

EVALUATION CONNECTIONS

CONTENT

A Message from the Vice President	1
Evaluation as Moral Philosophy in Action: an Editorial	2
The Last Frontier of Evaluation: Ethics	3
How Evaluation Can Support Positive Thinking and Action	4
Aspirational Aspects of Evaluation: Hope, Theory and the Nobel Peace Prize	6
Advocacy Evaluation: Lessons from Brazil (and the Internet)	7
The Use of Social Mechanisms in Evaluation Research: Some Methodological Issues Around the “What Works” Question	9
Triangulating the Results of Qualitative Comparative Analyses	11
Towards More Effective International Financial Institutions’ Assistance to Small and Medium Enterprises: an Evaluative Perspective	13
Evaluation Briefs	14
The Authors	15

A MESSAGE FROM THE VICE PRESIDENT



Dear members and colleagues,

Happy EvalYear 2015!

On December 19 2014, for the first time ever, a standalone United Nations Resolution on national evaluation capacity development was adopted by the Second Committee of the United Nations General Assembly. Twenty six European countries co-sponsored the resolution offered by Fiji.

Marco Segone’s keynote speech at the 11th EES conference in Dublin last October, announced the start of EvalYear in Europe (<http://ees2014.eu/keynote-video-presentations.htm>). By kicking off the International Year of Evaluation in Dublin we joined a global community that shares a single vision: using evaluation to improve people’s lives through better policy making.

Evaluation events will take place in all regions of the world throughout 2015 under the umbrella of EvalPartners. All European evaluation societies will join the celebrations. EES is no exception. The International Organisation of Cooperation in Evaluation (IOCE) will provide a global forum for the initiatives to network. We expect to contribute to EvalYear in the following ways.

Future of evaluation: We are providing a platform for a web-based consultation

designed to contribute a European voice to the Global Evaluation Agenda 2015–2020. (<http://www.europeanevaluation.org/ees-blogs/riitta-oksanen/join-debate-about-global-evaluation-agenda>). We look forward to your active participation either individually or through your national voluntary organisation. We are also encouraging our Thematic Working Groups to intensify their activities and feed inputs into the global debate. This consultation process will culminate in Kathmandu at an event in Nepal hosted by the Parliament in November. EES will be represented.

Democracy and evaluation systems: EES will help organize a meeting at the European Parliament. We will bring together key EU institutions and leading evaluators to discuss the role of evaluation systems in Europe and beyond. We wish to take stock. Specifically we wish to exchange views on the role of participation, partnership and professionalism in evaluation and highlight the distinctive role of independent evaluation in democratic decision making.

Emerging evaluators: We need to build the future with the evaluators who will own it. A group of emerging evaluators has set up an EES Thematic Working Group on Emerging Evaluators after the Dublin conference. The objective is to support professionals who have

recently entered the evaluation field. In search of new solutions the next generation of evaluators is contributing ideas and it has begun to inject energy in new kinds of activities.

Professionalization: We evaluate in an increasingly complex and demanding environment. This puts increasingly high requirements on the quality of evaluations and evaluation professionals. The EES laid the foundations for piloting a Voluntary Evaluator Peer Review

(VEPR) in a seminar in 2014 and at the EES Dublin conference in 2014. We invite volunteers to get this initiative move forward.

All this and many other activities around Europe will take place in a festive mood. We should, however, wait to congratulate ourselves until we succeed in using the momentum of EvalYear to promote, expand and strengthen our discipline. I look forward to engaging with all of you to make this happen.

Let's make sure that by working together we will be able to accumulate strong evidence that we have lived up to the promise of EvalYear 2015. I very much look forward to sharing our common achievements with the rest of the global evaluation community at the Nepal closing event!

Riitta Oksanen
EES Vice President

EVALUATION AS MORAL PHILOSOPHY IN ACTION: AN EDITORIAL

Robert Picciotto

All seven articles in this issue of *Connections* take for granted that the function of evaluation is not only to inform but also to guide decision making towards the public interest. In the leading article of this edition Michael Scriven asserts that evaluators are duty bound to grapple with ethical dilemmas. Since assessment of value is what makes evaluation unique they cannot shirk the task of distinguishing right from wrong and they should do so without fear or favour: ethics is the last evaluation frontier.

Unfortunately the evaluation world today is dominated by fee dependent, utilization oriented, client controlled evaluations. In such contexts highlighting performance shortfalls requires extraordinary courage and dedication. Furthermore doing so may induce risk avoidance and inhibit positive action. Thus drawing on the psychology of individual learning Burt Perrin, Nicoletta Stame, Sanjeev Sridharan and David MacCoy argue that positive reinforcement and a bias for hope are often more effective than blame and criticism in achieving wholesome changes in behaviour.

On the other hand, Peter Dahler Larsen's subtle meditation about the Nobel Peace Prize process recognizes that social learning differs from individual learning. Public opinion matters in shaping policy and this is why the Nobel Committees do not shy from controversy. Indeed they deliberately strive to tip the balance of public debate towards progressive outcomes. Towards this end rec-

ognizing achievement is not enough. Awards are also driven by aspiration, a legitimate driver of the evaluation enterprise.

This suggests that the boundary between advocacy and evaluation is blurred. At their contested intersection advocacy organizations resort to evaluation to enhance the efficacy of their information campaigns. They recognize that the contest for "hearts and minds" takes place in complex, decentralized and diffuse networks of influence in which advocacy often faces sophisticated, dedicated opponents.

The complex methodological problems that result are probed in William Faulkner's article which illustrates with ingenuity how multi-step diagrams and sophisticated mapping software can help civil society organizations generate shrewd, adaptable and timely sequences of judiciously targeted messages designed to enhance the impact of advocacy campaigns.

In turn this special case illustrates a generic evaluation challenge which is aptly evoked by Eric Melloni, Flavia Pesce and Cristina Vasilescu in their article about the effectiveness of social mechanisms in diverse operating contexts. When is the past prologue? Why does an intervention that succeeds here fails miserably there? How can summative evaluation help identify useful formative recommendations? Shifting attention from 'did it work?' to 'why and how?' leads the three authors to propose empirical catalogues

of case studies designed to link policy outcomes with salient characteristics of social processes, mechanisms and contexts.

Rick Davies' perceptive article is similarly concerned with the dynamics of social change. It highlights the potential and limitations of Qualitative Comparative Analysis (QCA). So as to better identify the multiple configurations of conditions that can generate an outcome of interest it proposes triangulation of QCA with decision tree algorithms and ethnographic methods.

In a nutshell, diverse meta-evaluation methods hold the potential of transforming case study findings into valid evidence for policy making. Demonstrating this proposition Elsa de Moraes Sarmiento and Fredrik Korfker's article shows that aggregation of project level evaluation findings can help design development assistance policies that achieve more with less.

In sum the seven articles included in this issue display diverse concepts, approaches and methods that help evaluators judge not only the outcomes of whatever is being evaluated, but also the working mechanisms that make the social world tick. They do so for the sake of knowledge to be sure – but also to better serve the welfare of society as a whole. It is in this pragmatic sense that evaluation is driven by values. In Ludwig Wittgenstein's words "ethics is the inquiry into what is valuable.". If so, evaluation is nothing less than moral philosophy in action.

THE LAST FRONTIER OF EVALUATION: ETHICS

Michael Scriven

In order for evaluation to deal appropriately with the ethical aspect of its work, it needs to complete three steps. First, it needs to do a better job of the step it is now on, and rightly thinks of as important, namely recognizing and addressing the interests and perspectives of the impactees of an evaluation, even if they are not still around (as is the case for many historical and distanced evaluations). While such an approach and the recommendations that flow from it, have generally improved, they can still be seriously wrong even if they do not result in the intensification of exploitation. For example, some of the usual remedial formulas are overcorrections.

A case in point: it is sometimes suggested that there is a moral obligation to have evaluation impactees represented on the management team for the evaluation, possibly in at least equal numbers with the initiating evaluators. That may be a seductive idea but it is no more defensible than the idea that homeless people should be represented in a state's legislature by as many homeless people as there are non-homeless in that body – since it controls much of their welfare. An equally indefensible remedy is a variety of empowerment evaluation, i.e. giving the job of doing the evaluation to either the impactees or the staff of the organization running the evaluated program. These solutions may have useful social consequences especially for formative evaluations and in the early stages of policy design but this article is not about political processes. Rather it concentrates on long-term ethical solutions to evaluative problems.

