals' access to resources so that the adaptive capacity of societies also depends on a collective ability to act in the face of the threats posed by climate variability and change.

In many instances, the most vulnerable populations to climate-related shocks and stresses are women and the poor. Particular vulnerabilities for women are identified with regard to access to health, dependence on agriculture and access to water, formal and informal labour, climate-related disasters, displacement and conflict. The vulnerability of the poor is generally seen as resulting from limited access to assets combined with physical exposure to predicted climate-related hazards.

Evaluation approaches have so far developed in response to the different interventions to tackle the challenges of adaptation. Broadly categorised, these tend to be either upstream actions to develop the capacity of government and institutions to understand climate change and integrate adaptation into decision-making; or downstream,

to build coping mechanisms, adapt systems to incremental climate change and explore transformation of agricultural systems.

The Helsinki panel presentations explored these challenges from two perspectives. Paula Silva Villanueva looked at the downstream problem of framing adaptation at a local level and argued for monitoring through a basket of process indicators based around four dimensions:

- **Adaptive capacity** – changes in perspectives, attitude, behavior and practice at all levels.
- **Underlying causes of poverty and vulnerability:** access to land, markets, technology ...
- **Climate risk management** – dealing with uncertainty.
- **Context** – interdependency across scales, multiple interests and levels of decision-making.

She argued that little or no attention has been given to capturing the decision-making processes at household level. Adaptation cannot be evaluated without taking into

consideration social dimensions of change and decision-making. For this reason, indicators must reflect multiple interacting processes and dimensions of decision making beyond programme/project deadlines and “target” areas. This brings interdependency across levels and a concern for the broader operational environment within which communities live, as this will affect how capacity translates into action.

Gil Yaron looked at emerging experience in Ethiopia where DFID is supporting the Government's Climate Resilient Green Economy Strategy, an upstream initiative developing two strands of projects. One deals with organisation and institution building; the other to develop clean technology and climate change resilient projects.

The M&E approach is trying to develop a theory of change to identify how organisational and institutional capacity building and learning could transform macro-level climate funding to micro level impacts. A major part of the approach has been to develop a process monitoring tool to track performance of government organisations to feed back

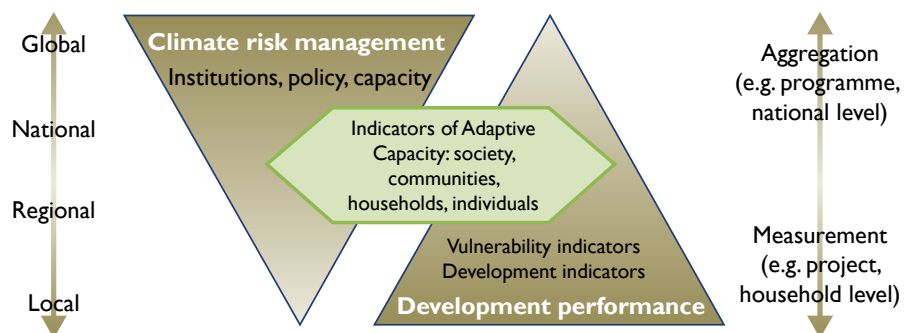


Figure 1
Source: Adapted from Brooks et al, 2011²

how their climate change capacity is developing.

Robbie Gregorowski built on the two examples to argue that climate change M&E frameworks need to:

- combine top down and bottom up approaches
- create a link between integrating climate change into policy/institutions with actions at the household/community level

Combining upstream climate risk management and downstream development outcomes requires indicators of adaptive capacity – at society, community, household and individual level. It is only through collective societal action to bring about the transformative change required that climate change adaptation will be successful. This ultimately comes down to adaptive capacity at all levels and not a dichotomy between climate risk management at the top and development outcomes at the bottom.

Evaluation should not separate the two streams but current experience suggests there is a missing middle between analysis of the policy environment and of development outcomes. The missing middle reflect

a gap in assessment of the adaptive capacity and behavior of people and society as illustrated in Figure 1.

‘Adaptive capacity’ draws together the macro and micro levels: ‘capacity and preparedness to confront longer-term adaptation issues at the institutional level and, flexible and resilient systems, institutions and governance that allow societies to respond to climate change and to uncertain and evolving risks.’

A simple theory of change would be: ‘Climate change adaptation is complex, operating across multiple scales, sectors and involving multiple processes. The M&E of climate change adaptation revolves around how best to ‘bring the people’ at all levels into the scheme. The one unifying factor is adaptive capacity – for society, institutions, communities, households, and individuals.’

The major determinants of adaptive capacity should be relevant and applicable across stakeholders and across scales and across all sectors. Potentially, this approach provides a basis for the development of a set of indicators of adaptive capacity that would

form the ‘glue’ that links upstream and downstream climate change adaptation efforts and connects climate risk management at national and international levels (institutions) to resilience-building for the climate vulnerable poor (the people).

The key question going forward is: can such a set of indicators be used to demonstrate whether people (societies) are adapting and innovating in response to the long term threats posed by climate change? In other words – is society as a whole building sufficient ‘adaptive capacity’.

-
- 1 This article reflects the deliberations of the 2012 EES Biennial Conference panel in Helsinki chaired by Derek Poate and funded by the Rockefeller Foundation. It included presentations by Paula Silva Villanueva, Gil Yaron and Robbie Gregorowski. Piva Bell commented on the presentations.
 - 2 Brooks, Nick., Simon Anderson, Jessica Ayers, Ian Burton and Ian Tellam (2011) Tracking adaptation and measuring development. *IIED Climate Change Working Paper No. 1*. IIED

INNOVATIVE APPROACHES TO IMPACT EVALUATION

Elliot Stern, Nicoletta Stame, Kim Forss and Barbara Befani

Two sessions at the EES Helsinki Conference offered a glimpse into the thinking behind a recent report prepared for the UK’s Department for International Development (DFID) *Broadening the range of impact evaluation designs and methods*¹.

The study was expected ‘to test the case that alternatives to experimental and quasi-experimental designs can have equivalent robustness and credibility – they are different but equal’ and to develop a broader range of designs for impact evaluations (IEs). This was against a background of new kinds of complex international development interventions (e.g. governance, post conflict reconstruction and state-building, democratization and climate change mitigation)

which have proved difficult to evaluate with traditional IE approaches.

Traditional approaches to IE such as RCTs tend to search for particular short-term effects and to attribute them to a single ‘cause’ (the intervention). This is suitable for simple interventions in circumstances where a counterfactual logic can be tested and where a narrow evaluation question such as ‘does the program work?’ is being asked. The study looked for evaluation designs suited to other circumstances, e.g. where a larger array of questions more relevant to complex programmes are being probed.

The study drew on philosophy of science findings (causation and explanation) and

social science research methodology and theory as well as on the evaluation literature. It built on OECD/Development Aid Committee definitions of ‘impact’ which is widely shared in the development community. Finally it acknowledged that programme effects are often indirect; that results can be unintended; that contribution can be as important as attribution; and that policy makers need explanation as well as measurement of outcomes if they are to learn.

