THE EES NEWSLETTER

EVALUATION
CONNECTIONS

Robert Picciotto Editor. **Ole Winckler Andersen** Associate Editor. **Tereza Zvolská** Editorial Assistant.

EES EUROPEAN
EVALUATION
SOCIETY

FEBRUARY 2016

ARTICLES – NEWS – EVENTS

CONTENT

PRESIDENTIAL MESSAGE

Riitta Oksanen

Dear EES members and friends,

In the Global Evaluation Agenda approved in Nepal last November the evaluation community has put forward an evaluation vision for 2020. The agenda is structured along four dimensions that are essential for a functioning evaluation system.

First, a favorable enabling environment that reflects an appreciation of the value of evaluation in all sectors of society. An understanding that evaluation is not a monopoly of the public sector but a partnership involving government, civil society and private actors is critical. It generates a broad based commitment to accountability and learning embedded in evaluation policies, governance and regulatory instruments. These should be backed up by adequate financial and human resources and they imply open access to data by all stakeholders.

Second, adequate institutional capacities on the demand and supply side are necessary. This may require capacity building in Voluntary Organizations for Professional Evaluation (VOPEs), public institutions, civil society organizations, and academia. These institutions play a key role in adapting evalu-

ation resources and processes as the world around us evolves.

Third, evaluation can only be independent, credible and useful if it is built on the sound foundation of individual capabilities. This encompasses evaluators as well as evaluation commissioners, managers and users. For all evaluation actors a core disposition involves internalizing that evaluation is first and foremost about shared values.

The fourth dimension of the Global Evaluation Agenda has to do with inter-linkages among the first three dimensions including joint events, a common set of evaluation terms, collaborative learning opportunities and a shared commitment to the Sustainable Development Goals.

The Global Evaluation Agenda gives VOPEs (including EES) a useful and fresh framework for benchmarking our collective endeavors. Are we fit for the future we want? To begin answering this question the EES Board will start the year by mapping our on-going activities against the EvalAgenda 2020. This will give us a baseline for Society wide consultations on our priorities. We look forward to your engagement through the website, the social media, contributions to the EES Maas-

tricht Conference and Connections, your Newsletter.

What can EES achieve by getting actively involved in the implementation of the EvalAgenda? In 2015 we contributed to its design by identifying where earlier versions were thin. We propose to continue focusing on the challenges that result from: (i) increasingly rapid decision making in volatile and complex environments; (ii) the expanding role of the private sector in society; and (iii) the on-going information technology revolution.

Marco Segone, in the leading article of this issue, stresses the on-going re-balancing of the global evaluation leadership. His analysis makes me ask if the European voice should be amplified. How can we reach out to other VOPEs to create new opportunities for co-operation: to project European approaches and practices and to learn from others?

The Global Evaluation Forum in Nepal confirmed that the time for joint work on evaluation professionalization is now. EES will reach out to other VOPEs for joint or parallel piloting of the Voluntary Evaluator Peer Review (VEPR) initiative pioneered by our Thematic Working Group. The resources we developed in this connection will be shared through the International Organization for Cooperation in Evaluation (IOCE).

In order to be relevant for the future we will work with future evaluation leaders. Mobilising emerging evaluators has been and remains a core EES priority. Facilitating their professional development, tapping their ideas and emulating their ways of interacting will make our Society more nimble and more open to innovation.

Does evaluation have a future? To me this is like asking: will citizens, political leaders, managers, private investors need value-based, independent and credible evidence to guide decisions and action in the future? They will. To respond to their growing and diverse demands we should rise to the challenge of the Global Evaluation Agenda.

EVALUATION IS ON THE MOVE AN EDITORIAL

Robert Picciotto

Propelled by the events of the International Year of Evaluation, energized by the "evaluability" challenges of the Sustainable Development Goals and faced by an exploding demand for evaluation services evaluation is on the move. But where is evaluation going?

Evaluation is like a compass that can tell us where north is. But it does not tell us where we want to go. This is the province of participatory decision making. In the first article of this issue Marco Segone, a prime mover in the largest ever consultation process in evaluation history, describes how the International Year of Evaluation culminated in a Global Evaluation Agenda in November 2015.

This is when evaluators worldwide came to Kathmandu (Nepal) to reason together and craft a vision for the future of their discipline. The new agenda incorporates the universal aspirations of the Sustainable Development Goals. They imply assessment metrics focused on shared prosperity, reduced inequality, gender equity and environmental sustainability.

In this context a determination to serve the public interest emerges as a fundamental disposition for evaluators. This essential pillar of ethical practice is grounded in the recognition that all citizens, especially the poor and disadvantaged, have legitimate interests, needs and rights. In order to respond to them the evaluation discipline acts as a transmission belt that links the citizenry, the knowledge professions and decision making.

Striving for rigor and responsiveness in knowledge creation is another ethical imperative for evaluators. Extracting truth from facts helps evaluators bring reality to bear on the comfortable assumptions and convenient rationales on which public, private and civil society interventions are all too often built. Figuring out impartially what works, why, how and for whom is what evaluators are mandated to worry most about.

In a nutshell 'telling like it is' is central to the evaluator's identity. But understanding what makes interventions effective also requires the ceaseless pursuit of evaluation excellence. Exploring the quality frontiers of the evaluation discipline has never attracted so much energy – or so much controversy as now. Four articles in this issue illustrate diverse aspects of evaluation quality.

Based on World Bank experience Nick York and Geeta Batra demonstrate that evaluation utilization is not a litmus test of quality. Beyond the quality of evaluation reports a host of organizational factors shape responsiveness to independent evaluation findings. Evaluation quality does not guarantee evaluation use or influence. Lessons drawn are only used when "the stars are aligned": it takes two to tango.

Lawrence Salmen's contribution highlights the social validity dimension of evaluation quality. Grounded in case studies of market led interventions it stresses the critical but often neglected role of respectful listening as an evaluative tool. The article demonstrates that amplification of beneficiaries' voice is critical to the design of effective and socially conscious interventions. It suggests that participatory evaluation often praised but rarely practiced should finally become the norm.

The article authored by James Copestake and Fiona Remnant reaches beyond experimental approaches to demonstrate how qualitative impact evaluation can help pinpoint valid findings and put them to work. The recommended approach reduces risks of bias in data collection, streamlines data analysis,

shares findings and facilitates quality assessment and use. This innovative evaluative approach confirms that development evaluation and developmental evaluation have begun to converge: timeliness and nimbleness are new markers of evaluation quality.

