THE EES NEWSLETTER

EVALUATION CONNECTIONS

EES EUROPEAN EVALUATION SOCIETY

JANUARY 2012

ARTICLES – NEWS – EVENTS

PRESIDENT'S MESSAGE

Dear Members,

First of all, let me wish you a happy, fruitful, and evaluative 2012!

Europe is facing difficult times. Our continent is going through a deep economic crisis. The state of national finances has in several countries lead to severe budgetary restrictions. But the crisis is no longer financial only; and has extended well beyond Europe's borders. Wealth distribution is increasingly lopsided and we are facing a crisis of values. Social cohesion, equitable development, environmental sustainability, equality, fairness and justice no longer hold sway in the public sphere. At the same time, we find hope in a new, awakening civil society mobilising in streets and squares, and in information technologies that allow new ways of communicating and networking across the planet.

What does all this have to do with evaluation and with our Society? Everything! Evaluation is meant to improve what we evaluate. It is also aimed at finding new ways to solve society's increasingly complex problems, and at making politicians, public administrators, entrepreneurs, managers, and civil society leaders accountable for their performance.

European evaluation must be grounded in democratic values and the EES and its members have a major task in spreading evaluation culture throughout institutions and society. Evaluations should not only be rigorous, sound, and useful. They should also avoid capture by vested interests and they should allow the unheard to be heard. If we do our job well, evaluation can ultimately be seen as an essential tool for informing important decisions in these times of austerity and difficult policy choices, so we must be conscious of the importance of our job and the responsibility we have in our hands.

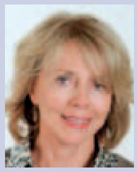
Please join me in thanking Ian Davies for his exceptional leadership during 2010 and 2011. Let us also congratulate Claudine Voyadzis who has been selected by acclamation as Vice President. Finally, let us welcome, our three new board members: Kim Forss, Liisa Horelli, and Barbara Befani. We are also grateful to Bob Picciotto and Sandra Speer, who have agreed to be co-opted again in 2012. All together with Bastiaan de Laat, Peter Wichmand, Karin Aström, and Ian Davies who will remain on the board for another year as Past President, we will all give our best to serve the EES and the evaluation community.

I feel honoured to serve the evaluation community as President of the European Evaluation Society (EES) in these challenging times. Together with the board and an increasingly involved membership, our Society will continue to promote the theory, practice, and utilisation of high quality evaluation in the year ahead.

María Bustelo, EES President

CONTENT

EDITORIAL

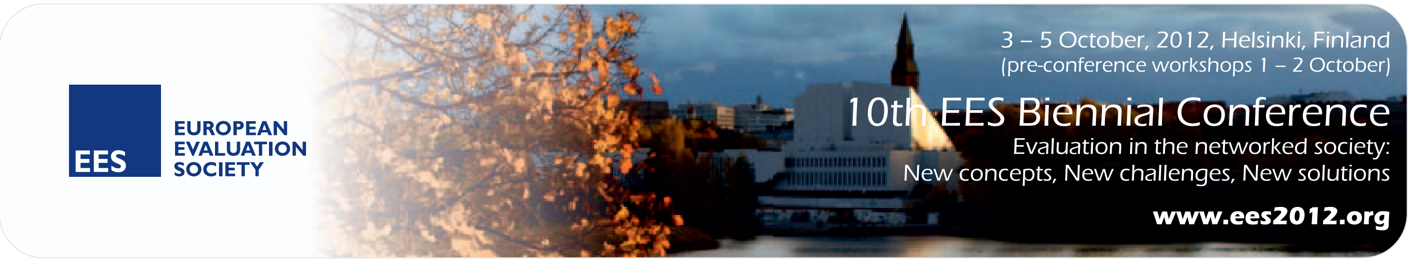


Claudine Voyadzis,
Vice President

Firstly, let me join our President in wishing you all the best for 2012!

We welcome you to the sixth edition of "Connections", the Newsletter of the European Evaluation Society. In this issue, we first take the opportunity to present our three new Board members.