Linking causes and effects involves what scientists call drawing a ‘causal inference’. From this perspective IE moves evaluation closer to the world of scientific research. Four main approaches to causal inference were reviewed:

- Regularity or statistical association
- Counterfactuals – assessing what would have happened without a programme or policy
- Multiple-conjuncture – where programmes contribute rather than have effects on their own; and
- Generative causation that tries to identify causal mechanisms in context.

Each of these approaches was assessed in terms of its pre-requisites especially in terms of how far a programme is necessary and/or sufficient to generate observed effects. It quickly emerged that all IE designs also have strengths and weaknesses. This is why combining methods is important in IE.

Different types of IE questions were therefore considered, each of which leads to a distinctive IE design. The first is the traditional IE question:

1. To what extent can a specific impact be attributed to the intervention or programme? (experimental logic, mostly quantitative)

The next three are more relevant to IE of complex programmes:

2. Has the intervention made a difference e.g. added value? (contributory cause, mostly qualitative)
3. How has the intervention made a difference? (explanatory analysis)
4. Can this intervention or programme be expected to work elsewhere / when? (generalisability and transferability)

The study analyzed contemporary international development programmes and identified some typical complex programme attributes, e.g. overlapping interventions with similar aims, multiple activities within the same programme, customized non standard activities, programs working indirectly through agents, programmes where impacts were likely to be long term and even intangible. Five IE designs were identified (see box below), that in different ways can help answer particular questions about particular programs.

Five IE Designs

Experimental: RCTs, Quasi Experiments, Natural Experiments

Statistical: Statistical Modelling, Longitudinal Studies, Econometrics

Theory-based

- a) Causal process designs: Theory of Change, Process Tracing, Contribution Analysis, impact pathways,
- b) Causal mechanism designs: Realist evaluation, Congruence analysis

'Case-based' approaches

- a) Interpretative: Naturalistic, Grounded theory, Ethnography
- b) Structured: Configurations, QCA, Within-Case-Analysis, Simulations and network analysis

Participatory

- a) Normative designs: Participatory or democratic evaluation, Empowerment evaluation,
- b) Agency designs: Learning by doing, Policy dialogue, Collaborative Action Research

Designs have to be matched to evaluation questions and program attributes; and to different programme stages and different programme levels. Multi-level programmes may require multi-level 'nested' designs. The study also emphasized the limited range of IE designs and methods now in use and the urgent need to test innovative methods in real world settings.

Highlights of the Helsinki Conference debate triggered by the above findings included:

- **The challenge of external validity:** Tackling this challenge is one reason why alternatives to experiment are needed. A "rigorous" experiment can show the effects of an intervention in a specific situation (here, for us, in these circumstances). RCTs rank high on internal validity – but say little about what works in another time and place – external validity. Yet policy learning requires generalization which in turn depends on understanding the mixes of contextual and policy elements that work in different situations; and building typologies of contexts and mechanisms. Moreover, policy learning is not simply transfer for use elsewhere.
- **What is meant by rigorous methods?** Every methodology has its own quality criteria. But there are also process requirements for rigour: a method should be systematic, logical, transparent, documented and open to scrutiny.

- **What is impact?** Most participants expressed views about impact that challenge the value of a simple indicator of success, e.g. a programme may reduce a threat (e.g. violence against women) which goes beyond ascertaining a positive outcome. Impact depends on the nature of the phenomenon: it is a reflexive outcome: 'you change me, I change you'. It is the combination of factors (elements of a causal package), interacting among each other. It also emanates from partners countries collaboration. What beneficiaries experience or do may differ from what international agencies intended.

- **How to develop capacity?** There are many actors in this drama! Capacity requires mastering a variety of research methods needed to apply a chosen design. Can the single evaluator possess all necessary skills? To what extent are teams essential and how can academic resources improve individual and team competencies? Capacity goes beyond the evaluator: evaluation commissioners have to be sensitive to the range of IE questions that need to be answered; and be flexible enough to recognize 'emergent' programme properties in order to adapt designs. How beneficiaries participate in the evaluation, will shape much of the impact – all actors have 'agency'. The evaluator has to facilitate this overall process.

- **The UK policy context:** Andrea Cook, of DFID, emphasized that whilst the UK aid budget has increased, there is intensive scrutiny of development policy in the media, hence the government's interest in development results. A new "Independent Commission on Aid Impact" now reports directly to Parliament. Programs have become more ambitious, and the program portfolio is more diversified. Development evaluators work in highly complex environments, where more people participate, which makes controlling implementation difficult. This is ultimately why the study was commissioned.

-
- I The study team, led by Elliot Stern, included Nicoletta Stame, Kim Forss, Barbara Befani, John Mayne and Rick Davies. See: <http://www.dfid.gov.uk/Documents/publications/design-method-impact-eval.pdf>

CONTRIBUTION ANALYSIS AT THE HELSINKI CONFERENCE

Jacques Toulemonde

Contribution Analysis (CA) implements the principle of theory based evaluation through a pragmatic six-step approach. First, it sets out the cause-and-effect issue to be addressed. Second, it puts forward a postulated theory of change. Third, it gathers evidence in order to confirm or refute the theory. Fourth, it develops a series of contribution claims that make up a contribution story. Fifth, it identifies and bridges the remaining gaps in the evidence base. Sixth, it revises and finalizes the contribution story. This approach ensures that all causal links of the theory of change are examined, reports on whether the intended changes occurred or not, and analyses the main contributing causes to such changes, including the programme under evaluation.

Over the last ten years, CA has attracted visible interest in international events, including the Prague 9th Biennial Conference in 2010. However, even by then, instances of rigorous CA implementation were surprisingly scarce and the theoretical foundations of the approach remained fragile.

In July 2012, a special issue of 'Evaluation – The Journal of Theory, Research and Practice' was devoted to CA¹. This issue consisting of eight articles was edited by John Mayne, who had initiated the idea of CA over a decade before. The preparation of the Special Issue offered a unique opportunity for reflection and sharing of practical experiences. Highlights of the Special Issue were presented at the Helsinki EES Conference in 2012 in a panel session gathering several of the authors: Sebastian Lemire (Ramboll Management), Tamara Mulherin (Results LAB), and Jacques Toulemonde (Eureval and Lyon University), as well as Rob van den Berg who is using CA at the Evaluation Office of the Global Environment Facility (GEF). The session attracted 70 participants and triggered lively discussions summarised below.

Strengthening theoretical foundations

In the Special Issue, John Mayne and other authors undertook to clarify and deepen

some key concepts on which the approach is based. A first example relates to eliciting the *theory of change* (Step 2 of CA). It was acknowledged that the analyst cannot just rely on a logic model displaying the intended effects and the mechanisms through which they are meant to derive from one another. For every link in the causal chains, the analyst should also identify other influential factors, plus feed-back loops and alternative explanations. Refining the theory of change into such complicated causal packages is the price to pay for identifying all the contributory causes that need to be analysed.

Furthermore, the findings of a CA are drafted (Step 4) and finalised (Step 6) in the form of a contribution story that assembles a series of *contribution claims* that follow the successive links of the causal chains. For instance, a given contribution claim may state that A and B did occur, that A contributed to B in conjunction with X, Y, and Z, and that A was the second contributing cause of B by order of influence. In the same analysis, some contribution claims may be supported by statistical analyses and others may derive from case studies or from other analytical methods. Since various methods and tools may be used in the same analysis, CA cannot be considered as a method or tool on its own, but should rather be seen as a framework approach.