In a similar vein Geoff White and Mark Matthews argue that evaluation "bolted on after delivery" (rather than "built in" from the start) does little to enable social learning. They recommend insertion of a tracking mechanism *within* the intervention that regularly updates the odds of different hypotheses being true. Will the logic of Bayesian inference become a standard feature of theory based monitoring systems?

Spirited debate is yet another marker of vitality and dynamism within the evaluation community. This is why Connections provides a platform for the exposition of diverse views. Whether evaluation should privilege learning over accountability – or accountability over learning – has been a chronic source of controversy in evaluation conferences.

Contrasting perspectives as to whether or not there is a trade-off between these two dimensions of evaluation practice are presented by Ian Davies, Caroline Heider, Indran Naidoo and Jacques Toulemonde in our new Forum section. Feedback from readers (e.g. letters to the Editor) would be most welcome.

The Forum section of this Newsletter confirms that evaluation is "a big tent". The policy directions of the new evaluation agenda do not aim at uniformity. Without diversity evaluation would not be adaptable. Without tailor made approaches evaluation would not be responsive. Without judicious combinations of methods evaluation would not be rigorous or effective.

Yet a golden thread runs through all efforts to assess merit, worth and value. It is the notion that evidence matters to principled decision making not only in government but also in the private and voluntary sectors. This belief is spreading throughout society and across borders. Ultimately this is why evaluation is on the move.

A CALL FOR ACTION TO LEAVE NO ONE BEHIND

Marco Segone

In 2015, the "evaluation torch" celebrating the International Year of Evaluation linked 86 events all over the world to discuss the Global Evaluation Agenda 2016-2020 (EvalAgenda). EvalAgenda, officially launched at the Parliament of Nepal in November 2015 in the presence of the Nepali Prime Minister and Parliament Speaker, 100 Parliamentarians, and 450 leaders of the evaluation community, is a call for action to ensure evaluation is fit for the 2030 Sustainable Development Agenda endorsed by 193 Heads of State at the UN General Assembly in September 2015.

The overriding message of the 2030 Sustainable Development Agenda is "to leave no one behind," to ensure "targets are met for all nations and peoples and for all segments of society." How can evaluation help to achieve this commitment? The vision of the thousands of evaluators who took part in the EvalAgenda's 14-month global consultation is that evaluation is an integral part of all efforts by governments, civil society, and the private sector to improve the lives and conditions of their fellow citizens. The vision is that evaluation will become such an integral part of good governance that a decision-maker cannot imagine *not* incorporating evaluation into her or his toolbox;

that no decision-maker would dare make an important decision without having reviewed all the relevant evaluation information; and that evaluators use whatever methods and approaches are most appropriate to the situation to generate the information needed for those decisions.

At the same time, EvalAgenda envisages that evaluation will help to raise the voice of all stakeholders that are impacted by such decisions, particularly those of the marginalized and disadvantaged. Evaluation should therefore be driven by values of human rights, gender equality and social equity.

EvalAgenda, which builds on the success of the International Year of Evaluation, is a further impetus to the global trends described below.

Rebalance of leadership in the global evaluation community

Until 15 years ago, the best national evaluation systems were in the Global North (i.e. USA, UK, Canada). The big majority of Voluntary Organizations for Professional Evaluation (VOPEs) were in the Global North.

The demand for evaluation was in the Global North. Today, this is changing dramatically. Several excellent national evaluation systems are in the Global South (i.e. Mexico, Colombia, Chile, South Africa, Morocco, Benin, Kenya, Uganda, Malaysia). The big majority of the existing 150+ VOPEs are in the Global South. Regional Parliamentarian Fora for evaluation now exist in Asia, Africa, the Middle East and Latin America (but not in Europe or in North America). In the future, demand, supply and use of evaluation will truly be universal.

Stronger demand for evaluation, in particular for equity-focused and gender-responsive ones

The first ever Global Evaluation Event held in a National Parliament is a strong signal of the new trends of policymakers' demand for evaluation. The official launch of the Global Parliamentarian Forum at the Parliament of Nepal is a clear indication that Parliamentarians are becoming new evaluation leaders. In parallel, the strong call by the 2030 Sustainable Development agenda for leaving no one behind and, accordingly, the positive response by the evaluation community with equity-

focused and gender-responsive evaluations, is another positive development. This led to the creation of EvalGender+, a multi-stakeholders movement to ensure evaluation will meet the expectation to inform policies that leave no one behind. EvalGender+, led by UN Women, the International Organization for Cooperation in Evaluation (IOCE), the United Nations Evaluation Group (UNEG), and 33 other VOPEs, UN agencies, Multilateral Banks and the Global Parliamentarian Forum for evaluation, are launching an inclusive process to develop guidance to evaluate the Sustainable Development Goals with an equity-focused and gender-responsive lens. In the future, demand for equity-focused and gender-responsive evaluation will be the norm.

Complexity is the new normal

The 17 goals and 169 targets included in the 2030 Sustainable Development Agenda are interconnected and synergic. While this is necessary, it also adds complexity. In addition, the focus on no one left behind means that socio-cultural-political contexts, as well as power relationships, are critical to evaluate sustainable and equitable development. In the future, evaluation theory and practice will decisively move to evaluation of complexity, adopting system-thinking and network analysis.

A stronger movement for professionalizing evaluation

The challenges above will accelerate the need to professionalize evaluation. Several VOPEs (including the European Evaluation Society, the UK Evaluation Society, the Canadian Evaluation Society, the Japanese Evaluation Society, and the International Development Evaluation Association), are leading this process. UNEG is exploring this area too, and UNEG members (i.e. UN Women and the ILO) already launched professionalization initiatives. In the future, evaluation will become a mature profession.

A more and more diverse multi-stakeholder community

In the past, national Governments, the European Union, VOPEs, UN agencies, multilateral Banks and bilateral donors were the main actors in the evaluation community. Today, new stakeholders are emerging. In addition to the above-mentioned Parliamentarian Fora, local governments (at the state and municipal level, i.e. San Paulo in Brazil and Johannesburg in South Africa) are developing and strengthening local evaluation systems. Impact social investment, which generates billions of dollars and is already more significant than official development aid, is also engaging with evaluation. In the future, the

evaluation community will be truly made up of multi-stakeholders.

To address the new opportunities and challenges of leaving no one behind in a complex world, multi-stakeholders partnerships will become the most meaningful, influential and impactful approach. EvalPartners, the global partnership for national evaluation capacities, brought together IOCE (the network of VOPEs), UNEG (the network of evaluation offices of UN agencies), EvalNet (the network of evaluation offices of OECD/DAC countries) and several other stakeholders. This generated the International year of Evaluation and EvalAgenda, among other things. It is now time to enlarge the partnership to welcome new actors, as well as new challenges.