We then provide an overview of EES activities in 2011. With the objective of stimulating innovating thinking about evaluation,



EES has launched a seminar *Vision and Logic of Evaluation* held in Brussels (December 2011). The objective was well achieved with brilliant and relevant presentations of which this issue presents two articles: *"Evaluation and learning in the current crisis"* and *"Four Waves of Evaluation Diffusion"*.

EES has also contributed to the EU Open Days conference (October 2011) through the co-sponsored workshop on ways of capturing the effects of EU funding. Two articles drawn from presentations at this workshop are included in this issue: *"The potential and limits of experimental methods"* and *"Just what is Evaluation 2.0?"* These articles question the relevance and appropriateness of the use of "classic" evaluation methods and explore new ways in a sometime provocative manner.

We greatly appreciate your views and opinions and keenly encourage you to comment on these articles in order to start off a lively debate on these very current and sometimes controversial issues; we wish thereby to set up a forum for discussion and include all relevant comments in subsequent issues of Connections. The next issue is planned for April 2012.

Please address your comments to Czech-In:
secretariat@europeanevaluation.org

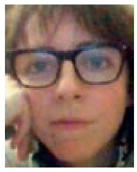
This year, the EES Board has the honour to welcome three new elected Board Members: Kim Forss from Sweden, Liisa Horelli from Finland and Barbara Befani from Italy/UK. The voters, almost one hundred EES members, were faced with a difficult choice among seven highly qualified candidates, all of them with strong experience in evaluation.

Kim Forss
is the Managing Director of Andante – tools for thinking AB, which focuses on evaluation research. Kim wishes to participate in the selection and definition of overarching themes of EES biennial conferences; he is interested in developing evaluation networks, and in promoting close and frequent interactions among national and regional evaluation societies and European institutions.

His experience with current research on evaluation as well as his practical know-how in conducting evaluation along with his chairmanship of the Swedish Evaluation Society will undoubtedly provide the EES Board with additional knowledge and enhanced capacity to further advance the cause of the Society.

Liisa Horelli
has a background in environmental psychology and works as adjunct Professor at Alto University in Finland. She is interested in actively participating in the next EES conference in Helsinki (October 2012) in order to help create close linkages and encourage exchanges of views among all people and parties involved, i.e., EES members, Finnish and national evaluation societies and international organisations.

Liisa has worked mainly with EU-funded programmes and her focus is on participatory planning and community development with special attention to gender issues. Currently President of the Finnish Evaluation Society (FES), she is engaged in the promotion of the development of evaluation theory and practice and the strengthening of an international evaluation culture. Sitting on the EES Board, Liisa's contribution will be most valuable in the preparation and organisation of the EES conference in Helsinki, and in further developing evaluation networks.



Barbara Befani
is an evaluation methodologist and works as an independent consultant. She collaborates with the journal "Evaluation: the International Journal of Theory, Research and Practice" and is interested in the evaluation-specific logic, small-n methods and agent-based modelling.

Barbara has been teaching and writing on evaluation approaches and methods, and wishes to make evaluation values and principles understandable to the larger public and to scientists of different disciplines. With these interests and with her experience of international organizations, the EU and of Italian and British national agencies, Barbara will adequately complement the EES Board's activities.

A last word on these elections: the EES Board seizes this opportunity to thank the candidates, **Luc Lefebvre, Ana Maria García Femenía, Lut Mergaert** and **Julia Brummer** for accepting their nominations, and cordially welcomes their willingness to contribute to EES activities; their help is very much needed and will be greatly appreciated.

The EES organised a new type of event at the end of 2011. This was the Seminar "The Vision and Logic of Evaluation" in Brussels on 2 December in which three reputed evaluators gave introductions on three current "hot" topics, which were followed by round table discussions with the audience. Two of the three presentations, by Nicoletta Stame and Evert Vedung, were reworked into brief contributions to the present newsletter. In the present issue, Prof Stame discusses the important role of evaluation in the current financial, economic and political crisis. Prof Vedung's contribution discusses four subsequent waves in evaluation diffusion, which all have left a sediment on which new contributions in evaluation were built. We're currently in what he calls the "evidence-based" wave – which wave will be next…?
More than hundred EES members attended the seminar, discussions were lively and of excellent standard. The event was a great success, and similar events are planned in the future, especially in non-conference years.