A given contribution claim may rely on an attribution analysis involving a counterfactual, but other claims in the same analysis may just derive from an inquiry into contributing causes that are deemed to be the particularly influential. On the other hand, the contribution story as a whole is free of any counterfactual reasoning. In this sense, attribution analysis and contribution analysis are truly alternative options.

Towards setting quality criteria

In his paper, Michael Patton concludes that the quality of a CA cannot be a matter of methodological standards and validity thresholds but rather one of *rigorous thinking*

secured by an *open critique*. Building on all papers of the Special Issue, it can be said that the rigor of the analysis may be assessed by asking at least four sets of questions:

- Was the theory of change elicited appropriately? Is it in line with available knowledge and free of logical gaps? Does it pay sufficient attention to influencing factors and alternative explanations?
- Have the evidence gathering efforts been sufficient? Does the evidence cover all links of the causal chains? Does it cover all influencing factors and alternative explanations identified in the theory of change?
- Are the contribution claims credible enough? Do they hold up in terms of traceability, appropriateness and validity of the arguments? Do they rely on judicious triangulation of information sources?
- Has the draft contribution story been submitted to a proper critique? Was the critique thorough, timely and open?

These questions sketch the set of criteria that will have to be developed in order to assess the quality of the approach. A pending issue is the extent to which quality benchmarks could be set for such criteria in the future.

Delineating the area of relevance

The Special Issue describes 14 examples of CA implemented in various settings (governmental, non-governmental, and international organisations). In all these examples, useful lessons have been learned from impact analyses in circumstances where no credible counterfactual could have been developed. For example, Michael Patton shows how CA was used in the evaluation of the Paris Declaration a landmark international agreement aiming to improve the effectiveness of development aid. The International Reference Group supervising this worldwide evaluation endorsed CA as a credible approach to analysing impacts. Moreover, the Group members engaged in thorough discussions and constructive criticism on contribution claims.

In Helsinki, Rob van den Berg exemplified the area of relevance of CA by showing that GEF applies this approach to evaluating relatively small interventions that are intended to play a catalytic role at the first of many steps towards broad objectives set at a high aggregate level. In other words, CA is used for assessing whether up-scaling and mainstreaming occurs, i.e. whether the intervention acted as the spark that lighted the fire. This illustrates the potential of CA for analysing impacts in the case of long causal

chains, cascading mechanisms affecting successive groups or systems, and shifts from micro to macro level.

CA also has a potential to analyse impact which cannot be subjected to an attribution analysis because the intervention was not implemented in a stable manner, because causal mechanisms and contextual factors are too intricate, because available data do not reach a satisfactory statistical threshold, or simply because some key changes are not measurable.

A comparison of the debates held on CA in Helsinki (2012) and Prague (2010) shows that the approach has gained considerable credibility, clarity, and maturity over the past two years. In other words, and as Mayne stated in the Special Issue, “CA is coming of age”.

I Evaluation, *Special Issue: Contribution Analysis*, Sage, Volume 18, Number 3, July 2012

PAYMENT BY RESULTS: DOES IT WORK?

Burt Perrin and Peter Wallace

There is no agreement about a single definition of payment by results (PBR) but it usually refers to a form of financing that makes payments contingent on the independent verification of results. It has three key elements: (i) payments based on results; (ii) recipient discretion regarding *how* results are achieved; and (iii) verification of results as the trigger for disbursement. To learn from experience, the Evaluation Department of DFID commissioned an independent study of PBR in development assistance¹. The findings were presented at the 2012 EES biennial conference in Helsinki. This article presents a summary.

The study had three main objectives:

- To identify and synthesise evidence, to the extent possible, from evaluations of PBR approaches in development.
- To provide an analytical critique of the quality of existing evaluations.
- To provide guidance for approaches, including evaluation questions and methods, to future evaluations of PBR programmes.

Evidence base

Most readily identified research and evaluation studies of PBR reviewed by the study were in the health sector. They almost invariably aimed at enhancing incentives to service provider organisations and individuals rather than governments. The importance of an outcome (or results) orientation that focuses on actual benefits rather than inputs

and outputs (i.e. services provided) is uncontested. On the other hand, the evidence regarding the potential of PBR incentives to change professional practice is weak. Perhaps the most optimistic conclusion that can be drawn is that contracting out may increase verifiable short-term access and use of health services. Improved health outcomes are more elusive given other influences and unintended effects.

All in all, there is limited evidence to date that PBR approaches offer value-added compared to other modalities. Implementation of PBR has encountered significant challenges. Attention to such basic questions as the efficacy of incentives mechanisms, cost effectiveness relative to other approaches, impact on equity, sustainability, etc. has been limited. What emerges strongly from the evidence base is that PBR approaches never operate alone, but as part of a package of increased funding, technical support, training, new management structures and monitoring systems, and often in the context of a significant reform effort. Furthermore, the diversity and complexity of interventions reviewed precluded meaningful generalisation.

Quality of PBR studies

Practically all the “evaluations” identified were carried out by people who identify themselves as researchers. The studies did not comply with the OECD/Development Assistance Committee (DAC) Evaluation

Criteria or the Evaluation Quality Standards² that DFID along with other bi-lateral aid agencies have endorsed. In particular, there was limited attention to the five DAC evaluation criteria, with the exception of achievement of objectives (effectiveness).

Similarly as research, the quality of studies was considered poor by major systemic reviews and critiques. It proved impractical to implement sophisticated experimental or quasi-experimental designs. The resulting biases compromised the ability to draw valid conclusions from the data collected. Alternative evaluation approaches better suited to evaluation of complex situations, e.g. realist evaluation or contribution analysis, were not attempted. In order to be able to apply or adapt findings from one setting or situation to another, it would have been necessary to unpack and test the mechanisms at work.

Guidance for future PBR evaluations

The most important outcome of the study was to provide guidance for future PBR evaluations.

First, before considering potential methodological options, they should identify *the questions* that need to be addressed. Second, they should identify the mechanisms and the sets of circumstances under which PBR approaches are expected to make a positive difference. Third, attention should be

devoted to cost effectiveness issues. Fourth, systematic comparison with other potential approaches and strategies and their effects on incentives should be probed. Fifth, evaluation should explore and describe the *process* by which PBR initiatives are implemented in practice, and the reasons why changes from the original conception may be needed. Sixth and finally, un-intended effects and the means of minimizing them should be ascertained. This would likely lead to recommendations about improved PBR designs. PBR initiatives beyond the health sector as well as schemes aimed at governments also deserve to be evaluated.

No single method is “best”. There are significant opportunities for theory-based approaches that can identify and document the mechanisms at play. Realist evaluation approaches that seek to identify what works

for whom in what circumstances may be particularly suited to evaluation of PBR schemes. A mixed method approach is probably optimal.

In all cases, articulation of a theory of change would aid in the selection of evaluation questions. Given the complex context in which PBR schemes operate and the multiplicity of explanatory factors, a contribution analysis approach rather than a linear cause-and-effect model is likely to be the most appropriate. Preference should be given to evaluation approaches that can best inform policy and programming decisions on a timely and cost-effective basis.