The fundamental problem with our present situation in evaluation ethics is that all the good work that has been done on improving cross-cultural ethical methodology simply leaves us with ethical relativism, i.e. ethical nihilism. In other words, this attempt to be more ethical results in destroying the validity of ethics. Relativism entails nihilism because if all ethical codes have the right to be treat-

ed equally (i.e., relativism), no one of them is right when they disagree, so there is no such thing as the correct answer to a disputed ethical question (i.e. nihilism). Pluralism, an attempted middle ground, is either too mild to salvage a single solution or too strong to avoid nihilism.

Of course, that makes it impossible to talk of evaluation as a science since science believes in absolutism about truth: there is only one right answer to any scientific question at any point in time (though there are some problems reconciling this principle with quantum theory at the subatomic level, it's a reliable guide at the macro level). So if we cannot buy relativism because the price is too high, how do we justify absolutism, i.e. ethical right and wrong? That will be the third and most difficult problem that we address here, but it has an essential preliminary.

The second step is to move beyond the ballpark of the common ethical problems the evaluator encounters – e.g. issues about privacy and transparency and conflict of interest – and get into the struggle over the big ethical questions with which the whole society beyond evaluation is wrestling – e.g., problems about the ethicality of abortion, war, torture, suicide, same-sex marriage, police killings, state killings, prostitution, addiction. The first reason for getting into these is that, sooner or later though not commonly, duty requires it. For example, we cannot answer the question whether a particular clinic or health plan is good, or better than its predecessor, or better than obvious alternatives, or worth what it cost, etc., unless we can say whether it does what it should be doing; and we cannot decide on that unless we can say whether it should be providing or avoiding the provision of abortion and birth control services. At the moment, we tend to give only conditional answers to such questions, e.g., by passing the buck back to the client, or to the community: in other words, avoiding direct answers to such questions. There are situations where this stance at can

be justified, but there are also many, e.g., doing evaluation in the public health or policy analysis fields, where doing that is simply dodging part of the question evaluators are called on to answer, i.e., failing to do your professional duty. If evaluators cannot lead in developing defensible standards of human behavior, they become like many psychotherapists of the mid-20th century who dealt with homosexuality as a curable sin or disease – being either biased or ignorant.

Getting into these battles is a 'big ask' for evaluators, who have often been brought up to think that normative value issues are not part of science, either because they cannot be dealt with using scientific method, or because they are part of religion's turf. Those defenses are evaporating under closer scrutiny than they were given earlier, i.e. we will have to deal with a broader and more combative playing field on the ethical side of evaluation, just as we have had to do in the case of ethical issues in society such as the legalization of some recreational drugs besides alcohol, nicotine, and coffee, and of some sexual or marital preferences other than those of early European missionaries. The playing field has changed, and we have to change our game to cope with the new reality, if we want evaluation to be a science or indeed a discipline of any kind.

But that is only two steps, and the third one is the hardest. We are all used to the idea that there are sub-fields of evaluation, including the specialties of program evaluation, personnel evaluation, product evaluation, etc. We've been a little slow to pick up on the fact that one of these sub-divisions, perhaps the most important of them all, is ethics. It is no more than – and no less than – the most far-reaching and fundamental type of evaluation of human behavior, attitudes, and thinking. As an important branch of Western philosophy since Socrates' time, more than 2000 years ago – and a topic for sages and leaders long before that, it has a huge literature, but still very extensive

disagreement about both its theory (meta-ethics) and its application (normative ethics). It is indeed a discipline in its own right, like property appraisal or judicial review or literary criticism, but no less a sub-discipline of evaluation, which is a multi-disciplinary endeavor. We must try to contribute to this field, at least by sketching working solutions to the task of laying sound foundations for it, since without that, all the ethical conclusions we develop in steps one and two above, will

be unfounded. I make some suggestions as to how this might be done in the full version of this paper that will appear in a book edited by Stewart Donaldson and Robert Picciotto.

I suggest that the key to a new approach is first to ruthlessly minimize the axioms needed to justify ethical conclusions. I cut it to one here, the *prima facie* right of all citizens to be considered by all whose decisions affect them. Second, to justify that axiom. Third,

to uncover and verify or reject the empirical content and consequences of specific applications – for example, to show that parents who fail to vaccinate their children while still sending them to school or play with others will kill or cripple some innocent children of other parents, as well as some of their own. Whether by this route or another, we must solve this great problem, or else we will have built evaluation on sand. That works for a while; then the rains will come.

HOW EVALUATION CAN SUPPORT POSITIVE THINKING AND ACTION

Burt Perrin, Nicoletta Stame, Sanjeev Sridharan, and David MacCoy

In this brief article, we highlight the importance of an approach to evaluation that supports positive thinking and action, discussing various ways in which this can be done. As Perrin (2014) points out, a positive approach to evaluation is basic to the objectives and *raison d'être* of evaluation, which is to support use of evaluation and to contribute, generally indirectly, to social betterment. This recognizes that evaluation is only of value if it is used in some way. Consistent with the literature on *utilization-focused evaluation*, this requires the engagement and motivation of stakeholders to act upon evaluation. The approach to evaluation that we advocate contrasts with how subjects of evaluation often view it, rightly or wrongly: as overly focused on fault finding and on the inevitable problems, glossing over what is working well and what has been achieved.

A positive approach to evaluation, focusing on strengths rather than on deficits, draws from learnings in psychology and related fields. A key principle from the psychology of learning is that *positive reinforcement* is more effective in achieving learning – and changes in behaviour – than punishment. But the latter is how negative evaluation reports are typically received, which is more likely to create resistance rather than to facilitate doing things differently.

Another key finding from psychology and the organizational leadership area is that *intrinsic*

motivation, involving internalization of core values, is necessary for commitment and a desire to make changes, ownership and follow-through. This is strongest when people feel that they are acting upon ideas and plans that they have come up with themselves rather than imposed by others, such as from an evaluator. A positive or constructive approach throughout the evaluation process, rather than externally imposed rewards or sanctions (extrinsic motivation), is most likely to result in meaningful and sustainable actions.

Stame (2014) discusses how traditional approaches to evaluation of programmes based on a “direct help” perspective, involving externally-identified solutions to address deficits, tend to yield the depressing finding that “nothing works.” Programmes based on an “indirect help” perspective, supporting people’s active involvement, require evaluation approaches that are able to value innovation and change, such as *positive thinking approaches (PTA)*.

Positive thinking approaches to evaluation are based on social theories that emphasize people’s active involvement and the thrust toward betterment that can result from acknowledging past successes. Stame discusses a range of alternative PTA to evaluation including: Appreciative Inquiry (Preskill and Catsambas), Success Case Method (Brinkerhoff), Most Significant Change (Dart and Davies), Positive Deviance (Sternin), Evalua-

tion of Innovation (Perrin), and Developmental Evaluation (Patton).

The main tenets of all PTA are “overcoming the dependency syndrome” and “there is always something that works”. A major strength is the ability to account for unexpected positive consequences, emerging outcomes, and all that is generated by people’s empowerment. However, the approaches vary regarding methodologies for discovering successful cases, eliciting people’s motivations, mobilizing latent energies, how “failure” is treated, and innovating on the basis of past success.

While some PTA may be most suited to evaluating cases of indirect help programmes, there are many situations where they also can be useful for evaluation of direct help programmes, such as in providing a means of surfacing values, challenging as need be negative assumptions that frequently underpin prescriptive, top down externally imposed interventions, and providing a means of identifying locally based solutions that often are masked in traditional evaluations.

As Sridharan and Warren (2014) discuss, the *realist evaluation approach* shifts focus from “does a programme work?” to “what is it about a programme that makes it work?” A fundamental aspect of conducting realist evaluation is to explore contexts, mechanisms and outcomes underlying pro-

grammes. Realist evaluation approaches are especially relevant for evaluating complex, multi-component initiatives.

