



EVALUATIONS CONNECTIONS

CONTENT

A Message from the President	1
Editorial: The pursuit of contextual objectivity	2
Evaluating complex realities	3
It's the very core of evaluation and makes or breaks our work: so why is it in hardly anyone's toolkit?	4
The tricky triangle – Evaluator, Evaluand, Evaluation Commissioner	6
Adding rigour to the evaluation of multi-actor processes: some reflections on the non- use of game theory and related actor analysis methods	7
Lessons from performance measurement in assessing research	8
Evaluation Use in the Organisational Context – Changing Focus to Improve Theory	10
Challenges of evaluating environment in international development	12
Sistematización: what's that? Introducing a multi-stakeholder, reflexive action-learning evaluation approach	13
Have you participated in an evaluation that influenced people's lives and / or led to social betterment?	15
The authors	15

A Message from the President

Dear EES Members,
greetings and welcome to the first 2014 issue of your Newsletter!

First of all I congratulate our newly elected members: Riita Oksanen, Vice-President and Jan-Eric Furubo, Board member. I also thank Murray Saunders for staying on as a co-opted member in 2014 along with Maria Bustelo who as past president remains on board for another year. Together with Barbara Befani, Julia Brümmer, Ramon Crespo, Bastiaan de Laat, Kim Forss, Liisa Horelli and Bob Picciotto our team is all set to work with you to serve the cause of evaluation in Europe and beyond!

In particular we are doing our best to ensure that the 11th Biennial EES Conference to take place in Dublin on 1–3 October 2014 will once again be the evaluation event of the year. Pre-conference workshops are on 29 and 30 September. They are an integral part of the event. The overarching theme we selected ("*Evaluation for an Equitable Society: Independence, Partnership, Participation*") has drawn a lot of interest. We are seeking to trigger debates on what evaluation should do to overcome inequality and contribute to ending absolute poverty; help assert its independence; involve poor and neglected groups in the evaluative process; and create partnerships across the public, private and voluntary sectors and at multiple levels to strengthen its social impact. Once more the EES confe-

rence will provide an opportunity for evaluators hailing from all over Europe and the rest of the world to share evaluation knowledge and experience across borders and to engage in reflection about the challenges that evaluators face in a changing world.

The conference will provide a platform for workshops made possible by EVALPARTNER grants focused on increased evaluation cooperation and on extending the reach of evaluation through creativity and innovation. Thus a *Parliamentarian Forum* will create space for dialogue about democratic evaluation while *Evaluation stories that make a difference* will illustrate the distinctive contribution of our discipline to the solution of real world problems. We will take stock of progress made in *Strengthening collaboration with national evaluation societies* following a successful workshop on this topic held in Warsaw in January. We also expect to further advance evaluation professionalization by following up on the joint UKES-EES workshop about a *Voluntary Evaluator Peer Review* (to be held in London in April 2014).

Needless to say we expect all existing Thematic Working Groups (*Gender, Professionalization, Private Sector, Sustainability, Fragile States*) to take advantage of the conference to take stock of their progress and plan their future activities. In addition, I would like to encourage all EES members to set up new Thematic Working Groups. Please do not

hesitate to contact me directly should you be interested in creating a TWG!

This issue of Connections confirms that evaluation is a vibrant and vital discipline that is critically dependent on knowledge exchange and a fulsome contest of ideas and opinions. To better contribute to this mission we have reconstructed the EES website. **The revamped website will be launched**

on the 6th of March! It will feature a blog section to facilitate exchange of views on controversial evaluation topics. It will make possible user-friendly participation in forums. It will enhance networking within and across Thematic Working Groups. Last but not least it will host a specific forum operated by the Network of Evaluation Societies of Europe (NESE) dedicated to interactive exchanges among evaluation societies.

I invite you all to take an active part in shaping the future of EES. Celebrating our diversity, exploring new evaluation frontiers and bringing us closer together is what I am after – and what Connections is all about. I very much look forward to hearing from you and to welcoming you in Dublin.

Claudine Voyadzis,
EES President

EDITORIAL: THE PURSUIT OF CONTEXTUAL OBJECTIVITY

Robert Picciotto

Validity and rigour are cornerstones of the evaluation endeavour. Accordingly all evaluation guidelines enshrine absence of bias and avoidance of conflict of interest as prime characteristics of ethical evaluation. At the same time fulsome engagement with stakeholders and intimate understanding of the context (including the force field within which evaluations are carried out) are prerequisites of high quality evaluative work. On the one hand evaluators cannot take sides in order to preserve their credibility. On the other hand they cannot afford to ignore the passions and interests of those likely to be affected by the evaluation process.

This is where critical systems thinking comes into its own since, according to Martin Reynolds, it is designed to tackle the complicated reality of interconnectedness, i.e. not only the *inter-relationships* of the entities involved in or affected by the program being evaluated but also their multiple *perspectives* and the conflict ridden complexity that arises from their contrasting *boundary* judgements. In particular, equity-focused evaluations force evaluators to confront the contest between ethical considerations and the realities of power imbalances.

Thus systems thinking induces evaluators to probe who gets what; who owns what; who does what; and who and what gets affected. How to go about this in practice is the focus of Jane Davidson's article. She stresses that answering questions about quality, value and/or importance lies at the core of the evaluation

process. She observes that too many evaluators get lost in indicators, metrics and stories so that they fail to visualize what worthwhile outcomes look like and often lack the inclination or the skills to (i) devise evaluation rubrics that generate unambiguous measures of programme quality, (ii) capture pertinent data derived from such rubrics and (iii) aggregate the resulting findings into higher level conclusions that serve summative or formative evaluation goals responsive to the needs and aspirations of evaluation stakeholders.

Bastiaan de Laat's contribution unpacks the diverse roles played by programme managers, evaluation commissioners and evaluators. They make up a "tricky triangle" that shapes evaluation within or between organizations. Depending on the governance configuration one of these actors may be able to capture another with important consequences for evaluation protocols. This holds major implications for the incentives framework that drives the evaluation process. In turn, Leon Hermans explores the untapped potential of game theory in evaluation and concludes that the limited resort to actor analysis can itself be explained by the complex games that evaluation stakeholders play.

Steven Hojlund goes even further by putting forward the paradoxical hypothesis that evaluations that are useful to society are unlikely to be carried out or used since most organizations view evaluation as a way to justify what they do. The dire consequences of distorted incentives also emerge in the ar-

ticle by Sirkka Immonen et.al. which explores the interface between monitoring and evaluation in international research organizations and confirms that a misplaced focus on input and output indicators that are relatively easy to measure discourages the creative and innovative behaviours without which research value cannot be generated.

Just as in the field of agricultural research, environmental protection programmes bring up tough evaluation dilemmas. Thus Juha I. Uitto stresses the critical role of time horizons, cross border effects, risk, uncertainty and complexity in the choice of pertinent methods in development cooperation. Echoing Jane Davidson's concerns about the need to carry out evaluative syntheses his article highlights the "micro-macro paradox" where individual interventions deemed to be successful by micro level evaluations do not appear to generate commensurate benefits at national or global levels.

Closing the March issue of Connections, the Esteban Tapella and Pablo Rodríguez-Bilella article reverts to the theme of complexity and uncertainty evoked by the multi-stakeholder, reflexive action-learning evaluation approach called *sistematización*. This approach is little known outside Latin America. Yet it shares with systems thinking a reluctance to adopt linear cause-effect logic models. Instead it captures and maps the dynamics and interdependencies that influence outcomes in specific cases and it encourages partners and communities to bring new interpretations to

the learning process by contributing narratives to the multifaceted story of the overall intervention.

To be sure no single evaluation model can lay claim to resolving all the tensions that

arise when detachment and objectivity are expected to be combined with sensitivity to the unique social and cultural factors that shape policies and programs. approaches. Evaluation commissioners and programme managers are not held to the same standards

of non-partisanship that evaluators are committed to uphold. Accordingly the search for contextual objectivity will long continue to challenge evaluators in the real world of evaluative practice just as it has in the domain of quantum physics where the term originates¹.