In the future, evaluation will be an agent of change for the world's 193 nations committed to achieving the 2030 Sustainable Development Agenda. To be sure, EvalAgenda for the next four year is broad and complex. But procrastination is not a sensible option. That is why 2016 must be the year of doing the "right thing": leaving no one behind.

◼

WHEN THE STARS ALIGN: HOW EVALUATIONS GET USED IN PRACTICE

Geeta Batra and Nicholas York

Numerous articles have been written about the factors that affect the use and influence of evaluations (Cousins and Leithwood, 1981; Shulha and Cousins, 1997). Yet evaluation findings continue to be under-utilized. This article builds on the literature and uses World Bank experience to probe the mechanisms that encourage or discourage evaluation use. It develops a theory of change for evaluation use that draws on survey results, case studies and structured interviews with a range of World Bank stakeholders.

The literature has identified a number of causal factors both on the demand and supply

side of evaluation use (Kirkhart, 2000, Henry 2003). World Bank experience confirms the importance of both sets of factors. Indeed, the evidence suggests that outcomes are favorable only when things come together (i.e. if the "stars are aligned").

On the supply side, researchers have pointed to such drivers as evaluation quality, credibility, relevance, communication quality and timeliness. In this context, evaluation independence and methodological rigor matter. Team leadership is another key determinant factor in World Bank experience. Ability and willingness to engage in constructive

dialogue on evaluation findings are especially important.

Beyond the supply side, evaluation use depends on the demand side. It has to do with decision making and policy formation – a complex process driven by competing forces and multiple actors, information needs of stakeholders, availability of competing information sources and the receptiveness and commitment to change of evaluation users. Personal characteristics also play a significant role. Receptiveness to change is also driven by organizational factors and evaluation can help shape a learning culture.

E3
External Stakeholders Acquire Information

O1B
Change in Knowledge and Attitudes Of External Stakeholders

O2B
Change in Behavior of Other Organizations

O3B
Improvements in Performance of Other Organizations

M2
Dissemination of IEG Evaluation Results

M3
Sense of Ownership of Evaluation Credibility of Evaluation Results Quality of Recommendations

M4
Stakeholder Pressure

A3
IEG Produces Evaluation Results

E2
CODE, WBG Managers and Staff Acquire Information

A1
IEG Conducts Evaluation

A2
IEG interacts with CODE, WBG, Managers and Staff during Evaluation Process

E1
CODE, WBG Managers and Staff Acquire New Way of Thinking

O1A
Change in Knowledge and Attitudes of CODE, WBG Managers and Staff

O2A
Change in WBG Organizational Behavior (Strategy, Structure, Policies, Practices, Products, and Project Design

O3A
Improvement in WBG Organizational Performance (Relevance, Output, and Efficiency)

I1
Improvement in Economic/Social Outcomes (Effectiveness and Sustainability)

M1
Interaction with IEG Evaluators

M5
Interaction between IEG Evaluator and WB Mgt. Timeliness of Evaluation Relative to Decision Point
Champions for Reform
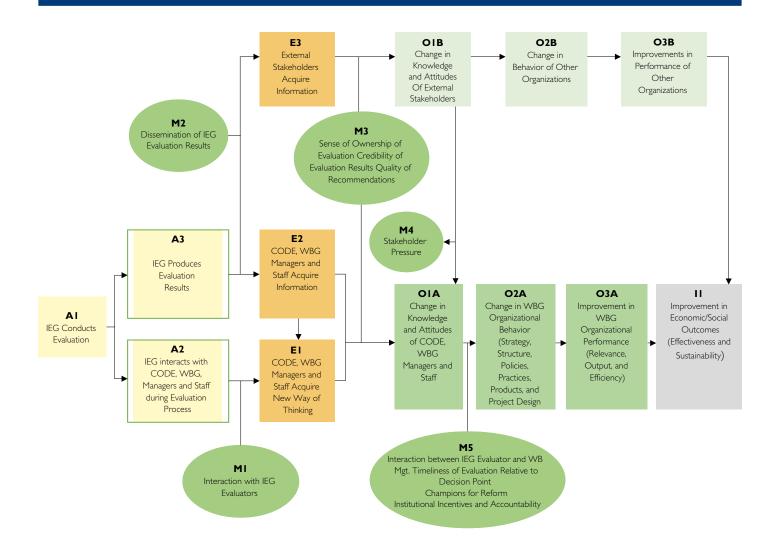Institutional Incentives and Accountability

Figure 1: Influence of IEG Evaluations on WBG Development Effectiveness.

This is why the Independent Evaluation Group (IEG) has sought to build relationships with key interlocutors and evaluation champions within the World Bank and act strategically to help improve the receptivity of the institution to lessons of experience and evaluation findings. Ultimately, then, evaluators can bring the horse to water but it will choose whether or not to drink. Decisions on evaluation use are driven not only by operational teams but also by senior management and the executive Board since leadership sends signals throughout the organization. IEG's assessments of self-evaluation systems recognize this reality.

Practical experience in IEG points to the fact that evaluators can usefully invest in building productive relationships and credibility

for a productive and constructive dialogue. Even small decisions can make a big difference – the length and format of a report, the way a meeting is conducted, the timing and location of a conference and the tone and language used in the communications, both written and oral.

Evaluators tend to focus on what comes naturally to them – e.g. quality of methods and reporting standards. They are not always skilled communicators and they often fail to draw the implications of supply side factors in selecting evaluation topics and selecting evaluation teams. This is problematic since the toughest issues and obstacles to change lie in the institutional context, receptivity to change and a host of demand side factors. This suggests that they are not always driven

by a sound theory of change or mental model in relation to evaluation use.

Straightforward linear models inspired by the traditional policy cycle assume that evidence based decision making is already in place and that the decision making mechanism is monolithic. But this is not always the case. The notion that independent evaluators can on their own identify questions, collate and analyze relevant evidence, present findings and recommendations and then disseminate them into the grateful and waiting arms of policy and decision makers is flawed. Nor even if this is the case are evaluators in a position to generate precise and targeted recommendation – the last piece in the jigsaw or the magic bullet which allows the policy or decision maker to solve a problem.

Evaluation is a collective enterprise and instrumental use is not the only way to influence outcomes.