One of the implications of a realist view to complex programmes is a recognition that program implementers need help to align the complex programming with the long term goals (such as redressing health inequities). The realist approach attempts to understand why programs work. “Realists do not conceive that programmes ‘work,’ rather it is action of stakeholders that makes them work, and the causal potential of an initiative takes the form of providing reasons and resources to enable programme participants to change” (Pawson & Tilley in Sridharan and Warren).

Using an example of a National Demonstration programme in Scotland, Sridharan and Warren demonstrate how realist evaluation has helped in the following aspects of policy implementation: exploring ‘loss in translation’ from policy aspirations to program design; interrogating the program design; developing a range of learnings from conducting the evaluation; and aligning the learning with policy priorities. This example also identifies conditions under which evaluations can lead to ‘positive’ influence on policy makers and on future policies.

The value of participatory, stakeholder, and learning-oriented approaches has been an important theme in evaluation for many years. *Appreciative Inquiry (AI)* is not evaluation per se, though it offers a philosophy and

method for conducting a full evaluation or various phases of an evaluation.

As MacCoy (2014) discusses, AI is the cooperative search for the best in people, their organizations, and the world around them. Evaluators using AI have found that its *appreciative questions, reframing,* and *generative* features set the stage for sound assessment of worth as well as potential for powerful solutions. AI questioning moves from determining what is valued and appreciated to combining strengths and activating people’s creative energy to ignite change.

The AI approach has been applied in a wide range of evaluation initiatives in the private, not-for-profit and public sectors. It has been used effectively in evaluations of social, health and education programs, international development, the arts, and human resource programs within many diverse organizations. AI has been used to focus an evaluation, conduct appreciative interviews, develop evaluation systems and build evaluation capacity.

The choice to use AI has often resulted from limitations encountered in using problem or deficit focused approaches. The problem-focused approach may be effective in solving existing problems and “fixing” them, but often is less effective in identifying what is going “right” and applying this for positive change. Reframing is frequently used in AI to counter deficit thinking (e.g. focus on failure) and shift to a solution or asset-focused perspective (e.g. possibilities not yet considered).

Positive (or “strength-based”) approaches to evaluation, despite some misconceptions, do not focus just “on the good stuff” or avoid problems. What they do, however, is to constructively reframe inevitable shortcomings on future directions, acting as a generative process to facilitate the identification of creative ideas. As such, they represent solution-oriented approaches that can take into account local context and culture and are consistent with requirements for accountability and the obligation of evaluators to tell the truth.

References

- MacCoy, David. (2014). Appreciative Inquiry and Evaluation – Getting to What Works. *Canadian Journal of Program Evaluation*. 29 (2, fall), 104–127.
- Perrin, Burt. (2014). Think Positively! And Make a Difference Through Evaluation. *Canadian Journal of Program Evaluation*. 29 (2, fall), 48–66.
- Sridharan, Sanjeev, and Warren, Tim. (2014). The Utility of a Realist Evaluation Approach in Implementing and Evaluating Health Equity Policy. *Canadian Journal of Program Evaluation*. 29 (2, fall), 87–103.
- Stame, Nicoletta. (2014). Positive Thinking Approaches to Evaluation and Program Perspectives. *Canadian Journal of Program Evaluation*. 29 (2, fall), 67–86.

ASPIRATIONAL ASPECTS OF EVALUATION: HOPE, THEORY AND THE NOBEL PEACE PRIZE

Peter Dahler-Larsen

All evaluation takes place at a moment in time. Evaluation looks at past achievements and aspires to help shape the future. However, the nature of the link between achievement and aspiration is far from evident. The balance between these two elements varies from one form of evaluation to another and so may the exact way in which they connect. For example, a more or less explicit theory of change may help an evaluator make choices about criteria and methods in the light of how he/she wants to call forward particular reactions that help shape the future but the evaluator has very little control over these reactions. Counter-reactions occur too.

I am sharing my thoughts on this problem after reflecting on the official motivation for the 2014 Peace Prize to Kailash Satyarthi and Malala Yousafzay for their struggle against the suppression of children and young people and for the right of all children to education. I was curious about the Prize because I had just interviewed Geir Lundestad, General Secretary of the Nobel Peace Prize committee, whose generous participation made it possible for me to write an article about the Nobel prize as an evaluative institution (see reference below). The Nobel institution is an interesting case, not only because nominations, prizes and awards play an increasing role in society, but also because the underlying evaluative decisions and practices of the Nobel committee are complicated, fascinating and largely secret.

Geir Lundestad's fundamental view on the aspirational aspect of evaluation is this: *"No situation is ever fully resolved. History moves on. It moves on all the time. So the aspirational aspect is a least implicitly frequently there: you hope that something more will follow."*

The term *hope* is revealing. So is the confidentiality of the deliberative process and the *implicit* and *covert* nature of any particular

theory of change that may have guided the Nobel Committee's deliberations. I also observed that the Committee resorts to supplementary practices that are both interpretive and normative and that help shape and direct the potential consequences of the evaluative decision. Communication is key.

The General Secretary was very articulate about the kind of peace that the Nobel Prize promotes. While some totalitarian regimes may be able to produce peace, Geir Lundestad explained that *"you cannot really expect peace in the longer run, if you do not have a basis in human rights, in democracy... There will still be conflicts. We are not talking about absolute causal connections"*.

A strong normative component along with an aspiration component underlie the Committee's thinking regarding international peace. This helps explain why the Nobel Peace Prize is often awarded to controversial figures embroiled in conflict. Peace without further qualification is neither a desired, nor a direct or expected consequence of a Nobel Peace Prize. Controversy regarding the award cannot be avoided. Indeed it is often intended.

Publicly the Nobel committee goes at great lengths to explain the links between contemporary security issues and Nobel's original conception and vision of peace. Nobel's will was written in 1895, a hundred and twenty years ago. It referred to "peace conferences" and "fraternity between people". It could not predict the many contextual issues that influence peace in today's complex world. This opens the door wide to thoughtful interpretation. The view of the Nobel Committee is that such issues as climate change and the rights of minorities constitute important preconditions for a peace worth having. Causal and normative arguments can be adduced to support this view.

Official documents such as the press release that announces the Peace Prize not only celebrate the virtues of the person who has received the award but they also help direct the world's attention to the issue or the struggle that the person embodies. For example the press release that justifies the 2014 award to Kailash Satyarthi and Malala Yousafzay states that *"the struggle against suppression and for the rights of children and adolescents contributes to the realization of the "fraternity between nations" that Alfred Nobel mentions in his will as one of the criteria for the Nobel Peace Prize."*

This passage helps to defend the Prize against critics who argue that the Nobel Committee ought to steer away from divisive international politics and stick to a more conventional interpretation of peace. It explains why the focus on children's rights is fully justified on peace building grounds. The very allocation of the Prize to individuals that are the public face of a salient policy debate is an integral component of a communication strategy that puts the spotlight not only on those persons, but also on the cause and contest in which they are engaged.

The press release even mentions that *"The Nobel Committee regards it as an important point for a Hindu and a Muslim, an Indian and a Pakistani, to join in a common struggle for education and against extremism"*. Here, the Nobel Committee goes well beyond the recognition of documented achievements. Instead, it aims to promote cooperation among the religious and national categories that the laureates are said to represent. Here, hope and aspiration were evidently *added* to the evaluative criteria used when the award decision was made.

Yet, Geir Lundestad is careful to explain that *"there has to be a basis of achievement. You cannot just throw a prize and hope something*

will happen.” He goes on to give examples from the history of Nobel Peace prizes, where some (e.g. Mandela and de Klerk) had “almost crossed the finishing line”, whereas for others (Arafat, Peres, and Rabin) “much remained...”

The lesson of the Nobel Peace Prize process may be that a fulsome engagement with both the element of achievement and the element of aspiration is an interesting challenge in evaluation. In turn this would suggest that evaluation is more an art and a practice than a science.

Reference

P. Dahler-Larsen: Dynamite, Medicine, Peace, and the Nobel Art of Evaluation. *American Journal of Evaluation*, 35 (3), 2014: 377–386. ■

ADVOCACY EVALUATION: LESSONS FROM BRAZIL (AND THE INTERNET)

William N. Faulkner

Advocacy evaluation is still “stammer[ing], making its debut, creating itself...in direct confrontation with the world.” (Latour, 2004) The organizations doing advocacy are usually decentralized and non-hierarchical. The patterns of change pierce smooth regression lines, jumping around choppily. Also, many advocacy efforts have sophisticated, dedicated opponents; information is deliberately compartmentalized rather than disseminated. All of these aspects mean that advocacy tends to naturally repel the battery of methods on which evaluators typically rely.