Conclusion

PBR may be an appropriate mechanism in certain circumstances. However, it is unlikely to be effective in all situations. The val-

ue of PBR is, as of now, unproven. A broader range of research and evaluation approaches and tighter quality assurance arrangements should be adopted so as to reach more definitive conclusions about the potential of PBR to improve aid delivery.

1 The full report is available at: <http://www.dfid.gov.uk/Documents/publications/evaluation/payment-results-current-approaches-future-needs.pdf>.

2 OECD DAC Network on Development Evaluation. *Evaluating Development Co-operation: Summary of Key Norms and Standards*. Available at: www.oecd.org/dac/evaluationnetwork

TOWARDS MORE USEFUL COMPREHENSIVE EVALUATIONS OF INTERNATIONAL ORGANIZATIONS

Rob D. van den Berg and Robert Picciotto

Comprehensive evaluations focus on the overall performance and achievements of an organization, corporation or network. They have been especially prevalent in the international development domain. Costing anywhere between \$0.5 million and \$18 million and taking 2–5 years to complete, comprehensive evaluations have frequently been commissioned on an *ad hoc* basis by major stakeholders with the intent of influencing decision-making, improving resource allocation or secure ‘value for money’ for funding commitments.

About 17 comprehensive evaluations have been carried out in the last 10 years at a cost of more than \$30 million. A workshop held in Paris in June 2012 brought out important lessons based on desk reviews of readily available comprehensive evaluations and detailed case studies. It emerged that most comprehensive evaluations have proved use-

ful and influential internally and they have generated value for money, especially for their most powerful sponsors.

Comprehensive evaluations are sometimes a recurring event but more often than not they are commissioned as a ‘one-off’ initiative in response to an organizational crisis, an external shock or a shift in stakeholders’ priorities. Some development agencies escape comprehensive evaluations altogether. Others may be subject to reviews that reflect the strong policy predilections of one or more powerful stakeholders. As a result they may lack rigor, balance and/or legitimacy.

An overview of workshop findings was presented at a panel session of the EES 10th Biennial Conference that took place in Helsinki in October 2012. A major theme of the presentations was that the usefulness and legitimacy of comprehensive evaluations

have left a great deal to be desired. Most comprehensive evaluations have failed to reflect widespread public policy concerns regarding such well documented dysfunctional features of the development system as its increased fragmentation, rising transaction costs, poor coordination among development actors, inadequate harmonization of aid delivery processes, lack of alignment with developing countries’ priorities and processes, etc.

Similarly, there has been limited resort to benchmarking analyses due in part to lack of comparable data and disparate ways of assessing performance. Yet donor countries’ funding decisions require comparative performance data. Given the demand for more rigorous resource allocation decisions at a time of aid budget scarcity, a cottage industry of league tables produced by think tanks and individual donor development agencies has materialized in order to fill

the need. But here again questions have arisen regarding the rigor of their evaluative content and the validity of the resulting rankings.

In a nutshell, the weaknesses of the current comprehensive evaluation system are to a large extent caused by collective action dilemmas. Individual organizations tend to commission comprehensive evaluations to match the distinctive perspectives of their most influential owners and/or to respond to internal management needs. Silo thinking is the inevitable consequence so that individual comprehensive evaluations are conceived as *ad hoc* and unique with the resulting tendency to 'reinvent the wheel' and to avoid learning from previous evaluations.

The Paris workshop concluded that "guidance" on comprehensive evaluations would help improve the relevance and quality of future comprehensive evaluations. Such guidance would be directed at governing bodies, evaluators and stakeholders. It would disseminate knowledge and identify good practice. The Helsinki panel presentations and the debate that ensued focused on what this guidance might look like and how to achieve greater rigor and legitimacy.

Derek Poate highlighted the top-down nature of many comprehensive evaluations and their failure to achieve country level ownership of evaluation processes and find-

ings. He also pointed to the need to address corporate governance issues, to ensure independent quality assurance and to resort to benchmarking. He favored the linkage of comprehensive evaluations to decision making geared to organizational reform processes or replenishments in order to facilitate use of evaluation findings.

Bob Picciotto focused on the international context in which comprehensive evaluations take place. He referred to the asymmetry of power relationships embedded in comprehensive evaluation designs. While comprehensive evaluations should be tailor made to the challenges faced by individual organizations they should also comply with Paris Declaration principles. The current failure to reflect those principles in comprehensive evaluations has undermined their value and credibility.

Elliot Stern stressed the need for flexibility in comprehensive evaluations. In his view achieving comparability was desirable but intrinsically hard due to the different size, nature and mandates of international organizations. He therefore saw room for "action research" oriented comprehensive evaluations specifically geared to finding pragmatic solutions for emerging challenges faced by each organization.

Caroline Heider noted that transnational partnership programs were increasing

in number and influence and that they were in dire need of comprehensive evaluations. The World Bank's Independent Evaluation Group has published a source book to help guide such evaluations. They draw on international good practice and they have been disseminated by the DAC evaluation network. They need updating and in the process they should take on board lessons from comprehensive evaluations.

Indran Naidoo laid stress on the issue of subsidiarity. Comprehensive evaluations should be conceived as ways of gaps not filled by existing independent evaluations. Hence, they should be conceived and planned in the context of existing evaluation programmes within the overarching development evaluation architecture. In turn, this implies sound evaluation foundations grounded in solid and independent internal evaluation systems.

Finally, Rob D. van den Berg concluded and pointed Conference participants to the recently established community of practice on comprehensive evaluations. The GEF Evaluation Office hosts an online Comprehensive Evaluation Platform for Knowledge Exchange (CEPKE). The CEPKE (www.cepke.org) went live in October 2012. It allows evaluation specialists and other interested professionals to share ideas, information, and experiences related to carrying out comprehensive evaluations of institutions.

VALUING EVALUATION POWER AND THE POWER OF EVALUATION IN “SPEAKING TRUTH TO POWER”

Sandiran Premakanthan

“The truth that makes men free is for the most part the truth which men prefer not to hear”. Herbert Agar

The phrase “speak truth to power” was popularized by the Quakers during the mid-1950s. But its origins go much farther: Moslems assert that the *“Prophet Muhammad said that the best form of jihad is to speak truth to power”* while Jews *“are commanded by Torah to speak truth to power”*. As evaluators what does this slogan mean to us, our profession and our practice?

By speaking evaluation truth to power the public interest is enhanced. But this assumes that the power of evaluation is judiciously exercised. Valuing global evaluation power requires estimates of the volume of evaluation carried out and its effects. Such an exercise might be applied to governments, philanthropic foundations, financial institutions, government aid agencies, United Nations (UN) agencies, non-governmental organizations (NGOs) as well as to policy networks, both national and international.

Evaluation power may be defined as the ability or official capacity to exercise control or authority on the dollar investment in evaluation infrastructure or capacity building to meet policy requirements and evidence gathering to continuously measure, monitor and evaluate for informed decision making in program expenditure management (accountability) or strategic uses. The *power of evaluation* is a measure of the wealth of performance results: measurement, monitoring and evaluation evidence created for use by the evaluation community, the individuals and institutions vested with evaluation power, to inform and influence program, policy, and resource allocation and reallocation decision makers by speaking truth to power.