1 In quantum physics the measurement result of an observable depends on the arrangements made to measure it.

EVALUATING COMPLEX REALITIES¹

Martin Reynolds

Two parallel methodological developments and associated traditions of thinking have emerged within the global evaluation community. They deal specifically with evaluating complex realities through complexity thinking and systems thinking. Complexity thinking draws on complexity science and the revelations regarding the incidence of non-linear interconnectedness amongst entities. The tradition departs from Newtonian science which it caricatures as dealing with mechanistic linear relationships of simple cause and effect. Within the tradition of scientific practice the evaluand is characterised as being ‘complicated’ rather than ‘complex’.

Complicated situations have interconnected entities which can with various degrees of difficulty be subject to the certainties of design or programming. They are essentially predictable and can be controlled. Complex situations, on the other hand, are harder to plan. They are essentially uncertain, unpredictable, and significantly uncontrollable. Perhaps the most familiar illustration of complexity is the ‘butterfly effect’ – the proposition by meteorologist Edward Lorenz in the 1960s that a flapping of a butterfly wings in South America may lead to significant weather events, such as a hurricane, at a far removed distance, say, North America.

The effect works socially as well as biophysically. The on-going ‘Arab Spring’ events giving rise to national and international upheavals can be linked back to the personal action and tragedy of one disaffected individual in Tun-

sia being denied his livelihood opportunity to sell vegetables.

Systemic emergence is a term often used amongst complexity practitioners to describe effects of uncertainty, unpredictability, and uncontrollability. However, one of the concerns sometimes expressed by evaluators grappling with such issues is a sense of hopelessness arising from notions of systemic emergence.

Where events of an evaluand are regarded as ‘other-worldly’ what hope is there of making meaningful evaluations? Concern for systemic effects might be expected to invite greater attention to systems thinking. Curiously though there appears to be relatively little attention given to this longer tradition of thinking in practice for evaluating complex realities.

Where reference to systems is made it tends to be confined to an assumption that systems are merely sets of interrelated entities. The systems-based evaluation presented here takes a more nuanced-view of complexity and systems in emphasising the role of systems in not only representing sets of interrelationships, but as proxy to perspectives about such interrelationships and interdependencies. In other words, systems might be used as ontological devices to represent reality – as with complexity thinking – but also as epistemological devices or conceptual constructs to actively learn about and transform reality. Such a systems-based evaluation

will have explicitly subjective components that interplay with the real world. But how might these systems devices be different from other subjective devices used for evaluation such as the use of story-telling in, say, action research?

Drawing on the sub-tradition of critical systems thinking (CST), ‘reality’ is very much kept in check with a systems-based evaluation. The ‘real world’ of CST comprises not only the complicated reality of interconnectedness of entities – *inter-relationships*, but also the complex reality of multiple *perspectives* on the real world, as well as the conflict ridden reality of social and ecological tensions arising from contrasting *boundary* judgements associated with contrasting perspectives (cf. Williams, 2013). Boundary conflicts in this latter instance have expression in the ethical domain regarding contested ideas of what’s good and what’s right, but also in the political domain regarding, for example, how ‘goodness’ and ‘rightness’ might be circumscribed by social relations of power (gender, religion, ethnicity, sexuality, socio-economic status, etc. etc.).

Equity-focused developmental evaluation based on CST, at its simplest attempts to provide a handle for evaluators to unlock these complicated, complex, and conflict ridden realities. One particular handle from CST is a powerful reference system – critical systems heuristics (CSH) – based on the works of two systems philosophers; C. West Churchman from an American prag-

matist tradition, and Werner Ulrich based in Switzerland from the more European critical social theorist tradition. The evaluand chosen to illustrate the workings of CSH in the article is the Narmada Dams project in India. The article is based on a desktop evaluation undertaken with my colleague Bob Williams, which was commissioned by UNICEF in order to illustrate the helpfulness of systems thinking for equity-focussed evaluations (Bamberger and Segone, 2012).

The Narmada Dams were initiated in 1949 as flagship interventions for the newly independent India. Until now, they remain a source of considerable controversy and debate. Indeed, the projects were chosen for the UNICEF commission because of their widespread and longstanding ethically and politically contentious implications. With a focus on promoting equity three quite unique features of using critical systems thinking may be drawn. First, there is an explicit attempt at exploring the inter-relationships between values, power, knowledge and legitimacy, which are at play in any evaluand. Using the handle of CSH questions regarding who gets what are

linked with associated questions of who owns what, who does what, and who and what gets affected in the process.

Second, as against conventionally *applying* value judgements in an evaluation, there is an imperative towards *developing* value in a CST-based evaluation. In other words, the evaluator is regarded very much as an active player in the evaluand, rather than as a passive observer. S/he may be involved with either sustaining/reinforcing the existing status-quo or be part of meaningful transformation in developing value.

Thirdly, with a pro-equity focus CST invites possible convergence and compatibilities with other existing methods of evaluation. Hence the imperative is not to substitute a new 'method' to replace other methods, but rather to seek ways in which systems thinking may articulate purposefully with the existing skill-sets of practising evaluators. The result is a meaningful, though not necessarily exclusive, correspondence with, theories of change, programme evaluation, developmental evaluation and realist

evaluation. Moreover, contemporary systems thinking can, given its particular focus on inter-relationships, incorporate insights from complexity science. However, a key point of departure from complexity science remains with appreciating 'systems' more as (abstract) conceptual tools for exploring and designing purposeful change than as mere representations (obstructs) of complex realities.

References

Bamberger, M and Segone, M. (eds.) (2012). *Evaluation for equitable development results*. New York, UNICEF.

Reynolds, M. (2014). "Equity-focused developmental evaluation using critical systems thinking." *Evaluation* 20(1): 75–95.

Williams, B. (2013). "Three core concepts: inter-relationships, perspectives, boundaries." *Evaluation Connections: Newsletter of the European Evaluation Society* June 2013.

† A fuller treatment is in Reynolds, M. (2014).

IT'S THE VERY CORE OF EVALUATION AND MAKES OR BREAKS OUR WORK: SO WHY IS IT IN HARDLY ANYONE'S TOOLKIT?

E. Jane Davidson

It's shocking but true: there's a critically important 'core' to evaluation that absolutely makes or breaks our work. It's so essential that if it's missing, we can't defensibly put the word 'evaluation' in the title of a report or presentation. Yet the vast majority of those who identify as evaluators don't even have it in their toolkit. It is *evaluative synthesis*, arguably the most important of the methodologies that are distinctive to our discipline. Simply defined, evaluative synthesis is the systematic combination of evidence with definitions of 'quality' and 'value' to draw well-reasoned and defensible conclusions about performance.

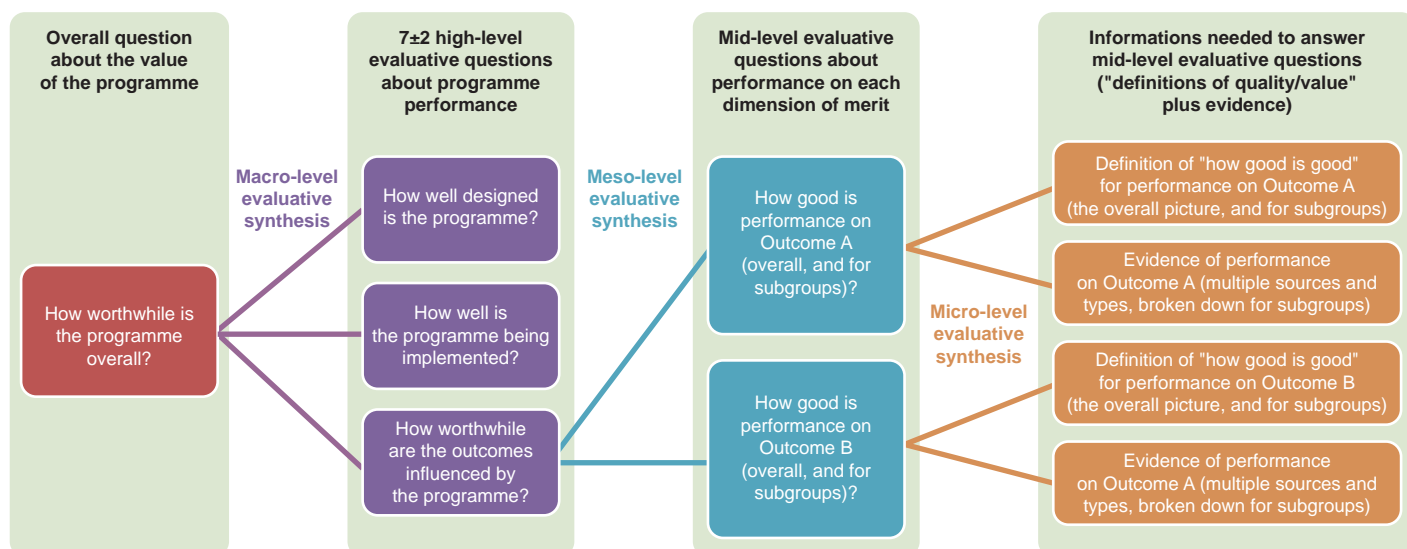
We can think of evaluative synthesis as having three levels – micro, meso and macro, as

the illustration below shows. Conclusions about overall programme value (or worth) are generally determined based on answers to a set of high-level evaluation questions. In order to answer them, we break them down into sub-questions, which are then 'operationalised' – from left to right. Synthesis is the process of packing these evaluative ingredients back together – from right to left – to directly answer the high-level evaluative questions.