While designing and implementing two projects surrounding international advocacy campaigns, PlanPP developed tools to cope with these peculiarities, specifically: a method visualizing data on connections between effective strategies and outcomes, and a way to organize qualitative data with mind mapping software. This brief article outlines what we learned from these projects in the hope that subjecting these techniques to public view may benefit others attempting similar ventures and also to articulate the more abstract claim that evaluation tools and technologies should be regarded as more than passive instruments. Evaluators need to respect their tools, right down to the particular software packages chosen, as active ingredients which shape and underpin their ideas – even about evaluation theory.

Advocacy evaluation in theory

Advocacy evaluation is an “elusive craft.” (Teles and Schmitt, 2011) Advocacy networks tend to eschew centralized, hierarchical structures in favor of diffuse webs. To the

extent feasible, advocacy evaluation results should be disaggregated by strategy, time-period, actor, and other pertinent dimensions so that stakeholders within the network can extract the information most relevant to their particular interests (Weiss, 2007).

Second, the pace of advocacy often follows a ‘punctuated equilibrium,’ whereby smooth change is replaced by long periods of stability followed by brief spurts of intense activity. Precise timing is therefore crucial; information may only be relevant for a few weeks, days, or even hours. (Coffman and Beer, 2011, p. 7).

The disclosure of results of an independent advocacy evaluation is a sensitive affair. Many advocacy efforts face formidable opponents. Disseminating results must consider the possibility of harming the evaluand by revealing strengths and weaknesses to the opposition. Sou da Paz, for instance, requested that a previous version of this paper prepared for the *Rede Brasileira de Monitoramento e Avaliação* (Brazilian Monitoring and Evaluation Network, RBMA) use a pseudonym for the organization and avoid references to specific organizations or groups.

Advocacy evaluation in practice

Under the assumption that there is no such thing as a universally ‘effective’ advocacy strategy, the PlanPP team sought a way of displaying information which would exhibit which combinations of strategies were most effective at contributing to which combinations of outcomes. Several questions on the survey questionnaire therefore aimed to elicit responses amenable to visualiza-

Insituto Sou da Paz (‘I am for peace’ Institute)

An NGO, Sou da Paz is a household name in their home base, São Paulo, due to their violence prevention programs in the city periphery as well as their protagonism in a national disarmament campaign. Several staff members have participated in advocacy networks surrounding UN processes regulating the trade, storage, and use of conventional weaponry since the 1990s.

WITNESS

WITNESS, an NGO based in Brooklyn, USA, developed and undertook the Global Forced Evictions Campaign between June 2011 and May 2014. The primary objective was to ‘protect the rights of poor and underrepresented communities to housing, livelihood and community from forced evictions by development.’

tion using Sankey diagrams, a special type of flow diagram which overlays stacked bar graphs (the columns) with connecting bars, the widths of which correspond to the percentage of respondents who paired the two connected options (see figure below). The production of these diagrams is a multi-step, multi-software process requiring significant computer literacy but only one paid software package which most evaluators already have: Microsoft Excel. The procedure is essentially an exercise in database preparation which transforms survey answers into a format

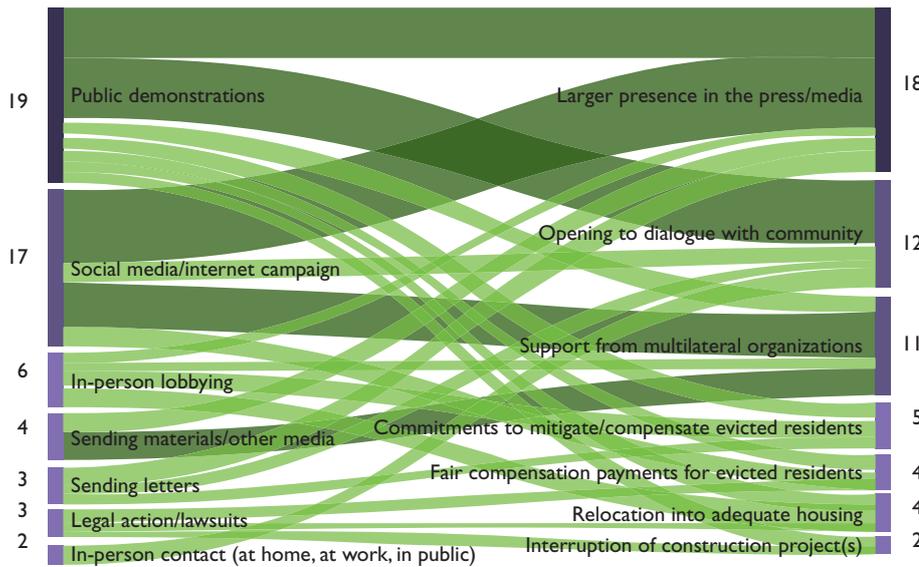


Figure 1: Example SankeyDiagram from WITNESS project.

which can be processed by an Excel template called 'cDataSet.'

Rather than serving as stand-alone results, these figures (a) guided probing questions during interviews, and (b) served as data sources to be triangulated to highlight areas of agreement and disagreement between different data sources.

Mind mapping for qualitative analysis

Another challenge the PlanPP team faced was synthesizing qualitative information from 30+ (Sou da Paz) and 23 (WITNESS) interviews into useful analyses for the clients while allowing novel, unexpected categories and results to emerge. PersonalBrain, a mind-mapping software, includes several features which allowed the team to partially reconcile 'top-down' and 'bottom-up' categorization. The following advantages do come at a price of a labor-intensive coding process, however:

- PersonalBrain excels at handling and visualizing infinite-level nested hierarchies which

can mimic the function of coding a single piece of text to multiple themes.

- The interface for PersonalBrain is fairly tactile, generally based on drag-and-drop commands in contrast with menu- or code-based packages.
- Although not free (US\$219 for the more capable Pro version), it is still much cheaper than licenses for many qualitative analysis packages.

Broader abstractions

Much has been written in the field of evaluation lamenting the haphazard and premature application of novel technological tools without proper grounding in theory – social, mathematical, or otherwise – but few devote much effort considering the other side of the coin. Novel theories and fields of evaluation often require new or repurposed tools to reach their full potential, or sometimes to be applied at all.

The lack of prior experience in the area forced the PlanPP team to relinquish in-

grained notions about the intended purpose of the tools at their disposal. Instead, the team focused on the tools' components and basic functions and then tried to reconcile the potential configurations of these functions with what advocacy evaluation theory requires. When tools lacked small but essential features, the generosity of the web community usually supplied them for free. Still, because few of these add-ons, plug-ins, and programs had been applied in evaluation, discovering them required figuring out the key search terms (e.g. "Co-occurrence Matrix" and "Array Formulas").

As an emerging field (particularly outside the Global North) advocacy evaluation requires creative solutions to the unique challenges which it presents in both theoretical and practical domains. Imagining a matrix of methods and tools applicable to advocacy evaluation crossed with all the possible contexts in which advocacy evaluation might occur, only a tiny percentage of the cells would currently be populated with concrete case studies for reference. In this relatively unexplored landscape, both the possibility of discovery and the risks of disaster grow. The PlanPP team hopes that sharing the spaces explored during two advocacy evaluation projects realized under conditions propitious for innovation may benefit others attempting similar undertakings.

References

Coffman, J., & Beer, T. (2011). Evaluations to Support Strategic Learning: principles and practices, (June).

Latour, B. (1999). Pandora's Hope: Essays on the Reality of Science Studies. Cambridge, MA: Harvard University Press.

Teles, S., & Schmitt, M. (2011). The Elusive Craft of Evaluating Advocacy. Stanford Social Innovation Review, Summer, 38–43.

Weiss, H. (2007). Advocacy and Policy Change. The Evaluation Exchange.