Use of the “value model” is illustrated by the evaluation power actually exercised within the government of Canada. An estimate of the supply of measurement, monitoring and evaluation services in Canada is displayed below using the most recent Treasury Board Secretariat (2011) report on the evaluation function.

| Resource Category | 2007–08 (\$ millions) | 2008–09 (\$ millions) | 2009–10 (\$ millions) | 2010–11* (\$ millions) |
|-------------------|--------------------------|--------------------------|--------------------------|---------------------------|
| Total Resources | 57.3 | 60.9 | 66.8 | 67.4 |
| % Annual Increase | N/A | 6.3 | 9.7 | 0.9 |

Table 1. Financial Resources Expended on the Evaluation Functions of Large Departments & Agencies (LDAs) in the Government of Canada From 2007–08 to 2010–11.

The investment of \$67.4 million in 2010–11 resulted in acquiring and maintaining an evaluation capacity federal government wide equal to 497 FTEs (Full-time Equivalent – Table 2).

| Evaluation Resources | 2007–08 | 2008–09 | 2009–10 | 2010–11* |
|------------------------------|---------|---------|---------|----------|
| Full-Time Equivalents (FTEs) | 409 | 418 | 474 | 497 |
| % Annual Increase | N/A | 2.2 | 13.4 | 4.9 |

Table 2. FTEs Working in Evaluation in the Government of Canada From 2007–08 to 2010–11.

The federal government evaluation capacity supplemented by contracted resources (consultants) provided annual evaluation coverage of 6.2% (2010–11) of direct program spending (\$9.93 billion) when compared to 14.2% coverage in 2009–10 (Table 3).

| Direct Program Spending & evaluations by Fiscal Year | 2007–08 | 2008–2009 | 2009–2010 | 2010–2011 |
|---|---------|-----------|-----------|-----------|
| Total Number of Evaluations | 121 | 134 | 164 | 136 |
| Direct Program Spending Covered by Evaluations (\$ millions) | 5,041 | 5,879 | 11,999 | 6,607 |
| Total Direct Program Spending From Main Estimates (\$ millions) | 77,617 | 79,327 | 84,665 | 99,325 |
| Annual Evaluation Coverage (%) | 6.5 | 7.4 | 14.2 | 6.7 |

Table 3. Evaluations of Federal Program Spending in Large Departments & Agencies (LDAs) From 2007–08 to 2010–11.

The evaluation coverage requirements of section 6.1.8 of the 2009 Policy on Evaluation demands a 100% coverage of all direct program expenditure of \$99,325 million or \$9.9325 billion (2010–11 figures in Table 3) on a five year evaluation cycle beginning in 2013. This means the Canadian federal government's evaluation power must meet the demand for annual evaluation coverage of 20% of direct program spending.

Table 3 suggests that evaluation coverage has fallen from 14.2% (2009–10) to 6.7% (2010–11) of the potential demand. The 2012 annual report predicts that the annual coverage rate will rebound significantly in 2011–12. But in a climate of government deficit reduction requiring improved effectiveness of scarce public resources, the question arises as to whether the current evaluation power, \$64.7 million with an annual average increase of 5.63% is adequate.

On the demand side, the model attempts to value evaluation information by its use for internal program management and for strategic purposes. For example, the 6.7% evaluation coverage in 2010–11 produced 136 evaluations. The annual reports rated the quality and use of the evaluations as “strong” or “acceptable” (over 85%).

The value of the use of the evaluation power from the 136 evaluations could be assessed from the contributions and direct attribution to expenditure management decisions. The TBS annual reports (2011) and (2010) claim that Treasury Board Secretariat analysts use evaluation evidence when examining and providing advice on funding proposals.

What would be the optimum dollar value of investment to meet the evaluation coverage requirements of section 6.1.8 of the 2009 Policy on Evaluation? Systematic use of the value model would provide “what if” scenarios for different levels of evaluation power. The Canadian government value model suggests an average evaluation FTE standard cost range of \$135K–\$140K and evaluation coverage range of 18–20 million per evaluation FTE.

However, the optimum would be dependent on validated standards of average cost of an evaluation FTE and of evaluation coverage. Based on the value model, the return on investment would hinge on the actual use of performance results evidence for informed decision making on policy, program expenditure management and program improvement including Cabinet Committee decisions and Parliamentary Reporting. Research is needed

to determine the true value of the use of evidence in program expenditure management and strategic decision making.

Estimates of the value of investment in evaluation compared to other uses are also required and this would vary according to the context. A more distant step would be to conduct international comparisons of evaluation power in both public and private sector institutions to gather evidence for setting international bench marks. The value model could estimate the evaluation power requirements for various public and private institutional and country scenarios for evaluation coverage of direct program expenditure requirements.

While further research is needed it is safe to assume that the return on investment from the evaluation function in any organizational setting is dependent on the value it adds in providing products and services, the right quantity and quality of advice, at the right time, to the right people all the time. For this to happen, the evaluator facilitated by an organization culture that promotes openness, frankness, honesty and truthfulness and supported by norms, policies, systems and processes should “seek the truth and fearlessly speak truth to power”.

STRENGTHENING EVALUATION FOCUSED PROGRAMMES IN EUROPEAN UNIVERSITIES

The USPE Network¹

Post-graduate programmes in evaluation began to emerge across Europe at the end of the 1990's. Since then, at least one new programme was created every year. About fifteen European countries now offer at least one evaluation diploma, in addition to the plethora of smaller modules dedicated to evaluation as part of other university programmes. This trend demonstrates that European universities are increasingly interested in and committed to delivering diplomas in evaluation, even if the current public finance crisis has led to the interruption of some programmes, notably in the UK.

At the London Conference of the EES (2006), the leaders of six university programmes came together to form an informal network – the University Study Programmes in Evaluation (USPE), which sponsored panel sessions on evaluation education at subsequent EES conferences in Lisbon, Prague and Helsinki (2008 to 2012). At all of these events it was possible to draw on Bern University surveys of the European study programmes in evaluation.

Survey responses from eleven programmes in nine countries were summarised at the Helsinki Conference². During the Conference, several other programmes expressed their willingness to join the survey which now covers fifteen master-level programmes in twelve countries (Belgium, Denmark, France, Germany, Greece, Italy, Netherlands, Romania, Spain, Sweden, Switzerland, and UK)³. Most of these programmes are devoted to policy and programme evaluation.

At the Helsinki Conference, the issue of education in evaluation at universities was discussed in a panel session called “*Building evaluation capacity through university programmes: where are the evaluators of the future?*” Three successive presentations followed by debates addressed the following issues: (i) audience of the programmes; (ii) training approaches; and (iii) prospect for Europe-wide cross-fertilization. The following lines summarise these debates and the related works of the USPE Network.

What are the current audiences of evaluation focused programmes?

Overall, the fifteen surveyed programmes deliver about 200 diplomas per year. Seven programmes are recruiting mid-career professionals only and the eight other ones are also targeting young students. Four programmes are multi-disciplinary while others have specific connections with academic disciplines (economics, education, management, policy analysis, public administration, social research), policy areas (development, education, health), or neighbouring professions (monitoring, policy design, financial analysis). The overall picture is that evaluation focused programmes deliver a small number of diplomas/degrees which is proportionate to the small ‘niche’ that evaluation activities occupy in the public sector and the still limited professionalization of these activities.