EVALUATION AND LEARNING
IN THE CURRENT CRISIS

Nicoletta Stame

The links between evaluation and politics are well known. These are times of general crisis: financial, political and social. The role of evaluation seems to be that of helping governments to make decisions about what to cut (rigour). But it could also help at surfacing society's latent energies (development, trust) and appreciating diversity (equity). This would imply a different way of understanding the learning function of evaluation: learning to learn.

Evaluation is familiar with single loop learning: correcting errors within a constant framework of norms. But the gravity of the crisis asks for double loop learning: correcting errors by modifying the organizational norms themselves, when new modes are tried and tested. This need is well captured by Patton's idea of a "developmental evaluation", that "specifically supports development" of the situation, by "helping those engaged in innovation examine the effects of their action". Conceived for interventions at the micro level, this approach could be relevant also at the meso or macro level: institutions could evolve by listening to the voice of those who have been able to face the current crisis by finding new solutions to old problems, or by reframing the problems themselves: centralistic attitudes inducing conformity could give up in favour of a more democratic way of finding alternative solutions.

In the following I will provide a few hints at how evaluation could help this twist in learning.

Ordinary public policies have developed in a "distributive" trend: to resist pressures from social movements, small concessions and subsidies have been dispersed in a fragmented way to one category after another, according to their bargaining power, and in order to get their electoral support. Currently the distributive mode works in a reverse way: linear cuts will be imposed on everybody, regardless of needs, effectiveness and efficiency, processes of innovation underway. While cuts could be decided even without evaluation, the opportunity is offered to detect those cases that are meeting the crisis by reorganization, new ways of addressing public needs and delivering services.

Ordinary policies are usually evaluated by *ranking, league tables, monitoring indicators*. They offer a static picture of a situation, where something is better than something else, according to a unifying criterion (or an index) established at the centre. People are supposed to want to be rated higher: once known their location on the table they will act accordingly to improve it. This implies that they will agree on the criteria, and that they know what to do. None of this is true. First, people have different motivations and values. Second, ranking systems do not inform about how a given position has been achieved (both upward in the case of improvement, and downward in the case of decline). Ranking systems could help development if they were supported by horizontal benchmarking, where different ways of tackling problems could be compared in order to see which ones had allowed an upgrading, how (based on what assumptions) and why. This would help the centre learn how to recognize local abilities, value alternative ways of doing things and empower people.

Programs, are the target of the Evidence Based Movement. Starting from the feeling that "nothing works", it advocates the need to provide "robust" evidence of "what works" so that policy-makers can take sound decisions about what to keep or to dismiss, across the board.

Even if decision-makers learned that way, and used evaluation instrumentally, the policy consequences would be that of reinforcing a centralistic practice that is responsible for thwarting society's energies, instead of helping society reflect about its potentialities.

Nicoletta Stame

Nicoletta Stame retired from teaching Social Policy at the University "La Sapienza", Rome. She was a co-founder and the president of the Italian Evaluation Association. She is a Past President of the European Evaluation Society. She is associate editor of *Evaluation*, and participates in the International Evaluation (Inteval) network.

Nicoletta is interested in the theory and methods of evaluation. She is author of *L'esperienza della valutazione*, editor of *Classici della valutazione*, co-editor (with Ray Rist) of *From Studies to Streams*, and author of many essays in books and journals.

She has evaluated programs of enterprise creation, social integration and R&D at the local, national and European level. Her work aims at enhancing the evaluation capacities of public administrators, program implementers and beneficiaries.

Attribution. It is assumed that policy-makers want to know whether their intervention can be attributed a positive effect (value for money): if not, the program could be suppressed. But most programs are conducted in cooperation among actors, and along levels, and evaluation could be better aimed at understanding how such synergies have been enforced. Moreover, effects can be obtained through many different causal packages, and evaluation could aim at comparing packages in order to allow making judgements about their respective value.

Robust methods. Contrary to the hierarchy of robustness (from experimental methods down to ethnographic studies) which is implied, current debates hinge around strengths and weaknesses of each method. What matters is selecting approaches according to their "responsiveness" to the situation, the characteristics of programs and the evaluation demands, from "how to improve a program" to "why some effects have been obtained". For instance, participatory approaches and approaches based on "positive thinking" could provide crucial understanding of innovative trends.