1 Available at the Excel Liberation website: <http://ramblings.mcpher.com/Home/excelquirks/downloadable-items>.

THE USE OF SOCIAL MECHANISMS IN EVALUATION RESEARCH: SOME METHODOLOGICAL ISSUES AROUND THE “WHAT WORKS” QUESTION

Erica Melloni, Flavia Pesce, Cristina Vasilescu

The success of public policies is often the result of a combination of good institutional endowment and adequate policy capacity. However, in order to provide satisfactory explanations of the extent of policy achievements and so as to contribute to policy learning, the analysis should also take into account contextual considerations, the nature of policy actors (types, goals, values and beliefs) and the way they interact to achieve results. This complexity is hidden in policy “black boxes” and requires evaluators to use new methods to understand what lies behind policy success or failure and contribute to better policy designs.

While neo-institutionalism is overconfident in the explanatory power of institutional incentives and structures, realistic evaluation (Pawson and Tilley, 2007) acknowledges agency as a crucial element for unveiling the success of a policy. However, a key question around the concept of policy results is not only whether goals have been reached, and to what degree; but also, whether the policy which stimulated wholesome results can be used as an exemplar (be it a model, or a source of inspiration, or something in between).

The point of evaluation is to understand what works, for whom and in which circumstances. The mechanisms approach, largely spread across many disciplines and diffused throughout theory-based evaluations, aims to trace the linkages among the complex bundles that make up a policy and its context, with the goal of providing insight about what works in policy making thus enhancing the likelihood of achieving better results. (Barzelay 2007)

Both Barzelay (2007) and Bardach (2004) stress the need to learn about smart solutions discovered and implemented by other policy makers. They also suggest working with social mechanisms, a concept which is certainly not new and is indeed widespread across scientific disciplines, among which social research and evaluation¹.

Evaluation research has been heavily influenced by the “mechanism” concept, starting with the seminal contributions of Pawson and Tilley (1997). Theory-based evaluation approaches as well as contribution analysis (Mayne 2012) also use the concept of mechanisms. According to Stern et al. (2012), theories based on mechanisms constitute a specific class among four main approaches to the understanding of causation, one of which is the “Generative” Causation framework (the other three being: the Regularity framework, the Counterfactual framework and the Multiple Causation framework). Compared to these, the Generative Causation framework allows in-depth understanding and fine-grained explanations of complex and context-related causal chains but it also suffers from potential weaknesses in terms of external validity and generalization of results.

The core assumption of all these approaches is that “*causation without explanation is insufficient for policy learning because policy makers need to understand why as well as how they are to use findings from research or evaluation for future policy making*” (Stern et al, 2012). This shifts the focus of evaluation from the question ‘did it work?’ to ‘why and how?’ giving rise too many methodological and practical

issues that scholars and practitioners must confront.

To answer the formative ‘why and how?’ questions we applied the social mechanism approach through the use of the “extrapolative case study in a number of evaluations. This peculiar learning-oriented approach was proposed by Michael Barzelay in 2007. We adopted it in pursuit of convincing explanations regarding policy implementation success and failure. The corollary – more ambitious – goal was to compile an empirical catalogue of mechanisms to explore alternative ways of functioning in real situations with a view to draw policy design and implementation lessons².

While using this method we focused on policy actors’ roles and behaviours. Specifically we privileged four main elements in the extrapolative case study approach:

1. *Policy outcomes* (with particular reference to behavioural changes *aligned* and *coherent* with the policy goal). This interpretation broadens Barzelay’s reference to “outstanding results”. It incorporates any outcome of interest provided it points to durable changes in the behaviour of one or more actors involved in the policy process;
2. *Policy processes* connected to major and relevant changes in behaviour. Three main types of processes may be involved (Busetti, Melloni, Dente 2013): (i) the process of generating and maintaining engagement of relevant partners/stakeholders/beneficiar-

1 In social sciences, mechanisms typically belong to the sociological vocabulary (Merton 1957, Hedström and Swedberg, 1996; Elster, 1998; Tilly 2001) but also fuelled the behavioural economy, the organizational studies (Bardach 2004; Barzelay 2007), the European studies (Eberlein and Radaelli 2010), and the policy studies (Scharpf 1997, Mayntz 2004, Dente 2011).

2 It is worth noting that the method of unpacking the “best practice” into “relevant processes” would allow expanding the interest beyond the specific policy domain in which the best practice refers to.

ies; (ii) the process of enhancing/maintaining/reducing the role of certain partners; (iii) the coordination of participants' activities within the policy network (i.e. focused on actors who hold some responsibility in the implementing process).

3. *Mechanisms* triggering specific policy outcomes as described by various scholars: attribution of opportunity or threat (Tilly 2001); certification of the role and reputation and blame avoidance (Edelman 1977; Hood 2011); performance feedback (Cyert and March 1992; Bandura 1977; Barzelay 2007); anticipation of preferences (Scharpf 1997); "Bandwagon effect" (Granovetter 1978), etc.
4. *Context features* (institutions; policy networks; rules; ...). As realist evaluators (Pawson 2006) point out, discovering the explanatory mechanisms represent only half of the issue. The relation between social mechanisms at work and outcomes is not fixed: it too depends also on the contextual features. Programme functioning is influenced by contextual constraints, interrelations, institutions and the structures in which the programme is embedded. The other part of the problem consists, thus, in providing a picture of how the context operates and how it hampers or fosters actors' behaviours.
5. *Linkages* among the above-mentioned elements through narrative reconstructions of how outcomes are related to policy features and policy context and

why (through which mechanisms) specific outcomes were produced.

The benefits of this approach are twofold:

1. *Ex post/learning*: the "mechanism" approach focuses on actors and policy processes which allows us to see beyond the formal design of a policy, and to deepen understanding on the dynamics of change and adjustment undergone in the course of time.
2. *Ex ante/design*: The "mechanism" approach raises questions about how to foster behaviours consistent with the policy goals including how capacity-building interventions should be shaped.

References

- Bardach, E.: Presidential address – The extrapolation problem: How can we learn from the experience of others?, *Journal of Policy Analysis and Management* 23 (2): 205–220, 2004.
- Barzelay, M.: Learning from Second-Hand Experience: Methodology for Extrapolation-Oriented Case Research. *Governance* 20 (3): 521–543, 2007.
- Busetti, S., Dente B., Melloni E.: A Theory of Practice for Collaborative Governance: The Use of Social Mechanisms in Understanding Policy Success, Paper presented to the 7th ECPR General Conference, Sciences Po Bordeaux, 4–7 September 2013.

Dente, B.: Understanding policy decisions, Springer Briefs, 2014.

Mayntz, R.: (2004). "Mechanisms in the Analysis of Social Macro-Phenomena." *Philosophy of the Social Sciences* 34 (2): 237–259, 2004.

Pawson, R.: Evidence Based Policy: A Realist Perspective, Sage, 2006.

Radaelli, C.M., Dente B., Dossi S.: Recasting Institutionalism: Institutional Analysis and Public Policy, *European Political Science* 11, 537–550, 2012.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., Befani, B.: DFID Working Paper 38. Broadening the range of designs and methods for impact evaluations. DFID, London, UK, 2012.