What are the training approaches?

The training methods are predominantly academic (thesis, papers, and seminars). However, seven programmes also call on their participants to undertake real life tasks such as doing a small evaluation project on their own, carrying out an empirical research, doing a meta-evaluation, writing the terms of reference for an evaluation, competing for public procurement of external evaluation services, identifying and discussing ethical challenges, and / or assessing the methodological quality of an evaluation. Three programmes include an internship (up to five months).

The survey does not show the extent to which the programmes are based on professional competencies or capabilities. This is however an area where the members of the USPE Network have worked hard over the last five years. Building upon a literature review, they recognised that there is a broad international consensus about evaluation competencies so that a distinctly European approach would not end in something different from what is being done

in other continents. On the other hand, they admitted that none of their six programmes could reasonably claim to cover the whole range of evaluation competencies. They also acknowledged that each university needs to freely develop its own priorities and learning objectives in a competitive environment and in its own national context.

Conversely, the USPE members considered that evaluation focused programmes should refer to a set of “core competencies or capabilities” that are both essential for and specific to the evaluation profession. In this respect, they established a first version of thirteen core competencies while recognising that there is still a long way to go before the profession could reach a consensus on that issue. Moreover, there is an even longer way to go before universities become capable to confirm that their graduates have a sufficient mastery of core competencies or capabilities in a harmonised manner across Europe.

What is the prospect for cross-fertilization across European programmes?

While preparing this paper, the USPE members acknowledged that (1) there is a growing demand from all around Europe for opening and enlarging their network and (2) the EES has established a Thematic Working Group (TWG) on the Professionalization of Evaluation having four priorities, one of which being “expanding the supply of high quality evaluation education and training”. It was then decided to invite all members of and applicants to the Network to join the existing ‘Professionalization TWG’, to create a sub-group on education and training within that TWG, and to stop the USPE Network. Hence, the issue of cross-fertilization among European universities will be discussed within that new framework.

Looking ahead, it is clear that the biannual survey of Bern University is a valuable tool that deserves to be continued and amplified in order to monitor and advertise the supply of education and training across Europe. The idea

of sharing experience in the area of teaching and training methods could also be considered (e.g. sharing case studies, textbooks, fundamental references). However, the prospect for such exchanges is limited for two reasons. First the programmes tend to be delivered in the national languages (among the surveyed universities, only one non-UK programme is delivered in English). Second, there is a wide diversity of connections with academic disciplines, policy areas, and neighbouring professions as noted above. For the same reasons there is probably limited room for exchanges of students and visits of teachers⁴.

On the other hand, the authors of this article consider that European cooperation would be very useful in developing competency or capability standards as well as methods for assessing these. In this area, universities are

on the supply side of the labour market and they could usefully cooperate with those who hire their graduates on the demand side of the market.

1 The University Study Programmes in Evaluation (USPE), involves programmes located in Bern, London, Lyon, Madrid, Odense, and Saarbruecken. The leaders of these programmes jointly authored this article: Georgie Parry-Crooke, London Metropolitan University; Peter Dahler-Larsen, Suddansk Universiteit, Odense; Maria Bustelo, Universidad Complutense de Madrid; Verena Friedrich Universität Bern; Jörg Rech Universität des Saarlandes; Jacques Toulemonde, Université de Lyon.

2 Friedrich, V. & Beywl, W. (2012). European university-based study programmes in evaluation: Eleven profiles. University of Bern: Centre for University Continuing Education. (The document can be requested by contacting verena.friedrich@zuw.unibe.ch).

3 All programmes targeting students are called 'Master' with the meaning of a post-graduate Bologna-like diploma. Things are less clear for programmes targeting mid-career professionals which may have other names, e.g. 'Certificate', 'Diploma', or 'Master' of Advanced Studies in Switzerland bearing 15, 30 and 60 European Credits respectively. The survey also covered one programme at PhD level.

4 Only one such exchange was experienced within the USPE network.

OBITUARY: CAROL WEISS

Robert Picciotto

The evaluation community has suffered a huge loss. Carol H. Weiss died on Tuesday, January 8, 2013 at the age of 86. She was born and raised in New York, graduated from Cornell University and earned an M.A. and a Ph.D. at Columbia University. She devoted her entire professional career to evaluation and left behind an extraordinary legacy. Her contributions span all facets of evaluation theory and practice.

At the Bureau of Applied Social Research at Columbia University, and later at Harvard, she conducted groundbreaking social research studies geared to public policy decisions. She had a solid grasp of the realities of policy making and stood firm on the critical need for strictly objective, evidence based evaluation work. She harboured no illusion about the limits of instrumental use of evaluation findings but took pains to highlight the myriad other ways of achieving results through principled evaluation.

Through numerous books and articles as well as through generous participation in workshops and conferences she challenged all evaluators to achieve excellence and influence. In a field fragmented by competing conceptions of evaluation Carol Weiss always stood above the fray. While she never engaged in petty doctrinal disputes she did not refrain from engaging in fulsome debate regarding important evaluation policy issues— and she did so with grace, subtlety and civility. She always looked for common ground

through patient dialogue across the multifaceted evaluation world.

Carol Weiss reached out to all the social sciences. By pioneering new approaches and crafting new concepts in programme evaluation, theory based approaches, qualitative and quantitative methods, evaluation use, etc. she vastly expanded the boundaries of the evaluation discipline and enhanced the stature and utility of our fledgling profession. She won many awards, including the Myrdal Award from the Evaluation Research Society as well as fellowships at the Centre for Advanced Study in the Behavioural Sciences and the Brookings Institution.

Carol Weiss won the admiration, respect and affection of all evaluators who read her, met her or debated with her. According to Michael Patton, “Evaluation lost one of its great pioneers when Carol Weiss passed away... Dialogues, discussions and, yes, debates, with Carol Weiss over the years have had an enormous impact on the field”. For Michael Scriven “a special virtue, besides the content of her work, was her respect for civility”... Eleanor Chelimsky will miss “her calm spirit of inquiry, her knowledge, her capacity to see other points of view”.

For Stewart Donaldson, she “was a profound and influential thinker about evaluation theory and practice and an exemplary role model”. Patricia Rogers will remember “not only her contributions but also

her generosity of spirit”. Finally for Jane Davidson “this is a huge loss for evaluation. Her brilliant ideas live on” and for Lois-ellin Datta Carol Weiss is still “here through her influence on our thoughts including the cascading discussions, debates, theories from her path-breaking 1970 article on politics and evaluation”.

Carol Weiss travelled, consulted and advised widely and frequently. Her empirical contributions to evaluation practice are legion. She worked with dozens of U.S. government agencies, international agencies, and governments in Africa, Asia, Australia, and Europe in fields as diverse as international development, public health, crime and delinquency, agriculture, mental health, and education. She greatly enjoyed the rough and tumble of evaluation practice.