All evaluations, by definition, should be answering questions about quality, value and/or importance, so all evaluations will need micro- and meso-level synthesis. This holds not just for summative evaluation, but also for formative and developmental evaluation.

The main steps involved in evaluative synthesis are¹:

1. Develop a set of high-level *evaluative* questions to guide the whole evaluation (e.g. how well is the programme designed? How worthwhile are the outcomes?), broken down into sub-questions as appropriate.
2. Identify the appropriate bases for defining what a "worthwhile outcome" and a "good programme design" (etc.) should look like in this context (i.e. where the 'values' come from, and why).
3. Devise one or more evaluative rubrics to define what the mix of evidence would look like if the answers to your evaluative questions and sub-questions are "excellent", "unacceptable", etc.



Source: Actionable Evaluation workshop materials (in-house training), ©2013 Real Evaluation Ltd., used with permission.

4. Design the right mix of data capture instruments and protocols to gather evidence that goes straight to the heart of quality (as defined in the evaluative rubrics).

5. Conduct micro-level synthesis, i.e. interpret evidence using the evaluative rubrics (may be done independently by the evaluation team or in collaboration with stakeholders).

The 'output' at this point is a set of explicitly evaluative conclusions pertaining to lower-level questions and/or specific outcomes or other criteria. The next steps involve aggregating these micro level conclusions up to give answers to your high-level questions that guide the entire evaluation.

6. Strengths, weaknesses, successes and disappointments are far more actionable if we know which are critically important, which are important but not key, and which are plusses or minuses but not that important. Use needs assessment and other relevant evidence to transparently determine which are which.

7. Having worked out the relative importance of evaluative findings, determine how the evidence on multiple dimensions should be synthesised to generate answers to the high-level evaluative questions. [This is not just a 'weight and sum' exercise; there are many different synthesis options

depending on the evaluative reasoning that makes the most sense.]

8. Drawing on relevant expertise and stakeholder input as appropriate, consider the evidence and conduct the meso-level synthesis.

9. If one of the high-level questions pertains to the overall worth of the entire policy or programme a macro-level synthesis is also required. 'Paint a picture' of what programme performance will look like overall at various levels (i.e. a whole-programme rubric) and then apply it to draw an overall conclusion.

Evaluative synthesis is the antidote for evaluations that get lost in indicators, metrics and stories. It helps get clear, straight-to-the-point – but not overly simplistic – answers to important evaluative questions. Building evaluative synthesis expectations into Terms of Reference is a critically important first step. Many evaluation commissioners will need expert advice in order to do so, and to ensure expectations are met as the evaluation project unfolds. So will evaluators still learning to master these methodologies themselves.

References

BetterEvaluation. *Synthesize data from a single evaluation*. <http://betterevaluation.org/>

plan/synthesize_value/synthesize_data_single_evaluation

[Includes a downloadable 2-page PDF summarising the main options for evaluative synthesis.]

Coyle, J., Hutchinson, K., & Davidson, E. J. (2013). *Evaluation rubrics with E. Jane Davidson*. Adventures in Evaluation podcast series. <http://adventuresinevaluation.podbean.com/2013/12/08/evaluation-rubrics-with-e-jane-davidson/>

Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.

Davidson, E. J. (2013). *Actionable evaluation basics: Getting succinct answers to the most important questions* [minibook]. Real Evaluation. [Also available in Spanish, *Principios básicos de la evaluación para la acción*, and coming soon in French, *Les essentiels de l'évaluation tournée vers l'action*.] <http://realevaluation.com/read/minibooks/>

Davidson, E. J. (2014, forthcoming). *Evaluative Synthesis*. UNICEF Office of Research Methodological Brief No. 10.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Thousand Oaks, CA: Sage.

I These are presented in more detail in *Actionable Evaluation Basics* and in a forthcoming UNICEF methodological brief on this topic.

THE TRICKY TRIANGLE – EVALUATOR, EVALUAND, EVALUATION COMMISSIONER¹

Bastiaan de Laat

The evaluation literature generally distinguishes between two main roles in the evaluation process: the *evaluator* and the *evaluand*. The former performs the evaluation. The latter is the evaluated entity, e.g. a project, a programme, a policy or an institution. The evaluator can be either internal or external to the organization responsible for implementation of the evaluand. Traditionally one therefore contrasts “internal evaluation” with “external evaluation.”

Within this dichotomy, the role of the *evaluation commissioner* is overlooked. This article proposes a simple analytical framework to analyze the evaluation commissioner role alongside the roles of evaluator and the evaluand: the *tricky triangle*. It is tricky because of the sensitive and sometimes conflicted relationships among the three entities. It is also tricky because the roles each of the three are expected to play often turn out to be quite different from what is generally supposed. In particular, externality does not guarantee independence.

Assuming three different roles for actors belonging to one, two or three different organizations, simple arithmetic leads to five possible configurations. Their treatment below is of course stylized and it does not exhaust all the possible arrangements within an organization.

Configuration I: Threefold separation. This is where roles are institutionally separated, i.e. evaluation commissioner, evaluator and evaluand belong to three distinct organizations. For example, a national ministry may commission an external consultancy firm to evaluate a program it is financing but which is implemented by a separate agency. This configuration is common in northern EU countries, where much policy implementation is traditionally performed by implementing agencies independent from, yet steered by, national ministries.

In terms of ensuring credibility and objectivity, this configuration is often considered

ideal, since the roles of evaluation commissioner and evaluand are separated thus enhancing the likelihood that the evaluation will be independent. There are pitfalls however, e.g., the evaluation commissioner may adapt the Terms of Reference (ToR) to its own agenda; influence the evaluation questions to be asked or prescribe the methodology to be adopted. Or the fee dependent evaluator may be induced to avoid sensitive issues or denied access to evaluand information.

Configuration II: Self-Evaluation. This is when the evaluator and the evaluand overlap, i.e. when the persons responsible for programme implementation are mandated to perform the evaluation. This is the central idea behind the *self-evaluation* concept and it often materializes when the evaluand reports to a third party. In such cases no evaluation commissioner is needed and self-evaluation practice is ruled by internal or external procedures.

The credibility of this approach is determined by the extent to which self-evaluators play the game seriously and honestly. The rationale behind it is that people and organizations want to learn and progress through self-examination and analysis of their achievements against set objectives. Therefore, this configuration is most appropriate for formative or developmental evaluation rather than accountability. It may also be used as a complement to other configurations, e.g. when an internal independent evaluation unit attests to the quality of self-evaluation products and processes.

Configuration III: Evaluation Commissioner = Evaluator. This configuration occurs when the evaluation commissioner and evaluator coincide but the evaluand is separate. While odd, this set up is not uncommon. It exists for instance within some bilateral development agencies, i.e. when the evaluand is the program or project in a partner country and the evaluation is carried out by an evaluation department located *within the funding agency*.

This often leads to hybrid configurations. First, boundaries may be blurred since part of the implementation responsibility lies in-house so that the agency may be vulnerable to bias given its *parental* role so that the configuration may be perceived as a variant of Configuration IV below. Hybridisation may also occur when external evaluators are brought in – reverting back to Configuration I. Finally, even if evaluation work is contracted out, the internal evaluator may retain a substantial role (e.g. in setting ToRs or taking responsibility for part of the fieldwork or the final evaluation report). In such circumstances, the internal evaluation department takes the intellectual lead and the resulting configuration differs from Configuration I and IV even though an external consultant is involved.