Figure 1 presents a program logic model (PLM) for IEG evaluations. It illustrates the causal chain from evaluation activities to improvements in development effectiveness, calling attention to the factors that mediate the strength of the links in the chain. Some of these factors, such as methodological rigor, are under the direct control of IEG; others, such as champions for reform, are not. Factors outside of the direct control of IEG still affect the influence of IEG evaluations. This has major implications for evaluation management.

IEG conducts various types of evaluations (**A1**) and, in so doing, interacts with the board committee on development effectiveness (CODE), managers and staff (**A2**). The extent and nature of the interaction (**M1**) during the evaluation process can have a major bearing on evaluation use. Judicious engagement of these actors in evaluation activities may help them think critically about programs (**E1**) as well as develop a sense of ownership of the evaluation process and subsequent results.

The principal output of an IEG evaluation is a report that presents results, including findings, conclusions, and recommendations (**A3**). Results are presented to CODE along with written responses from WBG managers (**M2**). Results are also disseminated more broadly within the WBG and among external stakeholders through publications, presentations, press releases, and peer-to-peer networking. In so doing, members of CODE, WBG managers and staff, and external stakeholders acquire information about the particular subject of the evaluations (**E2** and **E3**).

As a result of acquiring valid evaluation information CODE, managers and staff may gain knowledge which either confirms or changes dispositions with respect to an intervention being evaluated (**O1A**). Thus evaluations may influence decision makers' attitudes and shift operational policy *directions*, e.g., a program that was previously well-regarded may fall out of favor based on evidence that it is not meeting relevant objectives efficiently. Alternatively the salience of an issue may be revisited, e.g. it may suddenly be perceived as critical rather than peripheral. Evaluations may also affect the knowledge and attitudes of external stakeholders (**O1B**) who may, in turn, exert influence on CODE and WBG management (**M4**).

Thus the extent to which an evaluation affects knowledge and attitudes depends on several factors (**M3**) that relate to the characteristics of the evaluation and the environment in which the evaluation takes place, including a sense of ownership of the evaluation, the credibility of evaluation results, evaluation skills, knowledge and experience of the lead evaluator; quality of design, quality of data and adequacy of sample; methodological rigor; resources made available for the evaluation; credibility of evaluation results; quality of recommendations – practicality, clarity, coherence, feasibility and cost effectiveness.

Yet, new knowledge may not lead to changes as per IEG recommendations (**O2A**). There are several mediating factors that determine instrumental use (**M5**) of evaluations within WBG, including the extent of interaction between IEG evaluators and WBG management, timeliness of the evaluation relative to a decision point, champions for reform, institutional incentives and accountability, presence of sufficient WBG interlocutors with good understanding of how to use evaluation, etc. The demand side matters.

References

Batra, G., Nashat, B., Oldsman, E. (2012). Assessing the Influence of IEG Evaluations – A framework and Case Studies, *IEG Working Paper*.

Cousins, J. B., & Leithwood, K. A. (1986). Current Empirical Research on Evaluation Utilization. *Review of Educational Research*, 56(3), 331–364.

Henry, G.T. & M.M. Mark. (2003). Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions. *American Journal of Evaluation*, 24(3).

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 5–23. doi: 10.1002/ev. 1188. 2000.

Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, 18, 195–208.

LISTENING FOR CHANGE FOR AN INDEPENDENT EVALUATION OF FOUR PRIVATE SECTOR ORIENTED INTERVENTIONS

Lawrence F. Salmen

This brief article summarizes findings and lessons from four case studies (Independent Evaluation Group, 2012) conducted in as many countries for an evaluation carried out for the International Finance Corporation of the World Bank Group:

- A micro-credit and savings financial service in a large Central African country;
- A telecom project in a small East African country;
- A farm forestry project in a large South Asian country; and
- A water and sanitation project in a middle-sized East Asian country.

How people perceive their own capacity in relation to opportunities presented by a development initiative profoundly affects their motivation and ultimate effectiveness in escaping poverty. In successful development projects people become the architects of their own development. Tapping into this human dimension is best carried out by respectful listening used to amplify the voices of people.

The evaluation was carried out by local research teams who systematically listened to randomly – selected intended beneficiaries. The approach used two primary techniques: conversational interviews and focus group discussions. This directed, pragmatic, qualitative and yet quantifiable evaluation approach is akin to market research wherein the objective is to see how a product is adapted to the culture and capacities of the consumer. As applied to development work, this combination of conversational interviews and focus groups – along with participant observations – is a complement to project monitoring (Salmen, 2002). As Daniel Ticehurst (2013) elucidates, "monitoring…is about developing ways of being accountable to poor people …through listening, gauging how responses vary, feeding back interpretation and taking appropriate actions".

Local teams of individuals were selected on the basis of such criteria as recall ability, writing skills, fluency in the language of intended beneficiaries, and, crucially, the innate quality of being good listeners. To better assure field work quality, a guide was handed out to local team members. The two key qualitative research techniques were set forth as follows:

"The key to conducting a **conversational interview** *is establishing good rapport between interviewer and respondent. Once trust is created, information elicited should be valid. Good rapport will generate unsolicited data which may be more important than direct responses to predetermined topics in the interview guide. Respect for the intended beneficiary (respondent) is basic to the establishment of rapport. The timing of the interview, its duration and time of day, should be determined by what is most convenient and acceptable to the beneficiary. Generally, interviews should not exceed forty-five minutes to one hour at any one time; to cover the interview guide may require more than one conversation. Short interviews during repeat visits show respect for the beneficiary's time and facilitate increased rapport. It is important that the topics in each interview be explored with sufficient probing to elicit precise, useful responses. Once sufficient rapport has been established, efforts should be made to ascertain what specific recommendations the interviewee may have for project improvement.*

Focus groups *allow for a coverage of more persons in a given time period than one-on-one interviews. Focus groups are conducted by two persons, one of whom leads the discussion while the other takes notes. Generally the group is composed of 6 – 12 persons. The focus group helps elicit responses triggered by the presence of peers which may reveal values or attitudes not expressed in one-on-one".*

A total of over 1,500 persons were interviewed in all four countries, roughly one third individually, the large remainder in focus groups. The individual interviews were with key informants (community leaders, school principals, church spokespersons, etc.) and randomly selected beneficiaries. The sample was stratified by income and gender and, in the micro-credit case, proximity to a financial institution. In all cases control groups/ communities were included to better ascertain the impact of the project on intended beneficiaries.

The fundamental lesson emerging from these four case studies is that development projects work better when they are based on a sound and thorough understanding of the micro-economic and cultural conditions prevailing among the people for whom the projects are intended. A corollary, is that this kind of grounded understanding is best gained by systematic inquiry based largely on listening to intended beneficiaries in the communities where they work and live. Each of the four country case studies, or field studies, represented a different sector of developmental activity: finance, telecom, agriculture, and water.