Reports

DIAP Politecnico di Milano, Smart institutions for territorial development (Smart-Ist), 2013, http://www.espon.eu/export/sites/default/Documents/Projects/TargetedAnalyses/SMART-IST/SMART-IST_FINAL_REPORT_09.12.pdf

Istituto per la Ricerca Sociale, Gender mainstreaming in Committees and Delegations of the European Parliament, 2014, [http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/493051/IPOL-FEMM_ET\(2014\)493051_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/493051/IPOL-FEMM_ET(2014)493051_EN.pdf)

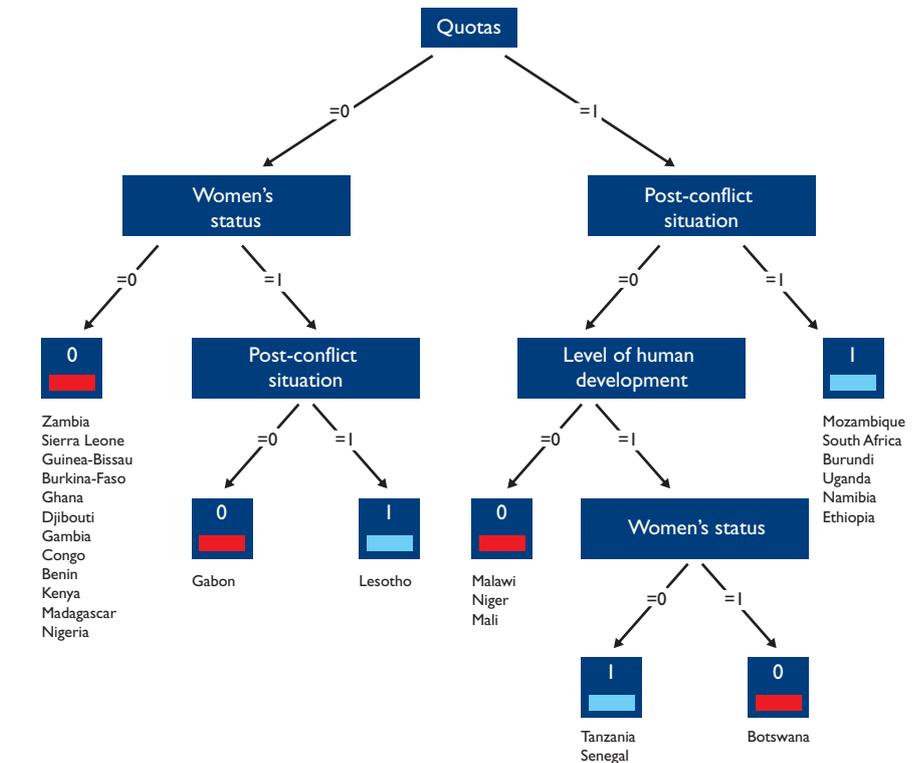
TRIANGULATING THE RESULTS OF QUALITATIVE COMPARATIVE ANALYSES

Rick Davies

Qualitative Comparative Analysis (QCA) has been the subject of increased attention in recent years especially from evaluators interested in alternative ways of identifying causal contributions (Stern et al, 2012). The use of QCA is worth exploring for a number of reasons. Foremost is its capacity to represent causal influence in a way that respects complexity. (Befani, et al 2007; Befani, 2013). Another appealing feature is the potential for QCA analyses to be tested and refined via replication: lack of replicability has long been a major concern in social research (Ioannidis, 2004).

Two aspects of QCA facilitate replication studies. First is the availability of data sets which are the basis of QCA analyses. QCA typically uses relatively few cases. These are often included in published papers, or in the truth tables derived for them. For example, Compasss, an online repository of QCA papers, indicates which published papers include data sets and provides online access to them¹. Second is the systematic nature of the QCA analytic process (Rihoux and Ragin, 2009; Schneider and Wagemann, 2012.) as highlighted in “good practice” reviews (Mello, 2012; Wagemann and Schneider, 2007).

There are however some downsides to the QCA process. For the lay person its complexity can be daunting. Some writers have sought to present QCA findings in a more user-friendly way) than the basic Boolean algebra statements (e.g. Schneider and Grofman, 2006). This makes them easier to read and appreciate. Other critics have questioned the value of the main QCA algorithm (Quine-McCluskey optimization) and assumptions about unrepresented configurations (Baumgartner, 2014). But broadly speaking, the basic approach to causality is



Data from Krook, M. L. 2010. “Women’s Representation in Parliament: A Qualitative Comparative Analysis.” *Political Studies* 58 (5): 886–908. Analysed using the Decision Tree operator in the RapidMiner software package.

Figure 1: A tree structure representing seven different but overlapping causal configurations.

widely valued. This has been described as “multiple conjunctural causation” (Rihoux and Ragin, 2009), i.e. an outcome of interest can be generated by multiple configurations of conditions.

My purpose in this brief article is to introduce alternatives that can be used to complement and triangulate QCA findings, which can test and enrich understanding of causal process-

es. Two alternatives have notable strengths. One is the use of Decision Tree algorithms as used in open source data mining packages such as RapidMiner². These are used in fields that are rich with data but poor in theory, e.g. Big Data sets generated by consumer purchases or peoples’ online behaviour. The other is the use of ethnographic methods, which privilege the knowledge of particular actors, and elicit their often tacit knowledge

1 <http://www.compasss.org/bibliography/allApp.htm>
 2 <https://rapidminer.com/>
 3 <http://mande.co.uk/special-issues/hierarchical-card-sorting-hcs/>

through structured comparison exercises. A particular application of interest involves a use of hierarchical card sorting³.

These two alternatives have one feature in common: the ability produce results which can represent causal configurations in the form of tree structures, which are relatively easy to read and understand. Figure 1 above is an example. It is derived from a QCA data set, using Rapid Miner. Each branch represents a different configuration of causal conditions. At the end of each branch are the cases associated with that configuration of conditions.

Specifically, red and blue “leaves” represent groups of countries where the level of women’s participation in parliament is low and high respectively. Branches may be made up of conditions that are present or absent, as signified by the =0 or =1 annotations to each branch. Conditions are attributes of the countries in the data set.

Given that there are different ways of analysing QCA type data sets the question arises as to how to compare and evaluate the causal models produced by the different methods. This can be done at three levels of scale: (a) the whole model, (b) specific configurations (i.e. branches) within the model and (c) specific conditions used in those configurations.

The three criteria can be applied at one or more levels. Simplicity can be seen in the number of configurations used in each model and the number of conditions in those configurations. Simplicity is not merely an aesthetic consideration. At a practical level a model with fewer or simpler configurations would be easier to implement and test through a specific kind of intervention.

Consistency is the proportion of outcomes correctly identified as such (e.g. as being absent or present). It can be measured at the level of individual configurations and for the model as

a whole. Both measures can also be tested in terms of their external validity where extra data is available. For example, data on other African countries outside the Kook dataset.

Coverage is the proportion of all outcomes of a given type (e.g. present or absent) that are identified by a given configuration. Individual conditions may also vary in their coverage, by appearing in one or many configurations. Conditions which appear in all configurations leading to one type of outcome (or the other) can be said to be necessary but not sufficient, and in that respect they have merit in terms of a simplicity criteria. More often conditions are likely to be INUS: Insufficient but Necessary parts of a configuration that is Unnecessary but Sufficient. For example, “women’s status” in Figure 1.

No matter what method is used it is unlikely that a single, unambiguous, best fitting solution will always be found. This is partly because the use of both methods involves the setting of specific parameters, which lead to different consequences.

More particularly, the three performance criteria mentioned above are often in competition with one another. Choices need to be made in particular settings as to which are the most important. Configurations with wider coverage of cases may be more practically useful than those with less, but at some sacrifice to consistency (i.e. accuracy with which they predict associated outcomes). In some settings (e.g. medical) consistency may be the most important consideration so that coverage needs to be sacrificed.

Other useful performance criteria have been suggested by a recent publication on the origins of innovation in biological systems (Wagner, 2014). This argues that a combinatorial perspective similar to that built into both QCA and Decision Tree models has much wider applicability (e.g. in models of genomes and metabolic circuits). One important feature of

such combinatorial models is their “robustness”, i.e. their ability to effectively perform under a variety of conditions⁴. This attribute of QCA and Decision Tree models should be measurable, by attending to the range of other conditions found alongside a given configuration of interest. Doing so may help evaluators identify the wider applicability and adaptability of particular causal models.

References

- Baumgartner, M., 2014. Parsimony and Causality. *Qual Quant* 1–18. doi:10.1007/s11135-014-0026-7.
- Ioannidis, J.P.A., 2005. Why Most Published Research Findings Are False. *PLoS Med* 2, e124. doi:10.1371/journal.pmed.0020124.
- Mello, P., 2012. A Critical Review of Applications in QCA and Fuzzy-Set Analysis and a “Toolbox” of Proven Solutions to Frequently Encountered Problems.
- Rihoux, B., Ragin, C.C. (Eds.), 2009. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Sage, Thousand Oaks, Calif. [u.a.].
- Schneider, C.Q., Grofman, B., 2006. It Might Look like a Regression Equation... But it’s Not! An Intuitive Approach to the Representation of QCA and FS/QCA Results.
- Schneider, C.Q., Wagemann, C., 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge University Press.
- Wagemann, C., Schneider, C.Q., 2007. *Standards of Good Practice in Qualitative Comparative analysis (qca) and Fuzzy-Sets*.
- Wagner, A., 2014. *Arrival of the Fittest: Solving Evolution’s Greatest Puzzle*. Penguin Group US.