Configuration IV: Internal Evaluation: All Within One Organisation. When evaluation commissioner, evaluator and evaluand are located within the same formal organization, the literature usually refers to *internal* evaluation. This is often the case within multilateral development banks (MDBs) and other multilateral organizations. Nevertheless, the three roles are generally well separated, with evaluation under the responsibility of dedicated evaluation *units* or *offices* reporting to executive boards, not to the management of the organization, and financing or implementation of programs under the responsibility of operational services.

It has been suggested this configuration causes role pressure for evaluators embedded in the very setting that they are evaluating. However, the likelihood of a positive bias has less to do with the fact that the evaluation unit is internal to the organization than to the adequacy of safeguards that are put in place to guarantee independence. “Good practice standards” that guarantee the independence of internal evaluation units exist² but strict adherence to them is not a given.

Configuration V: The Evaluand as Commissioner. This is when evaluation commissioner

and evaluand are part of the same organization and the evaluator is external. This structure is *the least desirable* from the point of view of evaluator independence and accountability as criticism is generally difficult for the evaluand to accept and evaluand influence on the evaluator is strong. A variant of this configuration involves evaluations commissioned and managed by *internal evaluation units* which belong to the same organization but are not directly involved with the evaluated programs. These act as a buffer between the external evaluator and program management with varying degrees of effectiveness depending on the degree to which the internal evaluation unit is independent.

Conclusion. The Tricky Triangle model provides a framework that facilitates systematic exploration of real-life situations. It brings out the risks that threaten evaluation independence. It identifies the strengths and weaknesses of different configurations and points to possible measures that enhance the integrity of the evaluation process. It demonstrates that evaluator independence can only be partly understood by applying the traditional internal/external distinction still widespread in the literature. Whether evaluation is internal or external is not the key to understanding the tensions inherent in the tricky triangle of evaluation governance. Resolving such tensions hinges on the roles actually played by each of the three parties.

Disclaimer: Any views and opinions expressed by the author do not necessarily reflect the views and opinions of the European Investment Bank.

1 A fuller treatment of this topic is in Chapter 2 of Marlène Läubli Loud & John Mayne (eds), *Enhancing Evaluation Use: Insights from Internal Evaluation Units*, Sage, 2013 (pp15–36).

2 In 2003 the Evaluation Cooperation Group (ECG) of the multilateral development banks adopted a framework for assessing evaluation independence along four dimensions: organizational, behavioral, avoidance of conflict of interest, and protection from outside interference.

ADDING RIGOUR TO THE EVALUATION OF MULTI-ACTOR PROCESSES: SOME REFLECTIONS ON THE NON- USE OF GAME THEORY AND RELATED ACTOR ANALYSIS METHODS¹

Leon Hermans

The evaluation community shows a great concern for analytical rigour in assessing impact in order to provide evidence-based information to policy makers about relevant, effective and efficient practices. Still, we should not want evaluations to stop there. True accountability and learning demand an *explanation* of observed impacts, for which we should strive to achieve similar levels of analytical rigour as for establishing impacts.

Many impacts come about through implementation trajectories that involve various organizations and different people in different roles. Explaining the outcomes of such multi-actor processes is different from establishing a link between some of the inputs (programmes and policy choices) into these processes and some of their outputs. Explaining why a link exists requires methods that look into the (multi-actor) processes through which inputs are transformed into outputs.

This is an area where important gains can still be made. For instance, Abhijit Banerjee and Esther Duflo make a compelling case for the use of RCTs to establish impacts of development interventions². But when it comes to explaining impacts, or arriving at testable hypotheses, they often fall back on anecdotal evidence and chance encounters with individuals. The resulting explanations may be rigorous and grounded in theory but they are not transparent.

Adding rigour to the *explanation* of observed impacts calls for methods based on models of actor interactions solidly grounded in theory and validated by prior social science research in a variety of contexts. Methods that meet these requirements may be labelled “model-based actor analysis” tools. The underlying models help to develop informed and test-

able hypotheses and/or to build explanatory models of observed impacts. Game theory methods fit in this domain, as do social network analysis, comparative cognitive mapping, argumentative analyses or transaction models.³ Surprisingly they have not been widely used in the evaluation domain. There are exceptions but far fewer than warranted⁴.

Why some programmes or policies are explicitly aimed at actor-processes, at building networks or strengthening communities. For these, actor analysis and social network methods are in fact being used and increasingly so... However, many other social programmes and conventional interventions aimed at increases in income, food security, health services or water availability work through actors and would benefit from the use of game theory tools and related methods. For instance, impacts in education come about as a result of interactions between teachers and students within a setting shaped by parents, school, the Education Ministry, inspectorates, the developers of educational materials, to name just a few. Shouldn't these processes warrant looking into? What assumptions regarding factors within the authorizing environment and the decision making processes help explain outcomes? Shouldn't we look into the “black-box of implementation”? How can the interactions that shape the outcomes of implementation processes be made visible?

Some evaluators recognize the importance of actor interactions but do not attempt to analyse them with any rigour. They view human interaction processes as irrational and emotional. They consider that the context is so rich and complex that it defies any attempt at rigorous analysis. To be sure such intricacies and contingencies are invariably part of

the reality mix, and they limit the accuracy of any actor analysis. However, this should not be a reason to stop trying. Even if there is only a small amount of evidence at hand, structuring available information in a model (e.g. for a game theory model) would help logical reasoning. Such models can be useful through detecting flaws in reasoning or ruling out certain outcomes out of the full range of possibilities. It may suggest that we cannot expect a policy or programme to achieve certain outcomes, once we take into account key actors' motivations and constraints.

Another possible explanation underlying the neglect of game theory and related methods is that it is not favoured by standard evaluation processes⁵. Evaluation commissioners do not normally request evaluations that look into actor interactions, or make use of game theory and related approaches. Nor are evaluators prone to propose such methods. Some may be unaware of their existence, or may feel unskilled in their use. Others may see it as risky since their intended audience is unfamiliar with or resistant to such methods. Their hesitation is understandable but game theory has been around for more than half a century and arguably predates contemporary evaluation practice. Important contributions to game

theory, network analysis and argumentative analyses were made in the 1950s and 1960s and they have long been standard analytical tools for such renowned policy analysis institutes as the RAND Corporation.

Why then aren't commissioners prone to request the use of such methods? And why aren't evaluators proposing these methods, even if not explicitly requested to do so? What underlies their reluctance to open up the black-box of implementation? This may well have to do, at least in part, with the well-known political dimension in evaluation. Decision-makers may not like to be confronted with the hidden motivations of their actions. They may prefer to contend that it is mainly the results that count. While they are ready to appropriate programme successes they tend to ascribe failures to others or to point to uncontrollable outside forces. In other words, explaining "why" may mean venturing into zones of potential discomfort. Thus and paradoxically game theory may well explain why game theory methods are rarely used in evaluation. It follows that evaluators intent on using game theory methods should adopt strategies that take account of the objective functions that animate programme stakeholders in order to help secure win-win outcomes and enhance public welfare.

- 1 This article draws on Hermans L., Cunningham S., and Slinger J, The usefulness of game theory for policy evaluations. *Evaluation*, Sage Publications, January 2014, vol. 20, no. 1: 10-25.
- 2 Banerjee, A.V. and Duflo E., 2011, *Poor Economics: A radical rethinking of the way to fight global poverty*, Public Affairs, New York
- 3 For overviews of actor analysis methods see Hermans & Thissen, 2009 in *European Journal of Operations Research*, 196, pp. 808–818; and Hermans & Cunningham, 2013, in Thissen & Walker (Eds.) *Public Policy Analysis*, Springer, pp. 185–213.
- 4 Stakeholder analyses as part of the early or preparatory phase of an evaluation are fairly common and very useful but they serve a different purpose: they are not meant to explain observed programme outcomes or impacts. Similarly, participatory approaches help to involve actors and focus evaluations on issues of interest, but do not seek to analyse the actor interactions as source of explanation of outcomes and impacts.
- 5 This hypothesis is my interpretation of a judicious comment offered by an anonymous Evaluation journal peer reviewer.

LESSONS FROM PERFORMANCE MEASUREMENT IN ASSESSING RESEARCH

Sirkka Immonen and Leslie J. Cooksy

In the past few decades performance measurement (PM) systems have been employed in the public sector with the aim of enforcing accountability and enhancing efficiency and effectiveness. PM has also been applied to research. Yet, the suitability of PM to research has been much debated, given that research is by its very nature exploratory, causality from activities to outcomes is uncertain and protracted, and often serendipity may lead to long term or unplanned effects.