Guidelines for implementation. For fear that an implementation failure may undermine results, it is recommended to provide precise protocols for implementation. This may impose conformity when discretionality and customization might be needed by the policy content. Moreover, implementation may follow different trajectories, and it will be necessary to learn how to motivate people that oppose initial resistance, or to maintain momentum in programs that tend to decline.

Terms of reference. It is often asked to employ specific "robust" methodologies, irrespective of the kind of program to be evaluated. It would be wiser to allow for a choice of methods open to an agreement between commissioner and evaluator, and to allow a change of design if the circumstances of emerging factors, and the reinterpretation of the program theory so requires.

Learning to learn

The important role evaluation may play in the current crisis is one of learning to learn how alternative routes for the solution of problems that have got us stuck could be available. This could provide trust in the "social nexus" and mobilize creative forces. To do that, it should pursue the presentation of innovative solutions proposed by different actors in their quest for development, scout even shy attempts at improvement, give alternative routes (and "causal packages") an equal chance, and offer tools for self-reflection. Moreover, and as a consequence, it should second a brokering role of the centre in recognizing innovations and strengthening them.

FOUR WAVES OF EVALUATION DIFFUSION

Evert Vedung

Evert Vedung

Evert Vedung is emeritus professor of political science, especially housing policy at Uppsala University's Institute for Housing and Urban Research and Department of Government. His works on evaluation in English include *Public Policy and Program Evaluation* (author, 1997, 2010), *Carrots, Sticks and Sermons* (1998, 2003, coeditor), and "Four Waves of Evaluation Diffusion" (2010). Please address correspondence to evert.vedung@ibf.uu.se or Uppsala University, IBF, PO Box 785, SE-80129 Gävle, Sweden.

Evaluation is an incredibly popular governance recipe. It seems to be virtually everywhere in most branches of public endeavors. Yet, the development since 1960 has not been straight toward more and more evaluation. Sometimes, one evaluation form has been heavily criticized and lost much of its support in favor of new forms which have been equally highly praised. Yet also these new forms have been attacked and lost their support in favor of other fads. Intriguingly, against all odds some earlier forms have gained new momentum and come back slightly changed and in different linguistic guises. (Furubo, Rist & Sandahl 2002).

Four Evaluation Waves with Depositions of Sediment

The whole thing can be likened to *ocean waves* that are roaring in and subsiding, roaring in and subsiding. We will therefore speak of evaluation waves. Four are the evaluation waves which have swept across the Atlantic world and particularly some countries of the Atlantic world, including the Scandinavian countries and the rest of Europe. In subsiding each wave has left behind layers of *sediment*. In due time, the evaluation landscape has come to consist of *layers upon layers of sediment*.

I have discerned four evaluation waves between 1965 and 2010:

1) the Scientific Wave
2) the Dialogical Wave
3) the Neo-Liberal Wave, and
4) the Evidence-Based Wave.

Scientific and Dialogical Waves

The *Scientific Wave* entailed that academics should test empirically, through two-group classic experimentation, appropriate means to reach externally set, admittedly subjective, goals. Evaluations should be carried out as scientific research, conducted by the Academic Man. They were carried out one off. Public decision-makers were then supposed to roll out the most effective means. The implied the triumph of means-ends rationality, i.e. the value doctrine of Max Weber.

Faith in scientific evaluation eroded in the early 1970s. Instead of being scientific, evaluation should be democratic in the sense of stakeholder-oriented, with information being elicited from users, operators, managers and other concerned parties. It should be carried out through discussions or deliberation *avant la lettre*. In this way, the *Dialogical Wave* entered the scene. Evaluation should be made by the Common Man, not the Academic Man. The dialogue-oriented wave communicative rationality, i.e. rationality by enlightened deliberation. It was Jürgen Habermas, not Max Weber.

Neo-Liberal Wave or New Public Management

Enter the *Neo-Liberal Wave*, also referred to as *New Public Management*, in the last years of the 1970s. Deregulation, privatization, contracting-out, efficiency and customer influence became key phrases. Evaluation as accountability, value for money and customer satisfaction was recommended. Evaluation became part of grand governance doctrines, such as purchaser-provider models, client-oriented and results-oriented management (management by objectives). The latter is the paradigm exemplar of what evaluation meant in New Public Management.