4 <http://www.investopedia.com/terms/r/robust.asp>

TOWARDS MORE EFFECTIVE INTERNATIONAL FINANCIAL INSTITUTIONS' ASSISTANCE TO SMALL AND MEDIUM ENTERPRISES: AN EVALUATIVE PERSPECTIVE

Elsa de Morais Sarmiento, Fredrik Korfker

The *raison d'être* of International Financial Institutions' (IFIs) assistance to Small and Medium Enterprises (SME) transcends the well-known economic rationale for public intervention in markets. Strengthening the role of IFIs in the provision of productive finance has to be tackled through a blend of direct lending (e.g. strengthening public sector capabilities, financing large infrastructure projects) with indirect lending (deepening the levels of financial inclusion for SMEs).

Well beyond the alleviation of capital market failures that affect small, fragile and financially constrained firms, the rationale of IFIs' SME interventions lies in the amplification of positive externalities generated by investments in SMEs active in public infrastructure, social sectors and export markets, and the mitigation of negative externalities (e.g. environmental and social effects of investments, financial crisis risks) through safeguard systems.

Thus IFIs resources can play a unique role in providing finance for activities that are not able to attract enough private capital. Furthermore tackling the "missing-middle" phenomenon can help to trigger total factor productivity growth¹. Linkages and spillovers between private sector support and inclusive growth² through targeted support to private firms can bring about significant developmental impacts, leading to higher employment levels (and job creation) and an acceleration of poverty reduction.

Most IFIs are project-based banks, providing finance to SMEs indirectly via financial intermediaries. Stimulating small business lending through effective and cost-efficient cooperation with local banks can make a noteworthy contribution to improving access to finance

by providing the former with capital and technical assistance. But this form of "subsidiarity" implies a lack of direct control over the targeting of ultimate beneficiaries (SMEs). Since local Banks and Fund Managers tend to play safe and keep transaction costs low, they often deploy IFIs resources towards larger SMEs, thus depriving capital to firms that are most in need.

Equally, IFIs need to make sure they apply strict lending standards. Securing Board seats in these institutions can help IFIs strike the right balance when combining financial prudence and judicious targeting. Improved targeting by IFIs is also a challenge due to the lack of standard SME definitions across countries and regions.

Strengthening relevance can be done through better targeting, aiming at precise firm characteristics, such as high-growth firms, venture capital firms, those more prone to contribute to youth employment, or those engaged in social businesses. This requires a full mapping of the business demography in the region and a clear definition of these concepts. It also needs to be coupled with the right type of selection criteria and incentives so that financial intermediaries actually hit the target group.

Another common problem concerns the difficulty in identifying results at the level of the ultimate beneficiaries. In the absence of results at the level of beneficiaries, proxies are used for loan sizes and repayment statistics. However, data access by IFIs can be facilitated by confidentiality agreements and good information systems installed in-house by local Banks, whereby micro-data can be easily anonymised and aggregated through newly available infor-

Development success

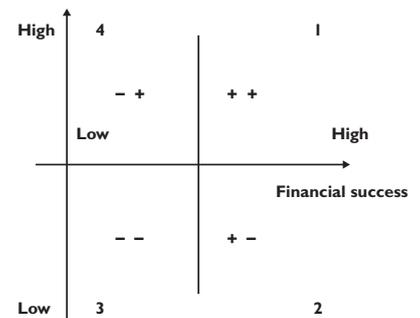


Figure 1: Development and financial success.

mation technologies. Meaningful evidence can also be secured through monitoring.

SME projects are traditionally evaluated against two main criteria, i.e. market-related and non-market-related, to be called financial success ("doing well") and development success ("doing good"). Financial success, measured by profitability of the investment performance, is considered positive for development. But *per se*, high financial returns are not a sufficient condition for generating positive development outcomes. "Doing well" does not necessarily imply "doing good" (Figure 1).

SME evaluation findings (IEG, 2014) have demonstrated that the share of projects that succeed in one dimension only, either developmentally or financially is higher than that of projects that are able to succeed in both. The evidence also suggests a small trade-off between development impacts and profitability and a stronger association between a higher degree of financial success and high degree of development success. How to

1 Total-factor productivity (TFP) growth relates to total output growth not accounted for by the traditionally measured growth in inputs (labour and capital), thus it can be taken as an indication of an economy's long-term technological change or technological dynamism.

2 Inclusive growth became a central idea in development and concerns the creation of opportunities for everyone to participate and benefit from the growth process, making sure that benefits are equitably shared.

better align projects on the high-high end of the axis of Figure 1, by focusing on improving projects' positive correlation along these two criteria calls for evaluation of causal relations so as to sidestep mixed outcomes (quadrants 2 and 4 of Figure 1).

Indirect forms of SME support assistance are common for IFIs, thus respective results should also be measured taking into account the need to obtain information through either secondary sources or third parties. Furthermore, evaluating third party effects (e.g. economic distortions, externalities) is critical. For this reason, private sector evaluation needs more than outcome data. Case-studies, derived from field work can help identify causal mechanisms and indirect effects.

The logic and rationale for SME intervention is not always adequately articulated. Little is known about synergies and linkages or about the most appropriate sequencing of interventions. For instance, identification of linkages related to skills development and technical training has been lagging behind.

SMEs are intrinsically part of the growth process, as both drivers and beneficiaries. The existing research and policy evidence (e.g.

DEGRP, 2013; Vandemoortele *et al.*, 2013; OECD, 2006; SEPT, 2005) indicates that there is scope for private sector and inclusive growth to go hand in hand, although this linkage is neither universal nor automatic. Sectors driving growth are not necessarily labour intensive and growth does not always occur in an inclusive fashion. It seems important to better explore the trade-offs between development outcomes and profitability, through systematic targeting and the usage of appropriate delivery channels.

Private sector interventions (currently less than 15% of total MDB operations) are likely to double by 2030 in most Multilateral Development Banks (MDBs). In order to meet this goal, IFIs will be confronted with changes in players, instruments and intermediation mechanisms. Additionality became an imperative, implying improving the allocation of financial and non-financial resources and a concerted effort amongst different actors: public institutions, national and regional development banks and local private sector actors. The financial sector has to serve the real economy. In today's world of scarce resources, beyond striking the right balance between support to sovereigns and targeted support to private firms, IFIs are increasingly compelled to reconcile "doing more with

less" with the enhancement of "doing well and doing good" in SME targeted assistance.

References

- IEG (2014) "The Big Business of Small Enterprise, Evaluation of the World Bank Group Experience with Targeted Support to Small and Medium-Size Enterprises, 2006–12", World Bank Group, IFC, MIGA.
- DEGRP (2013), "Sustaining growth and structural transformation in Africa: how can a stable and efficient financial sector help? Current policy and research debates", Dirk Willem te Velde and Stephany Griffith-Jones (Eds.), Growth Research Program, DEGRP Policy Essays, December.
- Vandemoortele, M., Bird, K. Du Toit, A., Liu, M., Sen, K. and Soares, F. V. (2013), "Building blocks for equitable growth: lessons from the BRICS", Overseas Development Institute, Working Paper 365.
- OECD (2006), "Promoting Pro-Poor Growth, Private Sector Development", in Promoting Pro-Poor Growth: Policy Guidance for Donors, Paris.
- SEPT (2005), "Private Sector Development and Poverty Reduction: Experiences from Developing Countries", Utz Dornberger and Ingrid Fromm (Eds.), SEPT working paper n. 20, Universität Leipzig.

EVALUATION BRIEFS

This section of the newsletter features concise summaries of recent evaluative work produced by EES members. It is designed to encourage professional interaction within the society. The authors would welcome feedback from evaluation colleagues interested in the issues described below.

Liliana Leone

email: leone@cevas.it

Social Capital and Innovation: a Theory-Driven Evaluation

The Community Foundation of Messina (CFM) was established in Sicily in 2010. Its aim is to promote social and environmental responsibility through local, civil society models of human development. A pilot programme ('Light is Freedom' – Luce è Libertà) is intended to benefit ex-inmates of the Judicial Psychiatric Hospitals (JPH) of Barcellona Pozzo di Gotto.