The potential problems of PM, particularly if mechanisms of direct rewards and sanctions

are put in place, include: perverse or unclear incentives; goal replacement away from the intended mission towards what is measurable; risk-avoidance; reduced ambition and innovation; less responsiveness to emerging opportunities; neglect of organizational learning; increased focus on competition, regulation and control; bureaucratization; and higher costs (Perrin, 1998, Smith, 1995, Van Thiel and Leeuw, 2002). Annual monitoring of performance indicators is a common practice in results-based management but in several respects, including research relevance, quality and intellectual influence,

research does not lend itself to effective monitoring through indicators (Cozzens, 1997; Feller 2002).

The experiences from a PM system applied to 15 autonomous international agricultural research centres of the Consultative Group for International Agricultural Research (CGIAR) are discussed by Immonen and Cooksy (2014)¹. This paper presents a brief overview of the lessons. The CGIAR has distinctive features: it engages in research done in partnership with a range of stakeholders. Its research is aimed at development

outcomes and the resulting causal pathways are fundamentally complex and non-linear. There are complex interactions among many autonomous and strategic actors.

The PM system installed at CGIAR was intended to become part of a streamlined monitoring and evaluation system and to enhance transparency, accountability, learning, and decision-making, including decisions about future funding (CGIAR, 2003). Its introduction was seen as following a global trend aimed at improving efficiency and accountability for results. The actual experience, elaborated below, shows that: (i) the large year-to-year fluctuations that were observed were probably related to the selected indicators rather than reflecting actual performance, thus undermining the validity of conclusions about performance; (ii) the fact that the indicators influenced resource allocation among centres had a strong negative effect on the way the indicators were perceived; and (iii) the PM information was not used for evaluation.

The PM system, developed under the oversight of the World Bank, was initiated in 2005 and operated for six years. Its development and objectives are described by Iskandarani and Reifschneider (2008). It was designed to capture the entire chain of causality from resources and inputs to outputs, outcomes and impacts. All indicators were common to all research centres irrespective of their line of research (whether, for instance, policy, crop breeding, natural resource management or biodiversity conservation).

Selection of judicious results indicators equally applicable across the diverse centres proved challenging. In the timeframe of annual monitoring, actual research results could not be reported meaningfully. Therefore, indicators for outcomes and impacts focused on the outcome and impact culture, such as monitoring along the impact pathway and the rigor of impact assessment. Those indicators were nevertheless mis-interpreted among some donors as reporting actual results. For outputs, bibliometric measures of research, such as rate of publishing, impact factors of journals and citations are a very partial proxy of relevant outputs for mission-oriented research programs. Therefore, the indicator for outputs was initially a measure of each centre's achievement of

its planned outputs (such as genetic materials, policy options, practices or capacity) targeted for the year of reporting. All centres reported near complete achievement of their output targets. This suggests that the selected output indicator generated perverse incentives since research always involves some risk of failure and degree of unexpected results, and it is not feasible to set a threshold for satisfactory achievement of outputs. Even an implicit expectation of full achievement of outputs may undermine the research process and discourage the identification of meaningful goals. It also potentially stifles the risk-taking behaviour that is essential to the scientific endeavour.

The results over six years showed strong annual fluctuation for most indicator results, which is typical in research. Notwithstanding the fluctuations and the acknowledged need to learn, and adjust the system, the results were presented to the donors annually for their use at face value. Intentionally, indicators had no benchmarks or targets so that donors could interpret the results and give weights to the indicators following their own predilections. At least one donor used the indicator results both for rating the centres as outstanding, superior or satisfactory and for determining its annual fund allocation on basis of its weighting of the indicators. While comprehensive conclusions cannot be drawn of the positive or negative effects of the PM experiment, its limited use by centres in decision-making, in evaluation and by donors (apart from one) suggests limited relevance. The fact that 13 out of 15 centres reported a drop in output achievement once the achievement-based output measure was moved out of the PM system, while achievement data were still collected, is a further indication that the indicator distorted incentives.

The centres identified several other pitfalls including lack of clarity regarding expected performance changes; misalignment between indicators and CGIAR's mission; promotion of parochial attitudes (single centre mindset); competition rather than partnership and collaboration; and risk of atomizing the centres' work rather than helping scale it up. Some centres indicated that the PM system helped direct more management attention to results: outputs, outcomes and impacts. However, for learning, there should

have been more internal analysis among researchers and managers of the linkages from activities to outcomes to increase their confidence on the performance measures. This is also emphasised by Poister (2010). Instead, donor requirement for accountability became the primary incentive for reporting on the results indicators.

There are inherent challenges in PM aimed at serving – with the same set of indicators and at the same time – multiple objectives, particularly accountability and learning, and multiple stakeholders, particularly research management and donors. Annual monitoring of some aspects of on-going research and operations is necessary and suited for management decisions and learning and for establishing data series for trends analysis, but of limited use for external audiences to act upon in a top-down process. Periodic studies of, for instance, adoption of new technologies and their effects on different beneficiary groups and environment are more appropriate for documenting results and generating lessons on the cause-effect linkages. For research that is aimed at changing both behaviour (for instance policy-makers, farmers, consumers, market actors) and state (for instance natural resource flows, productivity, nourishment) such studies provide a more credible basis for judging the effectiveness of long-term research programs. However, short-term performance monitoring and retrospective studies for results monitoring should not replace or compete with evaluation; at best monitoring and evaluation are complementary (Lepori and Reale, 2012). Both monitoring and evaluation should support analysis of the program from design to outcomes where learning and evidence of achievement contribute to enforcing or adapting the theory underlying the program.

References

- CGIAR (2003) *Towards Designing a Performance Measurement System for the CGIAR*. Washington, DC: CGIAR Secretariat. http://www.cgiar.org/www-archive/www.cgiar.org/pdf/agm03/agm03bus_wgpm_report.pdf
- Cozzens SE (1997) The knowledge pool: measurement challenges in evaluating fundamental research programs. *Evaluation and Program Planning* 20(1): 77–89.

- Feller I (2002) Performance measurement redux. *American Journal of Evaluation* 23(4): 435–52.
- Iskandarani, M. and Reifschneider, F.J.B. (2008). Performance measurement in a global program: motivation, new concepts and early lessons from a new system. *Science and Public Policy* 35(10): 745–55.
- Lepori B and Reale E (2012) S&T indicators as a tool for formative evaluation of research programs. *Evaluation* 18(4): 421–465.
- Perrin B (1998) Effective use and misuse of performance measurement. *American Journal of Evaluation* 19(3): 367–79.
- Poister TH (2010) Performance measurement: Monitoring program outcomes. In: Wholey JS, Hatry HP and Newcomer KE (eds) *Handbook of practical program evaluation, 3rd edition*. Jossey-Bass: San Francisco, 100–24.
- Smith P (1995) On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration* 18(2–3): 277–310.
- van Thiel S and Leeuw FL (2002) Public sector paradox. *Public Performance & Management Review* 25(3): 267–81.

I Immonen, S. and Cooksy, L.L. (2014). Using performance measurement to assess research: Lessons learned from the international agricultural research centres. *Evaluation* 20(1):96–114.

EVALUATION USE IN THE ORGANISATIONAL CONTEXT – CHANGING FOCUS TO IMPROVE THEORY¹

Steven Højlund

Introduction

Evaluation use is probably the most researched theme in the literature on evaluation and much of it confirms that evaluations rarely change policies. This constitutes a paradox, since the very objective of evaluation is to improve policy. And so, one may well ask, why evaluate if evaluation is not used after all?

Understanding the paradox

Evaluation was born at a time of belief in a better world achieved through rational interventions and social engineering (Vedung, 2010). Therefore evaluation is inherently rationalist, causal and evolutionary in nature (Sanderson, 2000). The often-stated purpose of evaluation and its rationale are rooted in assumptions of rationality and causality. Evaluation is commonly conceived as a tool informing policy-makers and civil servants of what works and what does not. This causal relationship between evaluation and policy improvement is apparent considering mainstream definitions of evaluation: '*evaluation is] systematic inquiry leading to judgements about program (or organisation) merit, worth, and significance, and support for program (or organisational) decision making*' (Cousins et al., 2004: 105; see also Vedung, 1997).