All four projects had, in addition to profitability, objectives of reducing poverty, three by increasing access to services (credit, communications, and water connections) and the fourth by forging a stronger linkage between farmers and the private sector whereby the former could benefit from the sale of pulpwood to the latter. The effectiveness of the poverty reduction strategy of each of these IFC-supported endeavors may be seen to vary in direct proportion to the degree to which the project product was adapted to the particular needs and conditions of the intended beneficiaries.

The lessons may appear obvious and commonsensical. But experience suggests that they are honored more in the breach than in the observance.

- Understanding the poor through demand assessment is essential to project success.
- Supply needs to be adjusted to market realities, and companies need to engage with beneficiaries in a positive manner.
- Facilitating access to a good or service provided by a project is a key element of project success.

- The affordability of what a project offers is a precondition of high performance in a development project.
- A final lesson which may be drawn from these four projects is the need for ongoing training to assist the beneficiaries to make full and proper use of the development opportunities provided by the project.

In a nutshell, good development projects resonate with the people for whom they are intended. The people's behavior and values should never be assumed, or taken for granted. Prior to and during project implementation, a project's intended beneficiaries should be consulted as partners in an endeavor which will ultimately become their own. The review of these four IFC-supported development projects shows the importance of integrating people with activities carried out in their name such that they are enabled and empowered to become the determining instruments of their self-realization.

References

Independent Evaluation Group, (2012). *The private Sector and Poverty Reduction: Lessons from the Field*, World Bank, IFC, MIGA.

Salmen, L.F. (2002). Beneficiary Assessment: An Approach Described, *Social Development Papers*, No. 10/August. World Bank.

Ticehurst, D. (2013). The lost prophets: A practitioner´s view of M&E, *Connections*, September.

CREDIBLE AND COST-EFFECTIVE IMPACT EVALUATION FOR COMPLEX CONTEXTS

James Copestake and Fiona Remnant

Moving beyond the RCT debate, how well are we doing in sharpening *other* tools for evaluating impact in a credible and cost-effective way? Contribution analysis, process tracing and other qualitative approaches to investigating causation have become more widely known and understood. But for many would-be commissioners there is still too much 'black box' uncertainty about what these approaches can yield, how reliably, how quickly and at what cost. In this article we share reflections on one action-research project that set out to raise standards and improve clarity about what qualitative impact evaluation can deliver. Our aim is partly to promote the specific protocol we have been developing (referred to as the QUIP), but more importantly to encourage others to invest in similar experimentation and to share their experiences.

Assessing Rural Transformations (referred to as the 'ART Project') was funded by the UK *Economic and Social Research Council* (ESRC) and the *Department for International Development* (DFID) between 2012 and 2015. While managed by the University of Bath it was designed from the outset as a collaborative venture, involving three NGOs (*Farm Africa, GSHA* and *Evidence for Development*), university staff in Ethiopia and Malawi, and an Excel design company (F1F9). Together we identified four rural development projects (two each in Ethiopia and Malawi) for piloting work. These included a longitudinal study to monitor changes in key food security indicators, and two rounds of QUIP studies designed to supplement the quantitative longitudinal evidence of change with qualitative evidence on causal factors behind the change. Much of this material is now available on the web at go.bath.ac.uk/art.

So what did we achieve? Here we focus on three points. The first is concerned with reducing risks of bias in data collection; the second with streamlining data analysis, and the third with how findings were shared, quality assessed and used. We then conclude by discussing problems encountered, issues outstanding and plans for further work.

Reducing risks of bias in data collection

We tackled the attribution question head-on, by relying mostly on self-reported causal statements from intended beneficiaries of the programmes being evaluated. One obvious advantage is that it avoids the extra costs and complications of identifying and interviewing a comparison group. On the downside, self-reported evidence may be susceptible to pro-project and confirmation biases. To address this risk the QUIP incorporates a process of deliberately *blinding* both interviewer and respondent to the specific purpose of the data collection. The interviewers are therefore not given any information about the NGO or project being assessed, and trained to conduct a semi-structured questionnaire that asks questions about changes in areas/domains of people's lives over a set period. Narrative responses that explicitly, or implicitly, link the activities being evaluated to intended outcomes can therefore be judged to have been volunteered rather than prompted. Blinding has the additional benefit of making the QUIP more open-ended and exploratory, potentially yielding insight into incidental drivers of change and unintended impacts (Copestake, 2014).

Streamlining data analysis

The key to this turned out to be smart use of Excel to turn narrative data into useful outputs quickly and transparently. This hinges partly on semi-structuring data collection instruments to reflect broad domains of expected impact. It also involves building clear analytical steps into the protocol. For example, in one exercise the analyst sorts data according to how much it *explicitly* supports the theory of change underpinning the project being evaluated, *implicitly* does so, or identifies drivers of change that are *incidental* to it. Table 1 illustrates one way in which this can be presented in summary form. Evidence

	Explicit		Implicit		Incidental	
	Positive	Negative	Positive	Negative	Positive	Negative
Food production	HI H2 FI	H4 F2	H3		HI	H2 H3 F2
Cash Income	H2 FI	H4	HI H3		H3	HI H2
Cash spending	H2		F2	HI		
Food consumption	H3			H2	FI F2	H3
Local relationships		F2	HI			
Asset accumulation	H2 FI		HI H2 FI	F2	HI	
External relations	HI H2 H3		HI H2 FI		HI	

Table 1. Index of changes reported by households and focus groups relative to the project theory of change (illustrative).
Note: H1–4 refer to semi-structured interviews, F1–2 refer to focus groups. Narrative text to which codes refer can easily be cross checked to source data.

of explicit impact can be likened to 'smoking gun' evidence in support of the programme theory of change, whereas implicit evidence can be viewed as more like 'hoop test' evidence (Punton & Welle, 2015), and incidental evidence can inform assessment of counter-theories.

Sharing and using findings

The production of draft QUIP reports provided an interesting opportunity for *strategic triangulation*. Meetings to discuss the reports can bring together field researchers (unblinding them), NGO staff and data analysts for the first time. With careful facilitation this serves not only as a form of triangulation but can yield additional insight into what worked, what didn't and why. For example, in the pilot studies timing both of data collection and of how interventions related to seasonal variation and market fluctuations emerged as a key issue.