The program is designed to create 'social capital' based on an approach inspired by Elinor Ostrom's Theory of the Commons: (1) Clearly defined boundaries (effective exclusion by ASD of external un-entitled members); (2) Ensure that those affected by the rules can participate in the decision-making process; (3) Develop a system for monitoring members' behavior; (4) Mechanisms of conflict resolution providing accessible, low-cost means for dispute resolution. A multiple linear regression model confirmed that the theory of change underlying the program is valid. Specifically the evaluation

found that the 'Scale of Participatory Governance' is highly correlated with an index of 'trustworthy relationships at work. A solar energy scheme associated with the program had created jobs, reduced fiscal burdens and induced local cooperation. UCINET 6 software was used to analyse survey data focused on tracking intervention effects on inmates' capabilities. Low rates of recidivism were validated by the evaluation. These evaluative findings have contributed to the diffusion and recognition of an innovative community welfare model in the European space and beyond.

THE AUTHORS

Peter Dahler-Larsen

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



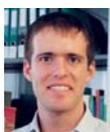
Rick Davies

Dr Rick Davies is an independent Monitoring and Evaluation consultant based in the United Kingdom (UK), focusing on the evaluation of international development aid programs. He developed the now widely-used tool known as “Most Significant Change” (MSC) as part of his PhD research on organisational learning. His website “Monitoring and Evaluation News” and its associated email lists connects evaluators around the world. See <https://richardjdavies.wordpress.com/> and <http://www.mande.co.uk>.



William Faulkner

William N. Faulkner is an evaluation consultant with a background in systems thinking and complexity theory. Until recently, William worked as an evaluation coordinator at Plan Políticas Públicas, an M&E consultancy based out of São Paulo, managing evaluation projects for a range of clients from local NGOs to multinational corporations to municipal governments. In April 2014, William moved into the Integrated Monitoring, Evaluation, and Planning team of the Collaborative Crops Research Program (CCRP).



Fredrik Korfker

Fredrik Korfker (Dutch nationality) gained experience in private sector evaluation during 17 years at the European Bank for Reconstruction and Development (EBRD) in London where he was Chief Evaluator. Since his retirement early 2011 he did short term consultant assignments for the UNDP, the World Bank and the Belgian Government. Fredrik’s background is in project finance/banking having worked for, the Dutch development finance company FMO, a Dutch subsidiary of the former Chase Manhattan Bank, a Dutch mortgage bank and the Inter-American Development Bank in Washington DC. He studied business administration at Nijenrode Business School in the Netherlands. After obtaining a masters degree in economics in 1972 from Erasmus University (former NEH) in Rotterdam, majoring in development planning, he worked five years for United Nations (FAO) stationed in several developing countries and at FAO headquarters in Rome.



David MacCoy

David J. MacCoy is a founding partner of First Leadership Limited, an organization development and coaching company, based in Toronto, Ontario. He is a practitioner of Appreciative Inquiry, an organization analysis and development approach to planning, implementing and evaluating change. With over 40 years of consulting experience in North America and Europe, his particular interest is assisting leaders and teams in the co-construction of solutions for improved performance.



Erica Melloni

Erica Melloni, MSc. PhD. She works at the Istituto per la ricerca Sociale (Milano). She is



a member of the AIV- Italian Evaluation Society. Her areas of interest include institutional capacity building, quality and performance of public administrations, and territorial development. In recent years, her work focused on the use of social mechanisms in evaluation research, with particular attention paid to the development of case studies oriented towards policy learning and design.

Riitta Oksanen

Riitta Oksanen is a senior advisor on development evaluation in the Ministry for Foreign Affairs, Finland. Riitta’s tasks include of evaluation capacity and systems. Riitta represents the Ministry in international initiatives, including the EvalPartners, aiming at stronger national evaluation systems in the partner countries in the global South. She recently chaired for two years the OECD/DAC evaluation network’s task team that focuses on evaluation capacity development. Riitta’s background is in development policy and management of development cooperation. She has previously worked in the Ministry as director for development policy and as an advisor on management and effectiveness of development cooperation. She worked in Finland’s permanent EU delegation as counsellor responsible for EU development policy and cooperation, and chaired the Council’s working group on development cooperation during the Finnish EU Presidency in 2006. Before joining the Ministry in 1999 she worked as a consultant specialising in planning, management and evaluation of development cooperation.



Burt Perrin

Burt Perrin is an independent evaluation consultant with over 35 years’ experience, with numerous publications and presentations about how evaluation can be practical and useful. The major focus of his current work is on the organisation of the evaluation function; planning, designing and aiding in interpretation of evaluations; and quality assurance.



Burt is a previous Secretary-General of the European Evaluation Society. He was awarded a lifetime membership for his outstanding contribution to the Society.

Flavia Pesce

Flavia Pesce holds a PhD in Political Sociology, University of Florence. In IRS since 1999, she is actually Director of the Education, Training and Labour Policies Area. Member of AIV Italian Evaluation Society of which she was a member of the Board of Director in the period 2008–2012. Member of EES European Evaluation Society and of EES Technical Working Group on “Gender and Evaluation”. Co-Coordinator of the NESE – Network of European Evaluation Societies in the period 2010–2012. Since 1999 she was Assistant Professor in Sociology and Sociology of Labour at the Faculty of Educational Sciences – University of Bologna. Her area of interest include labour market and education related policies, gender and social inclusion policies. She is an expert in evaluation especially with regard to Structural Funds programmes and projects.



Robert Picciotto

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



Elsa de Morais Sarmento

Elsa Sarmento is Principal Evaluation Officer at the Ope-



rations Evaluation Department of the African Development Bank. Previously, she worked with the World Bank, Organisation of Eastern Caribbean States, UNDP and the European Commission. She acted as policy adviser at the House of Commons (UK), for several African governments and held management positions at the Research Office of the Portuguese Ministry of Economy. She has lectured for over a decade in economics and econometrics and been a researcher at the CEP (LSE) and the European Parliament, amongst others.

Michael Scriven

Michael Scriven is professor of psychology at Claremont Graduate University and co-director of the Claremont Evaluation Center. Over fifty years of teaching he has taught and held chairs in mathematics, philosophy, psychology, the history and philosophy of science, law, evaluation and education. His 450+ publications have covered a wide range of disciplines including computer science, informal logic, cosmology, international philanthropy and technology. He has served on editorial or advisory boards for more than 40 scholarly journals and was president of the American Educational Research Association and the American Evaluation Association. He has been a Whitehead Fellow at Harvard as well as a Fellow of the Centre for Advanced Study in the Behavioural Sciences at Stanford and of the Australian Academy of the Social Sciences.



Sanjeev Sridharan

Sanjeev Sridharan is the Director of the Evaluation Centre for Complex Health Interventions at the Li Ka Shing Knowledge Institute at St. Michaels Hospital and Associate Professor of Health Policy, Management and Evaluation at the University of Toronto. Prior to his position at Toronto, he was the Head of the Evaluation Program and Senior Research Fellow at the Research Unit in Health, Behaviour and Change at the University of Edinburgh. His areas of interest are health inequities, evaluation methodology and evaluation of complex policies. He has worked on a large range of international development and global health initiatives. He is presently working closely with the China National Health



Development Research Centre to build evaluation capacity in the health sector in China. He is also working on an initiative to develop a post-graduate program in evaluation in five S. Asian countries and advising the Ministry of Health in Chile on utilizing evaluation approaches to redesign health policies in Chile. He is on the board of the Canadian Journal of Program Evaluation, New Directions for Evaluation, and the journal of Evaluation and Program Planning.

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.



Cristina Vasilescu

Cristina Vasilescu holds a Master degree in “Commercial and Institutional Relationships with East European Countries” (University of Macerata) and a bachelor degree in Political Sciences (University of Bucharest). She is an expert in: public policy analysis and evaluation of public policies/projects/programmes, with a particular focus on EU policies and programmes; performance management and performance evaluation for public institutions; evaluation of human resources in public institutions; governance; institutional capacity and capacity building; project design and project management; technical assistance to public institutions. Since 2007 she is researcher within IRS.