Evidence-Based Wave

Under the slogan 'What matters is what works' the *Evidence-Based Wave* implies a renaissance for scientific experimentation. It seems to mark the return of the Scientific Wave. Again, two-group double-blind randomized experimentation is the preferred approach. Yet, now comprehensive methods hierarchies are constructed in which experimentation is deemed best and professional and client knowledge worst. A second difference is the belief in systematic reviews of extant academic experimental studies as a means of furnishing decision-making with evidence based interventions that works, i.e. that causally produce the desirable outcomes (EBTs = Evidence-Based Treatments).

With time this strong interpretation of evidence as double-blind randomized controls has met massive criticism from academics and professionals in, i.e., social work. They have come up with alternative interpretations of evidence-based, summarized under the notion Evidence-Based Practice (EBP). In addition to scientific evidence, even professional knowledge and client knowledge, elicited through evaluation, is regarded as evidence that should be taken into account in public sector decision-making and deliberation.

Waves, Neither Generations, nor Trees

I have adopted the wave metaphor in contrast with the *tree metaphor* used by Marvin Alkin (2004) in his evaluation theory tree, and the *generation metaphor* used by Guba & Lincoln in their 1989 book *Fourth generation evaluation*.

Evaluation: a Minimal Definition

In my exposition, I have found it appropriate to define what I mean by evaluation. Thus, I have not followed the examples set by Michael Power in his book *The Audit Society* who did not define what he meant by auditing (1997:xvii–xviii), or Furubo, Rist and Sandahl who did not define the term evaluation in their *Atlas of Evaluation*. My minimal definition reads as follows:

Evaluation =$_{def.}$ careful retrospective assessment of public sector interventions, their organization, content, implementation, outputs, or outcomes, which is intended to play a role in future decision situations.

In one sense, this definition is narrow. It is limited to interventions, and to public sector interventions at that. It is equivalent to ex nunc and ex post evaluation, i.e. assessment of already adopted, decided, ongoing, or (just) finished interventions. It excludes evaluation ex ante (prospective e.), that is, consequence analyses of proposed and considered interventions, which are made *before* they are adopted.

In another sense it is wide. It is limited not only to the search for effects of decisions and activities at the outcome level (i.e. in society or nature) but includes the ascertaining of outputs and implementation processes as well. And it includes organization of interventions (Vedung 1997: 2 ff and 2006: 397).

The idea of *mimimal definitions* (opposite: *maximal definitions*) has been brought forward by the eminent Italian political scientist Giovanni Sartori. Yet references are difficult to find. In Sartori 1984 there are some clarifications, unfortunately too brief: 55–56, 64, 79 ("minimal definition") och 81 ("[p]arsimony (in defining)"). See also Sartori 1976: 58 ff. For an attempt by myself, see Vedung 1982b: 90.

Reference:

Vedung, Evert, 2010, "Four Waves of Evaluation", *Evaluation, 16(3), 263–277.*

http://evi.sagepub.com/content/16/3/263.full.pdf+html

Other references:

Alkin, Marvin, 2004, *Evaluation Roots: Tracing Theorists' Views and Influences* (pp. 12 ff), London: Sage.

Furubo, Jan-Eric, Ray C. Rist and Rolf Sandahl, 2002, eds., *International Atlas of Evaluation*, New Brunswick, New Jersey and London: Transaction Publishers.

Guba, Egon G and Yvonna S Lincoln, 1989, *Fourth Generation Evaluation* (pp. 50 ff), London: Sage.

Power, Michael, 1997, *The Audit Society: Rituals of Verification*, Oxford: Oxford University Press.

Sartori, Giovanni, 1976, *Parties and Party Systems: A Framework for Analysis*, vol I, Cambridge: Cambridge University Press.

Sartori, Giovanni, 1984, "Guidelines for Concept Analysis", i Giovanni Sartori, red., *Social Science Concepts: A Systematic Analysis*, 15–85, Beverly Hills, CA.: Sage.