Evaluation is inherently about overcoming information asymmetries among stakeholders. Who knows what and when affects not only the credibility of evidence (e.g. through blinding) but also trust, power relations and potential for future collaboration. Impact evaluation is never a politically neutral process, and can reinforce as well as reflect unequal power relations. But by being used to promote dialogue it can also alter relationships in a progressive way.

Problems, issues and plans

While the QUIP studies generated evidence on different factors contributing to change it threw less light on the relative importance of these changes. Our intention was to address this through micro-simulation: combining QUIP's identification of causal mechanisms with use of monitoring data to calibrate their magnitude. Doing this in a timely way requires close coordination of data collection. Sample selection was a second challenge. QUIP sampling can be more opportunistic than for RCTs because each case stands alone as an incremental contribution, both to accepting or rejecting the project theory of change, and to understanding within-sample variation. But the credibility of generalisation is enhanced if key characteristics of the entire target population, the monitoring sample, non-respondents and the 'Quipped' sample can be clearly summarised and compared.

Options for building on the lessons learnt in designing and piloting the QUIP are many. Looking ahead, we would like to test it on rural livelihood interventions in a wider range of contexts, as well as adapt it for use in assessing change in urban contexts and for interventions in other sectors, such as housing and education. This will also provide an opportunity to explore variations from the current model (based on ten person days of field work to conduct around 24 household interviews, plus four focus groups). More ambitiously we would like to explore how features of the QUIP (including blinding, streamlining data analysis and strategic triangulation) can be incorporated into organisational, value chain and systemic evaluations. At the same time we recognise the value of honing a range of tools that are often most useful when used in combination to suit differing contexts and needs.

More information on the QUIP methodology and examples of final reports from Ethiopia and Malawi can be downloaded from: go.bath.ac.uk/art

References

Copestake, J. (2014). Credible impact evaluation in complex contexts: confirmatory and exploratory approaches. *Evaluation*, 20 (4), 412–27.

Punton, M. & Welle, K. (2015). Straws-in-the-wind, hoops and smoking guns: what can process tracing offer to impact evaluation? Sussex: Centre for Development Impact, Practice Paper No. 10. April.

AN 'ODDS ON' APPROACH TO EVALUATION

Geoff White and Mark Matthews

In the December 2013 issue of *Evaluation Connections* we proposed a faster, smarter and possibly cheaper way of evaluating public policies. It is based on formulating succinct policy propositions against the available evidence and testing them as hypotheses in a structured, sequential manner as further evidence emerges. This would provide a stronger foundation for learning than the more audit-based approach – would be faster and could be cheaper.

In parallel, an assessment by the National Audit Office (NAO) of UK policy evaluations (2013) found that evaluation methods had not generally facilitated learning. The NAO highlighted a wide range of causes – narrow evaluation coverage, inappropriate methods, poor quality evidence, inadequate dissemination and limited use of the results.

We suggest the fault may run still deeper. The audit function has tended to dominate the learning role, with evaluations being bolted on after delivery rather than designed as an integral component of policy formulation and implementation. The result is that weak evaluation designs that do little to enable learning are often adopted. What is needed, especially when evaluation budgets are tight, is *innovation* in evaluation processes and methods – aiming for smarter and possibly cheaper ways of learning.

This article develops our earlier thoughts into what we call an 'odds-on' approach to evaluation since we take the view that policies rarely have a 100 % chance of success at the outset even when they are tried and tested (e.g. because the context has changed). Where the interventions are innovative and/or are set in complex and uncertain circumstances both risks and rewards are likely to be relatively high. Evaluation should provide the evidence to enable effective risk management and the intervention adapted where appropriate to strengthen the odds of success.

What we propose draws on the approach advocated by Patton (2010) to assist the development of policy in complex and uncertain contexts by setting up close to real-time feedback in a continuous learning loop which can incorporate both qualitative and experimental methods. It requires the forward-looking appraisal of a proposed policy intervention to be couched as a set of hypotheses about what it might achieve and then uses the monitoring and evaluation of evidence to test the relative strength of these hypotheses. At the start of this process available evidence about results secured in different contexts can be used to estimate the odds that each of these hypotheses will turn out to be correct. As experience unfolds and more evidence becomes available then these odds can be adjusted to reflect actual experience.

The use of competing hypotheses drawn from diverse theories of action and change within a single policy intervention or a set of similar interventions reduces the risk of misdiagnosis. These competing hypotheses can be tested and eliminated (or at least the odds of the hypotheses being correct can be reduced) via evidence gathering and analytical work. This may either confirm the initially assumed odds, or may lead us to revise these odds or, most usefully, develop updated hypotheses more consistent with the available data. The hypotheses testing would use diverse sources of evidence including case studies and other qualitative evidence building on the methods in Qualitative Comparative Analysis (Ragin, 2008) and process tracing (Collier, 2011) to draw robust inferences from such evidence sources.

The broad principle, that the odds of different hypotheses being true is updated when new information is received, underpins what is known as Bayesian inference. Whilst this principle is easily grasped, its expression in formal statements of method and analysis has often been impenetrable to non-specialists
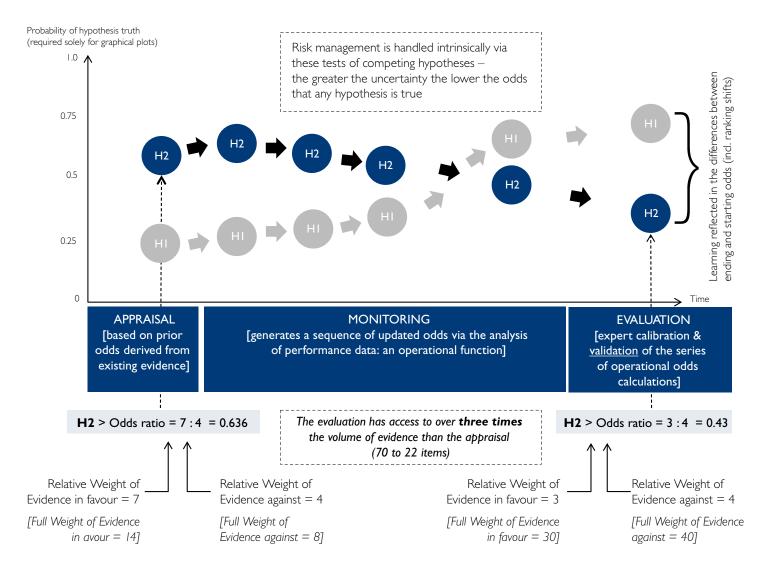
due to the overly complex mathematical approaches used. There are, however, simplified approaches to Bayesian updating based on the use of 'natural frequency' data that directly express the relative incidence of observations in a manner that is more readily intelligible and lends itself to odds-based expressions of risk, (Gigerenzer, 2002). As in clinical medicine, this approach allows the overall prevalence of diagnostic errors in hypothesis tests (i.e. false positive and false negative results) to be factored into specific evaluations, opening up an avenue to apply the diagnostic methods used in psychology and in signal processing [1] to policy development and evaluation.