When as Director General, Evaluation of the Independent Evaluation Group I was faced with a particularly complex evaluation challenge (the assessment of a community based poverty reduction initiative) I turned to Carol Weiss and she insisted on working directly with staff in order to assess the operational context, unearth the hidden assumptions and identify the key hypotheses underlying the programme. She exuded curiosity and quiet enthusiasm. She worked hard as part of the evaluation team and she won everyone over given her deadpan sense of humour, her infinite patience and her modesty. I too will miss her.

PROPOSAL FOR A NEW THEMATIC WORKING GROUP ON PUBLIC PRIVATE PARTNERSHIPS

Mehmet Uzunkaya

Given the ever increasing constraints on government budgets and growing demand for better infrastructure, public-private partnerships (PPPs) have become a widespread method of infrastructure service provision throughout the world for the last couple of decades. On a global scale, the total investment commitment of infrastructure projects with private participation was more than 1.8 trillion US dollars between 1990 and 2011¹.

PPPs are long-term contractual agreements between public and private sector that relies on the notion of combining and utilising respective strengths, different functions and comparative advantages of the two sides and sharing the risks associated with a public sector project between the partners, in such a way that each party assumes the risks for which one is better in handling compared to the other.

While PPPs offer promising benefits and have development impact potential, they also include complex multi-disciplinary processes and stages as compared to conventional procurement methods. This makes them subject to a multiplicity of risks, mismanagement of which may jeopardize all envisaged benefits and result in sub-optimal resource use. PPPs also include many different stakeholders with considerably varying objectives and motives, all aiming to maximise their own welfare. In addition, the long-term nature of PPP contracts raises inter-generational equity issues. These factors necessitate careful evaluation of PPP projects and programs, ex-ante and ex-post, from different viewpoints in an interdisciplinary way so as to realize their development potential and to avoid sub-optimal resource use, now and in the future.

In contrast to their development potential, widespread use, interdisciplinary character

and stakeholder variety, however, evaluation of PPP projects and programs have not been given its deserved attention so far, especially in developing countries. In such an environment, the lack of high quality ex-ante and ex-post evaluations of PPP projects and programs may result in recursive PPP arrangements, which are far from satisfying critical stakeholders' objectives.

EES members interested in joining a TWG on PPPs should contact me and the secretariat (secretariat@europeanevaluation.org). If it turns out that there is sufficient interest, I would be happy to be the coordinator.

My email address is:
mehmet.uzunkaya@kalkinma.gov.tr.

¹ World Bank and PPIAF, PPI Project Database. (<http://ppi.worldbank.org>)

EES 'PUBLIC HEARING' AT THE EU PARLIAMENT

The European Evaluation Society is proud to inform you that a 'Public Hearing' on "Evaluation in Democracy", sponsored by the EES and hosted by Tarja Cronberg, Member of the European Parliament, will be held on April 10, 2013 at the European Parliament Building in Brussels.

The main purpose of this one-day workshop is to stimulate debate and innovative thinking about the potential role of evaluation in the European Parliament and the European Commission. The event is intended to encourage collaboration among European evaluation institutions and the expanding European evaluation community of practice.

We will begin by probing the meaning of democratic evaluation. The debaters are two emi-

nent EES members (Elliot Stern, former EES President and Editor of the Journal Evaluation and Bob Picciotto, EES board member).

This will be followed by a session on the twin objectives of evaluation – accountability and learning. It will allow participants to explore issues raised in the EES Brussels event of 2011 when Nicoletta Stame stated that *"the current crisis calls for a change in attitude: public administrations should 'learn to learn', evaluation should become part of the business of government, and it should be incorporated into a 'responsible' public administration"*. The debaters are two former EES Presidents: Murray Saunders and Ian C. Davies.

The morning session will close with a panel, chaired by our host, Tarja Cronberg, MEP.

It will feature brief presentations of evaluations carried out in, on and by the European Parliament. The discussants will be representatives of STOA (Science and Technology Options Assessment), the Policy department and Library of the European Parliament.

The focus of the afternoon sessions will be on the use of evaluation within the Commission, including assessments of EU's regulatory performance. A roundtable will be chaired by Karin Attström, EES board member during which representatives of different evaluation departments will interact with the audience.

The one-day event will close with an open discussion regarding the future of evaluation. For more information please visit www.europeanevaluation.org.

EVALUATION IN DEMOCRACY

A 'PUBLIC HEARING' of the EUROPEAN PARLIAMENT HOSTED BY TARJA CRONBERG, MEP AND SPONSORED BY THE EUROPEAN EVALUATION SOCIETY (EES)

Purpose: To explore the potential of evaluation in the democratic process through enhanced accountability and organizational learning.

Audience: MEPs, Evaluation experts of the EU institutions, EES members.

Date: 10 April, 2013.

Place: European Parliament, ASP I E I, Rue Wiertz 60, B-1047 Brussels

PROGRAMME

8.30–9.15

Registration

9.15–9.45

Setting the stage

Tarja Cronberg,
MEP

Maria Bustelo,
EES President

Moderator: Liisa Horelli,
EES Board member

9.45–10.45

Session 1.1: What is democratic evaluation?

Two perspectives from Elliot Stern,
former EES President and Bob Picciotto,
EES Board member

Audience debate

Moderator: Claudine Voyadzis,
EES Vice President and President Elect

10.45–11.00 Refreshment break

11.00–12.00

Session 1.2: How can evaluation help strengthen accountability and learning in the European space?

Murray Saunders,

former EES President, on Accountability

Ian C. Davies,

former EES President, on Learning

Audience debate

Moderator: Maria Bustelo,
EES President

12.00–13.00

Session 1.3: Is the European Parliament benefiting from evaluation?

A Round Table with representatives
from Science and Technology Options
Assessment (STOA), Policy Department
and the Parliament library

Moderator: Tarja Cronberg, MEP

13.00–14.00 Lunch break

14.00–14.30

Session 2.1: Highlights of morning sessions

Barbara Befani and Liisa Horelli,
EES Board members

14.30–14.45 Refreshment break

14.45–16.00

Session 2.2: The New European Commission Communication on Regulatory Fitness – Implications for Evaluation in the European Space

Round Table: European Commission
representatives (tbc)

Audience debate

Moderator: Karin Attström

16.00–17.00

Session 2.3: Where do we go from here?

Final panel with all speakers and audience
participation

Moderator: Maria Bustelo, EES President

EES BOARD MEMBER ELECTIONS & BOARD MEMBER CO-OPTION 2013

Following the elections for the new Board member, which were successfully completed in January 2013, we would like to congratulate and welcome Robert Picciotto as our new elected EES Board member. We would also like to thank very much Julia Brümmer, Tamara Mulherin and Peter Wichmand for accepting their nominations and for their

willingness to contribute to the EES. Last but not least thank you to all nearly 140 EES members who casted their ballots.

Following the decision made on 21 February 2013 the EES Board is also happy to welcome back Murray Saunders, the former EES President 2008–2009, as co-opted EES Board

member for 2013 (1 year term, renewable). Murray will take on the responsibility for fundraising of the EES Events and the EES Biennial Conference in 2014 and will represent the EES and NESE voice on the board of the IOCE (collaboration through the Executive Committee of EvalPartners initiative).