It is clear from the above that evaluation use is an intrinsic part of the evaluation process. However, this brings out a logical paradox since most evaluations are unused and cannot therefore be called evaluations according to the definition. The root of the paradox lies in the fact that most evaluation models use the logic of cause and effect (see e.g. Mark and Henry, 2004: 38). This is ontologically and epistemologically linked to realist and positivist understandings commonly found in classical economic thinking and rational choice theory where all actors are perceived to be in rational pursuit of goal-attainment through utility-maximising behavioural patterns (Sanderson, 2000). In turn these assumptions contain an underlying positive and evolutionary assumption of progress and social betterment as the ultimate objective of evaluation (Mark and Henry, 2004).

What about the justificatory uses?

The fundamental assumption of linear cause and effect is intrinsically linked with instrumental use: ideally, evaluation improves policy through the instrumental application of an evaluation's results (conclusions and recommendations) to improve policy. However, there is a logical problem between the general assumption of linearity towards social

betterment and the fact that evaluations are also used to justify policies and organisations. There is no doubt that justificatory uses exist and should be kept as separate use categories apart from instrumental uses. Nevertheless, it is very hard to argue that symbolic and legitimising use of evaluation lead to social betterment. For example, water quality is not improved because an aid organisation uses evaluation to legitimise itself. The argument of rationality would need to be stretched very far to argue that symbolic and legitimising evaluation use can actually improve living conditions for human beings through learning etc.

Contrary to the main assumption that evaluations are undertaken to improve policies, evaluating organisations or actors are likely to have mainly justificatory uses in mind for their evaluations. It seems that the evaluation literature empirically has recognised justificatory uses, but theoretically has disregarded the implications of this acknowledgement for the overall rationalist and causal foundation of evaluation thinking.

Paradox – and so what?

The paradox explained above is well described by the literature on evaluation use. However, it remains to be seen how justificatory use types can be properly integrated

into a model of evaluation use. It seems unsatisfactory to empirically acknowledge justificatory uses of evaluation and widespread non-use of evaluations while not having a good explanation for the phenomena.

The problem might be that the evaluation use literature often focuses on the evaluation itself – its implementation, its outputs, its conditioning factors, etc. – and largely ignores the organisational context. However, on the organisational level, there might be forces at play that are so strong that they overrule a persuasive evaluation result and cancel instrumental use.

In this article, the claim is that organisational factors are more important, in terms of explaining evaluation uses and non-uses, than the evaluation literature so far has acknowledged. It is argued that an organisation's concerns of its external legitimacy are likely to have priority over evaluation findings and thus in certain cases lead to non-use of these findings no matter how relevant they might be to the organisation.

Looking for a solution

Selznik (1949) and a large number of later 'institutional' studies observed that organisations act contrary to the rational utility-maximising behaviour that is expected and contrary to the objectives organisations set themselves. Instead, organisations mostly act according to norms and values in their environment in order to legitimise themselves (Scott and Meyer, 1994). This is particularly the case if the organisation perceives itself to be operating in an unstable environment, where there is uncertainty about the distribution of resources between itself and rival organisations. DiMaggio and Powell (1991: 190) argue that organisations are concerned with reducing uncertainty and stabilising social relations and thus not primarily with maximising power or maximising outputs as, for example, a rational-choice perspective would be arguing. Hence, symbolic and legitimising uses of evaluation are in line with the most basic argument of organisational in-

stitutionalism that organisations have a need to legitimise themselves in order to survive as organisations.

The institutional context influences evaluation use and the organisational context has not received the attention this article claims it should have with regard to explaining evaluation use. Similarly, Boswell (2008: 473) argues that: 'Any account of how organizations use knowledge will inevitably be premised on a theory of organizations: a set of claims about the sources and nature of organizational interests, and how these translate into organizational action.' Along the lines of that argumentation, the argument put forward here is that evaluations are almost exclusively embedded in organisational contexts one way or the other.

Typically, evaluations are procured by organisations, carried out by organisations (enterprises or teams of individuals), may be read and acted upon by a third organisation, etc. Nevertheless, the majority of literature on evaluation use focuses on identification of uses, influence and factors and conditions related to the evaluation. Hence, as Ledermann (2011: 160) argues, maybe 'it is time to abandon the ambition of finding "the important" characteristic for use and to adopt a focus on context-bound mechanisms of use instead.'

Conclusion

This brief article investigates one of the old paradoxes of evaluation use, namely that evaluations are rarely contributing to policy-making even though this is the main reason for undertaking them². The paradox consists of a logical inconsistency between, on the one hand, the causal assumption that evaluation leads to policy-improvements, and on the other hand, that the empirical literature displays the widespread prevalence of non-use and justificatory uses that do not lead to policy-improvements or social betterment.

We learn that: 1) justificatory use types and non-use need to be better integrated

into a theory of evaluation use; and 2) that a theory of evaluation use should take into consideration the organisational and institutional context of the evaluating organisation.

References

- Boswell C. (2008) The political functions of expert knowledge: knowledge and legitimation in European Union immigration policy. *Journal of European Public Policy* 15: 471–488.
- Cousins JB, Goh SC, Clark S, et al. (2004) Integrating evaluative inquiry into the organisational culture: A review and synthesis of the knowledge base. *The Canadian Journal of Program Evaluation* 19: 99–141.
- DiMaggio PJ and Powell W. (1991) Introduction. In: Powell WW and DiMaggio PJ (eds) *The new institutionalism in organizational analysis*. Chicago: University of Chicago Press, 1–38.
- Ledermann S. (2011) Exploring the necessary conditions for evaluation use in program change. *American Journal of Evaluation* 33: 159–178.
- Mark MM and Henry GT. (2004) The Mechanisms and Outcomes of Evaluation Influence. *Evaluation* 10: 35–57.
- Sanderson I. (2000) Evaluation in Complex Policy Systems. *Evaluation* 6: 433–454.
- Scott RW and Meyer JW. (1994) *Institutional environments and organizations: Structural complexity and individualism*, Thousand Oaks: SAGE Publications.
- Selznik P. (1949) *TVA and the Grass Roots*, Berkeley: University of California Press.
- Vedung E. (1997) *Public Policy and Program Evaluation*, New Brunswick: NJ: Transaction Publishers.
- Vedung E. (2010) Four Waves of Evaluation Diffusion. *Evaluation* 16: 263–277.

1 This article is based on an article issued in a Special Issue of Evaluation: Højlund, Steven (2014) Evaluation use in the organizational context – changing focus to improve theory. SAGE: *Evaluation* 20 (1)

2 A full elaboration of this argument is included in the Special Issue of Evaluation: Højlund, Steven (2014) Evaluation use in the organizational context – changing focus to improve theory. SAGE: *Evaluation* 20 (1)

CHALLENGES OF EVALUATING ENVIRONMENT IN INTERNATIONAL DEVELOPMENT

Juha I. Uitto

Evaluation of environment in international development poses specific challenges. Some of these stem from the specificities of environmental phenomena and are thus common to evaluating any environmental programme (Birnbaum and Mickwitz, 2009). These may include factors, such as the different time horizons for observing changes in natural and social systems, exacerbated by the often short policy and programme cycles; or the disparities in geographical scales as environmental problems, such as deforestation or climate change, often do not match jurisdictional boundaries. There are also issues pertaining to data availability, especially at the local level. Furthermore, climate change necessitates a greater focus on risk, uncertainty and complexity (Picciotto, 2007). Because of these and other factors, environmental programmes in complex situations do not easily lend themselves to experimental or quasi-experimental designs (Vaessen, 2011). It has been suggested that a two-system evaluation approach that takes into account the different temporal and spatial frames of the human and natural systems may be appropriate (Rowe, 2012).

In evaluating environment in the international development context, it is essential to move beyond individual interventions, while project and programme evaluations will continue to be important for accountability. Evaluation must address the 'micro-macro paradox,' i.e. that while many individual interventions seem to be successful, there is little to show by way of solving the larger problems at national and global levels. Global environmental degradation and climate change fall into this category (van den Berg, 2011). There is a need to frame evaluation questions in terms of whether the interventions make adequate contributions at the macro level and whether the programmes and projects support national development efforts. International development agencies also have a responsibility to contribute to broader

goals and sometimes there are tensions between local, national and global priorities.