In this approach, the odds of different hypotheses being correct given currently available information are calculated by determining the extent to which information supports, or does not support, each of a set of specific alternative hypotheses. These odds are calculated by summarising all available evidence in a binary (true or false) manner in regard to each alternative hypothesis.

For example, the public funding for science and technology through institutes or programmes could be characterised as having a range of different 'yield' hypotheses depending on whether the anticipated spin-off process works through company formation, attraction of inward investment, increased licence income, new skills acquisition and development etc. In each case, and as experience unfolds, the goal is to be able to calculate the odds that each (or none) of these hypotheses is correct using currently available information – and expressed in betting terms as odds *for* relative to odds *against* (i.e. 3:1 for or 2:1 against etc).

If it turns out that the licence income hypothesis is exhibiting the most favourable odds of being the best, then this insight can be used to reshape the existing and future interventions. If the odds do not look good for any of the intervention yield hypotheses, then a case be-

---

1 Signal processing is concerned with identifying errors in communication and processing.

Probability of hypothesis truth
(required solely for graphical plots)

Risk management is handled intrinsically via these tests of competing hypotheses – the greater the uncertainty the lower the odds that any hypothesis is true

Learning reflected in the differences between ending and starting odds (incl. ranking shifts)

1.0

0.75

0.5

0.25

0

Time

APPRAISAL
[based on prior odds derived from existing evidence]

MONITORING
[generates a sequence of updated odds via the analysis of performance data: an operational function]

EVALUATION
[expert calibration & validation of the series of operational odds calculations]

H2 > Odds ratio = 7 : 4 = 0.636

*The evaluation has access to over **three times** the volume of evidence than the appraisal (70 to 22 items)*

H2 > Odds ratio = 3 : 4 = 0.43

Relative Weight of Evidence in favour = 7

Relative Weight of Evidence against = 4

Relative Weight of Evidence in favour = 3

Relative Weight of Evidence against = 4

[Full Weight of Evidence in avour = 14]

[Full Weight of Evidence against = 8]

[Full Weight of Evidence in favour = 30]

[Full Weight of Evidence against = 40]

The diagram above depicts how the odds-updating process could operate.

gins to emerge for either complementary support or indeed cancelation of the programme – with consequent lessons for the future.

The advantages of the proposed approach are three-fold:

1. Seamlessly connecting appraisal, monitoring and evaluation – which are too often approached as separate activities.
2. Integrating risk and uncertainty assessments into appraisal and evaluation rather than treating risk management as a bolted-on compliance activity.
3. Potentially providing faster and more cost-effective evaluations than the traditional audit approach. Evaluations can be updated frequently and at low cost because results can be expressed in a tabular form that summarises the evidence for and against

each competing hypotheses as odds. This succinct and fast-to-deliver format makes it easier to brief senior officials and ministers on progress achieved.

Political imperatives are driving devolution of policy capacity to the national and local levels in the UK. This could provide a useful opportunity to develop these faster, cheaper and integrated approaches to appraisal, evaluation and risk management. We are keen to discuss possible pilot implementations of this new approach with fellow evaluators and government departments, agencies and devolved authorities.

References

Collier, D. (2011). Understanding process tracing. *Political Science and Politics* 44, No 4.

Gigerenzer, G. (2002), *Reckoning with risk.* Penguin.

National Audit Office (2013), *Evaluation in government*, December.

Patton, M. Q. (2010), *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Guilford Press.

Ragin, C. C. (2008). *What is Qualitative Comparative Analysis?* Retrieved November 2015 from http://eprints.ncrm.ac.uk/250/1/What_is_QCA.pdf.

Swets, J.A., Dawes, R.M., Monahan, J. (2000) Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest.* I, 1–26.

This new section of Connections puts forward diverse perspectives on important evaluation questions. It is designed to provoke reflection and trigger debate. Readers are encouraged to respond (letters to the Editor). Looking ahead we would welcome ideas and contributions on salient evaluation topics worthy of thoughtful Forum deliberation in future issues of your Newsletter.

FORUM: IS THERE A TRADE-OFF BETWEEN ACCOUNTABILITY AND LEARNING IN EVALUATION?

Ian Davies

Accountability, in the institutional and organisational universes within which mainstream evaluation usually takes place, is conditional on the existence of a mutually acknowledged relationship between at least two distinct entities[1]. It may link organisations, constituent parts of an organisation, organisations and their funders, etc. The nature of the accountability relationship may be shaped by historical precedents and negotiations or such factors as statutory obligations, political imperatives, power dynamics, governance and management requirements, etc.

As such, accountability may exist in many forms. However, in the context within which a significant proportion of mainstream evaluation takes place, accountability is typically conceived and put into practice in a manner consistent with its classic definition: "the obligation to account for a responsibility conferred". This tends to reflect a hierarchical relationship in which one entity (the principal) has a degree of authority or power over the other (the agent).

When an evaluation is conducted for the purpose of accountability, whether or not it is conducted for other purposes as well, it implicitly accepts both the nature and the form of the accountability relationship it is superimposed on. And when the relationship is unequal, e.g. "donor" and "recipient" in the international development domain, evaluation often has little choice but to reflect, or at the very least be consistent with, the dominant entity's interests.

This is straightforward from an audit perspective since audit is a process embedded within an accountability relationship. Audit, *inter alia*, verifies compliance, provides assurance with respect to performance, financial or otherwise, and its approaches and techniques are developed and refined continuously to best serve the accountability relationship. Audit produces evidence based assessments which may be translated into knowledge and used by organisations in different ways, e.g. loop learning.

For evaluation however, especially when it purports to serve the public good, the accountability purpose constitutes a major threat to its ability to consider fully world views other than those embedded in the accountability relationship and to question the power demanding the accountability. And it is evaluation's ability to assess merit, worth and value fully and freely, without coercion, implicit or otherwise, that makes evaluation independence imperative and this is where the fundamental distinction between evaluation and audit lies. Evaluation unlike audit is not about finding out for accountability purposes whether the intervention achieved its intended results.