Vedung, Evert, 1982b, *Political Reasoning*, Newbury Park, CA.: Sage.

–, 1997, *Public Policy and Program Evaluation*, New Brunswick, New Jersey and London: Transaction Publishers. First paperback printing 20000228.

–, 2006, "Evaluation Research", i Peters, B Guy and Jon Pierre, *Handbook of Public Policy*, 397–416, London: Sage.

EVALUATING EU COHESION PROGRAMMES:
CATCHING SMOKE IN A NET?

Robert Picciotto

Evaluation was a major focus of the 9th OPEN DAYS European Union Conference. The contribution of EU funding to smart, sustainable and inclusive growth was its overarching theme. The Conference attracted 5,700 participants. It was held over four days (October 10–13, 2011) in Brussels, a period characterized by lively public debates about the systemic implications of the financial crisis, the European Commission's budget and the post-2013 proposals for a new cohesion policy.

Assessing the impact of EU funding is an extraordinarily tough challenge. EU programmes are typically multi-facetted, poorly defined and hard to evaluate. The Union's funding mechanisms (the European Regional Development Fund; the European Social Fund; the Cohesion Fund and the European Fisheries Fund) embody diverse policy directions.

Against this background this issue of Connections includes summaries of papers presented at an Open Days workshop co-sponsored by the European Evaluation Society. The session dealt with alternative ways of *Capturing the Effects of EU Funding*. This topic was highly relevant to all three major sub-themes of the Conference:

(i) *Europe 2020:* research and innovation; "digital Europe"; creative industries; low-carbon development strategies; employment; poverty reduction and social inclusion;

(ii) *Better delivery:* common strategic frameworks; exchange of good practice; the regulatory framework post-2013; more focus on results, new evaluation methods; etc.

(iii) *Geography matters:* usefulness of territorial approaches; transportation policies; diverse geographical/demographic challenges; regional strategies; urban research; etc.

Even though considerable resources have been invested in evaluation little is known about how policy levers, instruments and mechanisms should be combined and managed to achieve intended results. While attribution of results to individual programs is critical for accountability purposes, Europe also needs evaluations geared to knowledge creation about social cohesion and regional development.

"HOW TO CAPTURE THE EFFECTS FROM EU FUNDING"

Jenny Hughes

Just what is Evaluation 2.0?

Evaluation 2.0 is a set of ideas about evaluation that Pontydysgu[1] are developing. At its simplest, it's about using social software at all stages of the evaluation process in order to make evaluation more open, more transparent and more accessible to a wider range of stakeholders. At a theoretical level, we are trying to push forward and build on Guba and Lincoln's ideas around 4th generation evaluation which is a constructivist approach incorporating key ideas around negotiation, multiple realities and stakeholder engagement. But this is the first part of the journey – ultimately, I believe that e-technologies are going to revolutionise the way we think about and practice evaluation.

In what way do you think this is going to happen?

Firstly, the use of social media gives stakeholders a real voice – irrespective of where they are located. Stakeholders can create and publish evaluation content. For example, in the past I might carry out some interviews as part of an evaluation. Sometimes I recorded it, sometimes I just made notes. Then I would try and interpret it and draw some conclusions about what it meant. Now I set up a web page for each evaluation and I podcast the interviews using audio or video and put them on the site. (Obviously this has to be negotiated with the interviewee but so far, no one has raised any objections.) There is the usual comment box so any stakeholder with access to the site can respond to the interview, add their interpretations, agree or disagree with my conclusions and so on.

Secondly, I think it is challenging our perceptions of who are evaluators. Everyone is now an evaluator. Think of the software that you use every day for on-line shopping from Amazon or Ebay or any big chain store. If I want to buy a particular product I check out what other people have said about it,

[1] Pontydysgu is a small independent research organisation specialising in e-learning and education evaluation. It is based in Germany and in Wales and the name Pontydysgu approximately translates as Bridge to Learning.

"Dinosaurs were highly successful and lasted a long time. They never went away. They became smaller, faster, and more agile, and now we call them birds."

how many of them said it and how many stars it has been given. These are called recommender systems and I think they will have a big impact in evaluation. We have moved from the paradigm of the 'expert' collecting and analyzing data into a world of crowd sourcing – harnessing the potential of mass data collection and interpretation.