Two concrete cases are highlighted to demonstrate the above challenges in practice. The United Nations Development Programme (UNDP) is one of the main implementing agencies of projects funded by the Global Environment Facility (GEF). While it can be argued convincingly that environmental and development issues are closely related, there is a clear difference in the mandates of the two entities. UNDP's main mandate pertains to sustainable human development and poverty reduction in the countries it works in; the GEF goal is to address global environmental issues (including biodiversity conservation and climate change mitigation), while supporting sustainable development at the national level. The tendency of UNDP country programmes to substitute work on local environmental issues—such as sustainable management of natural resources and energy at the local level—with projects funded by the GEF has been found to have the effect of distancing environmental programming from the poverty reduction focus. The presence of this gap in integration has significance for evaluating UNDP work at the country level. It is important to be explicit about what we evaluate against. As worthy as both UNDP and GEF goals and mandates are, they are not identical. It is thus not adequate to evaluate GEF-funded projects implemented by UNDP only in relation to how they contribute to the generation of global environmental benefits. The expected value added by UNDP should be manifested in complementarities between the global, national and local benefits achieved through integrated programming.

The question of local versus global benefits is pertinent in the context of the GEF Small Grants Programme (SGP), also implemented by UNDP. Launched in 1992, SGP has funded over 17,000 projects in more than 120 coun-

tries providing direct support to local communities to take action in the global environmental focal areas. SGP being a GEF-funded programme, there has been a demand to measure its results at the aggregate level with regard to biodiversity conservation, carbon emissions avoided, international waters, land degradation and sustainable forest management, and chemicals management. The feasibility and meaningfulness of such aggregation can, however, be questioned, in particular because monitoring and collecting this data from the thousands of small community projects is very burdensome and by necessity based on self-reporting. Evaluative evidence suggests that SGP projects have allowed communities to move towards sustainable use of their resources and to embark on economic development activities that benefit both the communities and the environment. It would appear fair to include the results on the wellbeing and livelihoods of poor and vulnerable communities, as well as the replication and scaling up of the projects and approaches, into the evaluation of its results. This would give a better sense of the overall impact SGP has had beyond its direct investment into global environmental benefits. Arguably, the most significant impact could be achieved through policy influence.

Challenges of attribution become harder the further the results are distanced from the direct activities under the control of the programme. This does not mean that it is impossible to conduct rigorous evaluations. Theory-based approaches can be helpful in determining the contribution of the intervention, but it is also important to focus on unintended results and barriers to success (Woerlen, 2013) in order to understand what works, why and under what circumstances.

Aggregation of results to the global level is particularly challenging where an organization operates in a decentralized manner in many countries or areas, like UNDP or SGP.

Programmes and approaches need to be tailored to local contexts to be effective and they seldom should follow exactly the same format, even when aiming at the same goal, such as biodiversity conservation. It is possible to aggregate project outputs, but aggregating the broader results and outcomes is difficult.

It is also important to be clear at what levels we seek to verify results and impact. These can be defined at different levels ranging from the direct results that can be attributed to an intervention, to the higher level of results that may reduce the environmental stressors, to the final impact on the ecosystem. In the cases described above, evaluation focuses largely on the middle ground: the important social, economic, organizational and institutional dimensions that will eventually lead to positive changes at the ecosystem level.

In a constant pursuit for enhanced robustness of its approaches and methodologies, the Evaluation Office of UNDP tackles these issues in its programmatic evaluations at the country, regional and global levels. For further elaboration of the above arguments, please refer to an article in *Evaluation* (Uitto, 2014). All UNDP evaluation reports and management responses are publicly available at <http://erc.undp.org>.

References

Berg, R.D. van den. 2011. Evaluation in the context of global public goods. *Evaluation* 17(4): 405–415.

Birnbaum, M. and P. Mickwitz, eds. 2009. *Environmental Program and Policy Evaluation: Addressing Methodological Challenges. New Directions for Evaluation* No. 122. San Francisco: Jossey-Bass.

Piccioletto, R. 2007. The new environment for development evaluation. *American Journal of Evaluation*, 28(4): 509–521.

Rowe, A. 2012. Evaluation of natural resource interventions. *American Journal of Evaluation*, 33(3): 382–392.

Vaessen, J. 2011. Challenges in impact evaluation of development interventions: Randomized experiments and complexity. In: Forss, K., M. Marra and R. Schwartz, eds., *Evaluating the Complex: Attribution, Contribution, and Beyond*. New Brunswick and London: Transaction Publishers, 283–313.

Uitto, J.I. 2014. Evaluating environment and development: Lessons from international cooperation. *Evaluation*, 19(1).

Woerlen, C. 2013. The theory of no change. *Evaluation Connections*, September 2013, 9–10.

SISTEMATIZACIÓN: WHAT'S THAT? INTRODUCING A MULTI-STAKEHOLDER, REFLEXIVE ACTION-LEARNING EVALUATION APPROACH.

Esteban Tapella and Pablo Rodríguez-Bilella

A wide range of Latin American development experience is insufficiently known or not properly valued. Beyond facts, specific events or data they consist of vital processes that combine objective and subjective dimensions of diverse socio-historical realities, contextual conditions, particular situations, actors' actions, perceptions and interpretations, results and effects, and their interrelation.

Conventional evaluation approaches have applied linear cause-effect logic models focused on measuring performance and success in their attempt at demonstrating accountability to external authorities. These evaluations often come up short because development interventions are multifaceted involving different actors, interests and values. They constitute complex systems that emerged following turbulent scenarios so that many factors are likely to shape their outcomes

whether or not they were intended by the project.

In most cases the protagonists consider that learning is as important if not more important than accountability. A rich evaluation literature confirms that in such circumstances traditional evaluation practice cannot readily capture or map the complex systems dynamics and interdependencies (Patton, 2010; Funnell and Rogers, 2011; Williams and Hummelbrunner, 2010). This is where *sistematización* comes in. Its record over the past forty years deserves greater recognition.

The *sistematización* approach is a multi-stakeholder evaluation tool that emphasises in-depth comprehension of processes and shared learning among the participants of development experiences as they unfold. The approach is based on the notion that ex-

periences can be used to generate on-going understanding(s), and that lessons learned in real-time can improve both ongoing implementation and contribute to a wider body of knowledge.

Given its sharp focus on practice and despite its deep roots and long history the *sistematización* approach has not generated a full or formal prescriptive theory. This problem is not unique to Latin America. Other developing regions have also failed to disseminate effectively their comprehensive approach to development and their distinctive approach to evaluation. Thus the second edition of "Evaluation Roots" (Alkin 2012) lacks a chapter on development evaluation theories nurtured within low or middle income countries. In a recent article we explored the *sistematización* approach (Tapella and Rodríguez-Bilella, 2014). We will highlight here its main characteristics.

Although much reflection and work has gone into refining it as a method and more recently to describe its history and roots, the dialogue and interaction of *sistematización* with other evaluation approaches has been quite limited. Due to its practical origins, its links with social science theory have not always been made explicit albeit several authors have noted its affinity with dialectical materialism. On the other hand its potentially fruitful links with systems thinking have yet to be explored.

At the heart of systems thinking – closely linked with complexity theory and its concepts – lies the notion that making sense of the world implies approaching it as a whole and to embrace its interrelationships, instead of dismembering it into its component parts and analysing them in isolation. *Sistematización* shares with systems thinking a reluctance to adopt linear cause-effect logic models. Instead it seeks to capture and map the unique complex dynamics and interdependencies that influence outcomes in specific cases.

In general terms, *sistematización* is the participatory process of on-going description, analysis and documentation of the different aspects of a specific development intervention or experience: its actors, actors' interaction, results, outcomes, impacts, and problems. *Sistematización* is carried out with the active participation of the agents closely involved with the intervention, in order to draw lessons to improve the efficiency and effectiveness.

Sistematización involves a critical reconstruction and interpretation of a development experience, aimed at explaining the distinctive logic of the intervention, the external and internal factors that influenced it, and why it produced the results it did (Jara, 2006).

The *sistematización* method attempts to identify lessons from the experience by describing, organising and analysing a development activity's theory of change and the project approach, the institutional, social and historical context, the relationship between local and external actors, and describing the organizational process, including obstacles and facilitators as well as results and impacts of the experience (Morgan and Quiroz, 1988).