THE AUTHORS

Barbara Befani,

Barbara Befani is a London-based Evaluation Methodologist working as an independent consultant. She has a European PhD in Socio-Economic and Statistical Studies and her interests include small-n methods, the evaluation-specific logic, analytic sociology applied to evaluation; and most recently computational methods for complex systems modelling and the study of culture and micro-aggressions. Barbara collaborates on a regular basis with *Evaluation: the International Journal of Theory, Research and Practice*; she has joined international research networks and taught evaluation approaches and methods to PhD students, NGOs, UN and EU officers. She is a member of the American Evaluation Association, the Swedish Evaluation Society and is a Board Member of the European Evaluation Society.

**Maria Bustelo, PhD**

in Political Science, is President of the European Evaluation Society. An Associate Professor in the Department of Political Science and Public Administration at the Complutense University, Madrid (UCM), Spain she directs its Master on Evaluation of Programmes and Public Policies and leads its Quality in Gender and Equality Policies team. She also serves as a member of Spain's National Agency for the Evaluation of Public Policies and the Quality of Public Services Board Committee. She has worked as an independent consultant for several Spanish public agencies, the European Commission and NGOs. She has published widely and has vast evaluative experience in the areas of community development, health promotion, drug dependency prevention, and gender policies.

**Peter Dahler-Larsen, PhD,**

is professor of evaluation at the Department of Political Science and Public Management, University of Southern Denmark, where he is coordinating the Master Program in Evaluation. His main research interests include cultural, sociological and institutional perspectives on evaluation. He was the president of European Evaluation Society 2006–07. His most recent publication is “The Evaluation Society” (Stanford University Press, 2012).

**Kim Forss,**

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

**Dr. Verena Friedrich**

heads the postgraduate study program in evaluation at the Centre for Continuing Education, University of Bern. She has conducted evaluation and research projects related to the use of new technologies within higher education, to worksite health promotion and to tobacco prevention.

**Robert Kirkpatrick**

is Director of UN Global Pulse, an innovation initiative of the Secretary-General harnessing Big Data and real-time analytics for global development and crisis resilience. He was the founding CTO of the Silicon Valley global health and disaster technology NGO InSTEDD, and co-founder of Microsoft Humanitarian Systems. He has spent more than 15 years developing solutions with a focus on organizational change. He has done fieldwork in Iraq, Afghanistan, Kashmir, Uganda, Indonesia, Cambodia, and post-Katrina New Orleans.

**Georgie Parry-Crooke**

is Reader in Social Research and Evaluation at London Metropolitan University. In 1999 she set up and was course director of a Masters programme in evaluation and social research. She continues to carry out evaluations in the health and social care field as well as provide teaching and training in evaluation to a wide variety of audiences.

**Burt Perrin**

is an independent evaluation consultant with over 35 years’ experience, with numerous publications and presentations about how evaluation can be practical and useful. The major focus of his current work is on the organisation of the evaluation function; planning, designing and aiding in interpretation of evaluations; and quality assurance. Burt is a previous Secretary-General of the European Evaluation Society. He was awarded a lifetime membership for his outstanding contribution to the Society.

**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 was a member of the United Kingdom Independent Advisory Committee on Development Impact (2006–2010) and currently serves as a senior adviser to the International Fund for Agricultural Development, New Zealand Aid and the Rockefeller Foundation. He also sits on the Advisory Committee of Wilton Park, an executive agency of the UK Foreign and Commonwealth Office.



Derek Poate

served as President of the UK Evaluation Society in 2012. He co-founded ITAD, a prominent UK evaluation consultancy company, in 1984 and retired as a Director in 2011. His vast experience in international development evaluation spans all the continents. He has been especially active in the natural resources and rural sectors. He has also led comprehensive evaluations of large institutions – IFAD in 2005 and UNAIDS, in 2002 and 2009. His most recent assignments include work for IUCN, the Stockholm Environment Institute and the Rockefeller Foundation. Derek was appointed an Honorary Research fellow at Wye College and later at the Huxley School of Environment, Earth Sciences and Engineering, Imperial College of Science, Technology and Medicine.

**Sandiran Premakanthan**

is a senior public servant in the government of Canada. Prior to joining the public service for several years he managed his consulting practice, President and Principal Consultant, Symbiotic International Consulting Services (SICS), Ottawa, Canada. He has a Master's in Business Administration (MBA), University of Ottawa and is a Credentialed Evaluator (CE). He is the Chair/President of a web based initiative (IOCOM), the International Organization for Collaborative Outcome Management (www.iocomsa.org).

**Dr. Jörg Rech**

is a research assistant at the Centre for Evaluation as well as coordinator and lecturer of the 'Master of Evaluation' programme of Saarland University. He has vast experience in conducting evaluations and social research projects in the fields of education, labour markets and development cooperation.

**Marco Segone**

is responsible for the Decentralized evaluation function as well as the National evaluation capacity development portfolios in the UNICEF Evaluation Office. He is Co-Chair of the EvalPartners Initiative and the UNEG Task Force on national evaluation capacities. Over more than two decades of international development experience he worked in Albania, Bangladesh, Brazil, Colombia, Niger, Pakistan, Switzerland, Thailand, Uganda and the USA. In 2003 he was elected Vice-President of IOCE. In recognition of his contributions to the evaluation profession he was awarded the 2012 Alva and Gunnar Myrdal Practice Award by the American Evaluation Association.

**Nicoletta Stame,**

Nicoletta Stame retired as Professor of Social Policy at the University "La Sapienza", Rome. She is a Past President of the European Evaluation Society. Nicoletta is interested in the theory and methods of evaluation. Her work aims at enhancing the evaluation capacities of public administrators, program implementers and beneficiaries. She is author of *L'esperienza della valutazione* (Rome, 1998), editor of *Classici della valutazione* (Milan, 2007), co-editor (with Ray Rist) of *From Studies to Streams* (New Brunswick, 2006), and author of many essays in books and journals. She is associate editor of *Evaluation, the International Journal of Theory, Research and Practice*.

**Elliot Stern,**

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

**Jacques Toulemonde**

founded the Master's level Programme 'Evaluation and monitoring of policies and programmes' at Lyon University and he managed it until 2012. He has worked as evaluation expert at EU and national levels in many policy areas. He is a member of the International Evaluation Research Group (INTEVAL).

**Mehmet Uzunkaya**

Mehmet Uzunkaya from the Middle East Technical University, Ankara, Turkey has more than ten years of experience in various aspects of PPP policy making and project implementation, in the national and international context, and coordinated the preparation of a draft framework law for PPPs in Turkey.

**Rob D. van den Berg**

is the director of GEF's Independent Evaluation Office. He previously directed the independent Policy and Operations Review Department of the Dutch Ministry of Foreign Affairs and held various positions in Dutch development cooperation, including head of Dutch Development Cooperation at the Netherlands Embassy in Suriname and advisor to the European Commission in Brussels on development research. He has also chaired the Evaluation Network of the Development Assistance Committee (DAC) of the Organisation for Economic Cooperation and Development (OECD).

**Peter Wallace**

is an evaluation adviser based in DFID's Evaluation Department which supports DFID's spending teams to deliver a major expansion of high quality evaluations. His work involves strengthening the approach and design of evaluations commissioned by DFID, with a specific focus on health and private sector led interventions. Peter has previously worked with the Government of Vanuatu and the Scottish Government.