Thirdly, the explosion of Web 2.0 applications has provided us with a whole new range of evaluation tools that open up new methodological possibilities or make the old ones more efficient. For example, if I am at the stage of formulating and refining the evaluation questions – I put it out as a call on Twitter. It's amazing how restricting evaluation questions to 140 characters can sharpen them up!

I did an evaluation of a community capacity building project in an inner city area recently and spent quite a long time before I went to the first meeting walking around the streets, checking out the community facilities, the state of the housing, local amenities and so on, to get a 'feel' for the area – except I did it on Google Earth and with street-view on Google maps. There are about 20 or so other applications I use a lot in evaluation but maybe they will have to wait for another edition!

Fourthly, I think the potential of Web 2.0 changes the way we can visualize and present data. Why are we still writing long and indigestible text-based evaluation reports? Increasingly clients are preferring short, sharp evaluation 'articles' on maybe one outcome of an evaluation which they can find on a 'newsy' evaluation webpage – with hyperlinks to more detailed information or raw data or back up evidence if they want to check it out. We can also create 'chunks' of evaluation reporting and repurpose them in different ways for different stakeholders or they can be localized for different cultures – for example, I have started doing executive summaries as downloadable podcasts. I think evaluation 2.0 is about creating a much wider range of evaluation products.

Following on from that, I think Evaluation 2.0 breaks down the formative-summative divide and notions of 'the mid-term report' or 'the ex-ante report'. Evaluation 2.0 is continuous, it is dynamic and it is interactive. For example, I use Googledocs with all my clients – I add them as readers and editors on all the folders that relate to their evaluations. At any time of the day or night they can see work in progress and add their comments. I keep their evaluation website up to date so they get evaluation information as soon as it is available.

So do you think all evaluators will have to move down this road or will there always be a place for evaluators using more established methods?

Personally, I think massive change is inevitable. Apart from anything else, our clients of the future will be the digital natives – they will expect it.

There will always be a role for the evaluator but that role will be transformed and the skills will be different. I think a key job for the specialist evaluator will be designing the algorithms that underpin the evaluation. The evaluator will also need to be the creative director – they will need skills in informatics, in visualizing and presenting information, the creative skills to write blogs and wikis. They will need networking skills to set up and facilitate online communities of practice around different stakeholder groups and the ability to repurpose evaluation objects.

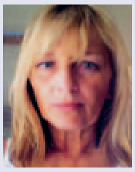
The rules of engagement are also changing – in the past you engaged with a client, now you engage with a community. We also have to think how stakeholder created content might change our ideas about copyright, confidentiality, ownership, authorship.

So do you think evaluators as we know them will become extinct!!

Well, as Mark Halper said

"Dinosaurs were highly successful and lasted a long time. They never went away. They became smaller, faster, and more agile, and now we call them birds."

Jenny Hughes

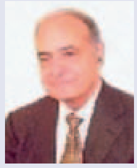


Formerly a chief education officer in local government, I have been working as an evaluator, trainer and researcher since 1996, mainly for government departments, universities and NGO. Much of my work has been for international aid organisations and, before I decided I was far too old to be spending my time in war zones, I worked in Central Europe, the Middle East and the FSU. My PhD was in something completely irrelevant and I haven't published anything remotely resembling an academic paper since I discovered blogging about five years ago (http://www.pontydysgu.org). I do, however, still write practical handbooks (such as *'The Project Manager's Guide to Evaluation'* or *'The Classroom Teachers Guide to E-learning*). I have a passion for radio and together with colleagues from Pontydysgu, make regular internet radio programmes – which I think is the medium of the future for evaluation and a lot of other things!

CAPTURING THE EFFECTS OF EU FUNDING: THE POTENTIAL AND LIMITS OF EXPERIMENTAL METHODS

Robert Picciotto

Robert Picciotto, AcSS,



Visiting Professor at King's College (London), is a graduate of the Ecole Nationale Superieure de l'Aeronautique (France) and of the Woodrow Wilson School of Public and International Affairs (Princeton University). At the World Bank which he joined in 1962 Robert Picciotto held several operational and corporate management positions including Director of Projects Departments in three Regions, Vice-President, Corporate Planning and Budgeting and Director-General of the Independent Evaluation Group (1992–2002). A board member of the European Evaluation Society and of the UK Evaluation Society he currently serves on advisory committees assembled by the World Bank; the Rockefeller Foundation; and Wilton Park, an executive agency of the UK's Foreign and Commonwealth Office.