Mutli-stakeholder involvement is crucial in any *sistematización* process, since critical reflection and the identification of lessons cannot be achieved in isolation. It needs to be accomplished with partners and communities. They bring new dimensions and interpretations to the learning process. They contribute narratives that capture the stories of how they lived their experience. The composite of these stories constitutes the multifaceted story of the development intervention. This story is told and heard by all the actors involved. Through interpretation and storytelling *sistematización* empowers stakeholders. In *sistematización*, the boundaries between the inquirer and the object of inquiry become porous and fluid.

Sistematización as a method proposes shared and participatory group dynamics. It creates a space where people can share, confront, and discuss opinions based on mutual trust. They critically engage in diverse interpretations of the development experience and generate mutual and collective learning. Thus it embodies a multi-stakeholder participatory approach that ensures that the voices of all are heard and if necessary amplified especially the poor and excluded.

In this way *sistematización* creates space for all actors to be heard and fully represented in the construction of the story – and its reconstruction. As a method it helps project

staff and stakeholders to carefully track and reflect on meaningful moments and illuminating events in the project's life thus reliving shared experience and producing new knowledge.

References

- Alkin MC (ed) (2012) *Evaluation roots* (2nd ed.). Thousand Oaks, CA: Sage.
- Carden F and Alkin MC (2012) Evaluation Roots: An International Perspective. *Journal of MultiDisciplinary Evaluation*, 8(17).
- Funnell SC and Rogers PJ (2011) *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco: John Wiley/Jossey-Bass.
- Jara O (2006) *Theoretical and Practical Orientations for systematization of experiences*, Electronic Library about Experiences Systematization: www.alforja.or.cr/sistem/biblio.html. Accessed: 23 May 2013.
- Morgan-T ML y Quiroz T (1988) 'Acerca de la Sistematización', en 'La Sistematización de la Práctica: Cinco Experiencias con Sectores Populares'. Buenos Aires, Argentina: HUMANITAS-CELATS.
- Patton MQ (2010) *Developmental Evaluation*. New York, NY: Guilford Press.
- Tapella E and Rodriguez-Bilella P (2014) Shared learning and participatory evaluation: the *sistematización* approach to assess development interventions. *Evaluation* 20 (1) 115–133.
- Williams B and Hummelbrunner R (2010) *Systems Concepts in Action: A Practitioner's Toolkit*. Stanford, California: Stanford University Press.

EES
EUROPEAN
EVALUATION
SOCIETY



Convention Centre Dublin, Ireland, 1–3 October 2014
Pre-conference workshops 29–30 September 2014

11th EES Biennial Conference

Evaluation for an
Equitable
Society

Independence, Partnership, Participation



EVALUATION NEWS

HAVE YOU PARTICIPATED IN AN EVALUATION THAT INFLUENCED PEOPLE'S LIVES AND / OR LED TO SOCIAL BETTERMENT?

Send in a Story Submission Form by April 15, 2014 for a chance to have your evaluation profiled in an international collection of stories about evaluations that make a difference.

Evaluations that Make a Difference: Stories from around the world is a recipient of an EvalPartners Innovation Challenge Award. The project's purpose is to demonstrate to evaluation users (e.g. governments and international organizations, NGOs and advocates) how evaluation can truly be a force for positive change. It will also help evaluators gain insight into evaluation approaches that can influence use and impact.

In May 2014, one to three stories will be selected from each of five regions: Europe, Africa, North America, Asia and Australasia. During the Story Development Phase (June – October 2014) the winning stories will be further developed, using engaging story development techniques. The final stories will be shared through an online platform during EvalYear 2015.

For more information, visit the Project website – www.evaluationstories.wordpress.com – and download the Call for Stories background document.

THE AUTHORS

Leslie Cooksy

is the Evaluation Director at Sierra Health Foundation in Sacramento, California. She has served as president of the *American Evaluation Association* and Associate Editor and Editor of the Ethical Challenges section of the *American Journal of Evaluation*. From 2004–06, she was a member of the Standing Panel for Monitoring and Evaluation of the Consultative Group on International Agricultural Research.

**Leon Hermans**

is an assistant professor in policy analysis at the Faculty of Technology, Policy and Management of Delft University of Technology in the Netherlands. Previously, he worked as an associate professional officer at the Food and Agriculture Organization of the UN in Italy. His research focuses on methods and approaches for monitoring and evaluation in



multi-actor systems. Areas of interest are water and environment, as well as international development.

Steven Højlund

is a Ph.D.-candidate at the Copenhagen Business School specializing in evaluation use and learning from evaluation practices in the European Union. Steven holds a MA in Political Science from the University of Copenhagen and a MA of European Studies from the College of Europe. He has five+ years experience with EU evaluations and policy analysis for COWI Consult in Copenhagen

**Sirkka Immonen**

is Senior Evaluation Office working for the CGIAR's Independent Evaluation Arrangement. In 1999–2013 she worked for the CGIAR's science advisory body coordinating institutional and program evaluation in



the international agricultural research centres of the CGIAR. She was also involved in designing a performance measurement system for research and analysis of performance results.

Robert ('Bob') Picciotto, (UK) Professor, Kings College (London) was Director General of the World Bank's Independent Evaluation Group from 1992 to



2002. He previously served as Vice President, Corporate Planning and Budgeting and Director, Projects in three of the World Bank's Regions. He currently sits on the United Kingdom Evaluation Society Council and the European Evaluation Society's board. He serves as senior evaluation adviser to the International Fund for Agricultural Development and the Global Environment Fund. He is also a member of the International Advisory Committee on Development Impact which reports to the Secretary of State for International Development of the United Kingdom.

Pablo Rodríguez-Bilella

is a consulting sociologist. He is a board member of IDEAS (International Development Evaluation Association), and has been board member of ReLAC (the Latin-American Network of Evaluation), and IOCE. He is a lecturer of Social Anthropology at the Universidad Nacional de San Juan, and researcher at the National Research Council of Science and Technology (CONICET), both in Argentina. Pablo blogs at <http://albordedelcaos.com/>, and his contacts are Email: pablo67@gmail.com; Twitter: @txtPablo

**Dr Martin Reynolds**

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

**Esteban Tapella**

is a lecturer and researcher at the Universidad Nacional de San Juan, Argentina. During the last 10 years he has been involved in the field of monitoring, evaluation and *sistematización*. He is a board member of ReLAC (the Latin American Network of Evaluation) and IOCE (the International Organization for the Cooperation in Evaluation). Esteban Tapella's contacts are Email: etapella@gmail.com; Skype: [etapella](#)

**Juha Uitto**

is Deputy Director of the Evaluation Office of the United Nations Development Programme (UNDP) based in New York, USA. Over the past 14 years, he has contributed to a large number of programmatic and thematic evaluations. He previously served the United Nations University, the Nordic Africa Institute and the International Fund for Agricultural Development (IFAD). He also worked as freelance consultant in international development. He was educated at the Universities of Helsinki and Lund, and holds a PhD in Social and Economic Geography. In October



2012, the European Evaluation Society awarded a prize for the paper on which the article in this issue of Connections is based. The book *Evaluating Environment in International Development*, that he edited will be published by Routledge-Earthscan in April 2014.

Claudine Voyadzis

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

**Dr. Jane Davidson**

is internationally recognized for her work on evaluation-specific methodologies that apply critical thinking and evaluative reasoning to evaluation, policy, strategy and programme design. She is author of *Evaluation Methodology Basics* (2004, Sage) and *Actionable Evaluation Basics* (2012). Jane offers evaluation coaching and professional development in person and online at <http://RealEvaluation.com>. She coauthors a popular blog with Dr. Patricia Rogers, at <http://GenuineEvaluation.com>. Jane was 2005 recipient of the American Evaluation Association's Marcia Guttentag Award and has keynoted at several conferences around the globe.

**Bastiaan de Laat (PhD)**

Bastiaan de Laat (PhD) has a long-standing experience in evaluation. Founder-director of the French subsidiary of the Technopolis Group (1998–2006) he led many evaluations for and provided policy advice to a great variety of local, national and international public bodies. He trained several hundreds of European



Commission staff and national government officials in evaluation and was involved in the design of monitoring and evaluation systems for various public organisations. More recently, he worked as Evaluator at the Council of Europe Development Bank and currently holds a post of Expert Evaluator / Team Leader at the European Investment Bank.