Capturing effects of EU programmes has become an overwhelming social and political imperative. Policy makers are clamouring for a better understanding of how economic growth combined with social inclusion can be brought about through EU funding. Despite major investments in evaluation, surprisingly little is known about how regional development programs should be designed, used and sequenced in the diverse and complex operating environment of the European Union.

Another driving force behind the commissioning of rigorous impact evaluations is the extraordinarily tight fiscal environment associated with the on-going economic crisis: taxpayers want to hold decision makers to account. EU funded cohesion programs are now expected to (i) include relevant dimensions of intended program benefits in the form of indicators; (ii) measure changes in these indicators with acceptable accuracy throughout the monitoring and evaluation cycle; and (iii) ascertain the extent to which the observed changes can legitimately be ascribed to the intervention.

Impact evaluations address the third dimension of this agenda. They tackle the attribution question, i.e. they seek to assign causality to the effects that are observed following an intervention. A *before- after comparison* design has been the traditional way of measuring effects for areas and groups. This is rarely a satisfactory way of addressing causality since influences other than the intervention may have affected the results. Contributory factors exogenous to the intervention could have improved or hindered the achievement of intended goals.

Hence, rigorous impact evaluation uses *with-without* comparisons instead of *before-after* methods. This requires the identification a *counterfactual*, i.e. a plausible and relevant scenario of what would have happened without the intervention. In the right circumstances experimental methods establish causality by providing a valid measure of what results would have been observed had the intervention not taken place. They achieve comparability between control and treatment groups as a result of random selection of beneficiaries and non-beneficiaries drawn from the same population through an explicit chance based process.

This approach addresses the issue of *selection bias* which arises when comparing impacts on two very different sets of beneficiaries that may end up falsely attributing the observed results to the intervention even though different known or unknown characteristics of the treatment and non-treatment groups may have been at work (e.g. in cases where those who accessed the program are richer, more powerful, more motivated or more educated). Thus, experimental methods exude precision and objectivity. Unfortunately, they are not always appropriate and they do not on their own tackle key evaluative questions, e.g. why a program has worked or not and who was responsible for the outcomes.

Experimental methods are redundant when no other plausible explanation for the results observed is available. They are costly, require large samples, highly skilled statisticians and specialized quality assurance arrangements. They are not feasible when no untreated target group can be identified, e.g. when an intervention is intended to be universal. They are poorly suited to the evaluation of complex programs. Their validity is threatened when the program changes during implementation, e.g. in response to evolving conditions. The behaviour of treatment and non-treatment group individuals may be affected by the experiment itself. Finally, experimental evaluations often raise ethical questions: depriving members of the control group of a useful treatment based on an arbitrary selection process can be perceived as discriminatory and may even be illegal.

In sum, where adequate resources and skills are available and ethical dilemmas can be resolved, rigorously designed and independently implemented experimental methods are the best way to assess attribution but only for relatively simple interventions the effects of which are realised in a short period of time and are large relative to other potential influences. Where these conditions do not prevail, one or more alternative ways of ascertaining attribution may be adopted: (i) regression and factor analysis; (ii) quasi-experimental designs; (iii) multivariate statistical modelling; (iv) participatory approaches; (v) surveys and sampling; (vi) expert panels validated by a Delphi process; (vii) benchmarking; etc.

Tools are just tools. Appreciating the potential and limitations of alternative evaluation methods is a critical competency for evaluators. No single methodology should be allowed to dominate what is first and foremost a creative, analytical and participatory process. In most real world situations *mixed methods* combining qualitative and quantitative approaches are the best way to ascertain what works and doesn't work; why interventions succeed or fail; whether design or implementation problems need to be addressed and who among partners is responsible for particular outcomes. In Michael Patton's words, "the only gold standard in evaluation is appropriateness".