For evaluation therefore the question is what are the tradeoffs between accountability and independence while contributing to the public good. These tradeoffs are implicit in the choice of evaluation approaches, methodologies and practices. To posit that, for example that evaluation requires monitoring data, indicators, or all manner of measurable results is but one way in which such tradeoffs are struck through methodological choices.

Evaluation scope statements, probably one of the most misunderstood and misused components of evaluation planning, rarely, if ever, consider the political implications of the choice of evaluation approach and methodology, i.e. what dominant world view is being served, whose voices are ignored and whose power is denied?

On "learning", the practical question is whether, and if so who, is trying to find out what for what purpose. Evaluation terms of reference and reports, especially in the development industry, are often replete with jargon about "lessons learned" and their treatment is mostly undistinguishable from what is found in a performance audit report. Yet these are put forward with little apparent understanding of functional distinctions between information, knowledge and organisational learning processes, as well as of their political dimensions.

Yet for evaluation to contribute meaningfully to "learning" it must first and foremost be itself a learning function, the realization of a learning paradigm, i.e. where the value of evaluation resides in the learning process.

For the commissioner of evaluation the question is what are the tradeoffs between the use of audit, evaluation or other processes, for the purpose of accountability. That many commissioners appear to prefer evaluation to audit for accountability purposes raises interesting questions of political economy, e.g. whose interests is the evaluation serving and what approaches and methodologies most reflect dominant interests?

Consider for example the OECD DAC criteria for evaluation of development interventions. Or the current rush of development industry organisations to come

---

1 This article does not discuss accountability or learning from an individual perspective.

up with ways to "measure" achievements of the Sustainable Development Goals. Do they foreshadow a commoditisation of the political process of which the Sustainable Development Goals are but a statement? Is this evaluative accountability response a way of domesticating the shared values and negotiated visions of *homo politicus* by reverting to a *homo economicus* stance?

One thing is clear: taking stock of the ongoing discourse on the relationship between "accountability" and "learning" in evaluation, we are still a long way off from the day when evaluation will be considered as the intelligence of an organisation, the wisdom of a society or something of more value than a set of "deliverables".

Caroline Heider

By three methods we may learn wisdom: First, by reflection, which is noblest; second, by imitation, which is easiest; and third, by experience, which is the bitterest.

Confucius

Colleagues along my long career have said: learning and accountability are irreconcilable. But, as a practicing evaluator, I have experienced learning and accountability as two sides of the same coin. Why does this debate matter so much? Because it gets to the root of what makes evaluation effective: when and why does it stimulate learning, how does it influences change.

In Jacques Toulemonde's article below some of the things counted as part of the learning function of evaluation – in particular "providing … lessons on how to better achieve the desired development outcomes, avoid undesirable effects" – serve accountability (were outcomes achieved?) as much as learning (how to improve?).

One theory suggests that accountability impedes or undermines learning; and as long as evaluation tries to do both, it will not be effective. The other perspective suggests learning and accountability are mutually reinforcing. In particular, learning is derived from an understanding of success and failure to determine how to be more effective in future. In this case, accountability is the foundation for learning.
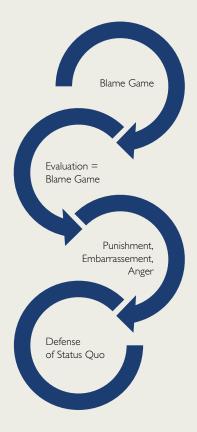
In short: the debate is, in my view, about the underlying theory of how people are motivated to learn and change behaviors. Much

Culture
of Accountability

Evaluation =
Opportunity

Reflect, Learn,
Change

Striving
for Excellence

Blame Game

Evaluation =
Blame Game

Punishment,
Embarrassement,
Anger

Defense
of Status Quo

research has gone into trying to understand exactly this question, albeit without coming to conclusive responses. Kahnemann's behavioral economics shows that people act on instincts, which they learn by doing (*Thinking Fast*), but find it much harder to learn from and act on evidence or knowledge (*Thinking Slow*). This view is contrasted by research to show how doing combined with reflection leads to deeper learning (Di Stefano et al., 2014), or the work of Bonner and Sprinkle which suggests that the effectiveness of in-

centives to learn or behavior change – such as rewards, recognition, or feedback – depends on the complexity of the task.

Empirical evidence of Collins suggests that unless companies "face the hard facts" – which includes accountability for failure – they are not able to build sustained success and grow from "Good to Great". But, others (Brookfield, 1986) have shown that adult learning is facilitated by a sense of personal power and self-esteem and positive rein-

forcement. The latter might be challenged by negative feedback, including from evaluation.

So where does leave us? There is no simple answer to the question whether evaluation works better when it has the carrot of learning for better results with or without the stick of accountability.

Instead of simply separating learning and accountability, I suggest we look at deeper issues that stand in the way of learning and achievement of results. Corporate culture or attitudes towards and of evaluation play an important part in shaping how evaluation can be effective in influencing change with a balance of learning and accountability.

Accountability is often confused with blame. When the "Blame Game" is the dominant culture, the focus is on culprits and scape goats rather than results. Learning – from evaluation or otherwise – is hindered because honest feedback, when critical, is associated with embarrassment and punishment.

By contrast, a culture of accountability accepts failure as an opportunity to learn, grow and strive for excellent. Critical feedback from evaluation or other processes are part of a culture that stimulates debate, learning, and change.

> **„Instead of simply separating learning and accountability, I suggest we look at deeper issues that stand in the way of learning and achievement of results.“**

Accountability derives from the assessment whether objectives set at the beginning were attained. It is the planner who sets his/her yardstick at the beginning of the process. If accountability is removed from evaluation, does it mean we do not care about the objectives set at the beginning and what we can learn from our efforts to achieve them?

What would be the alternative? Not make assessments or evaluate what has happened, or introduce new yardsticks at the time of evaluation? Neither of these options are satisfying: they redefine evaluation away from original purposes, move it closer to what program planners and implementers can do through internal review processes, and above all, it would not address the real crux of institutional learning.

References

Bonner, S. Sprinkle, G. (2002). The effects of monetary incentives on effort and task performance: theories, evidence and a framework for research. *Accounting Organizations and Society*, 27, 303–345. doi: 10.1016/S0361–3682(01)00052–6.

Brookfield, S. (1986). *Understanding and facilitating adult learning: A comprehensive analysis of principles and effective practices.* McGraw-Hill Education (UK).

Collins, J. C. (2001). *Good to great: Why some companies make the leap… and others don't.* Random House.

Di Stefano, G., Gino, F., Pisano, G. P., & Staats, B. R. (2014). Learning by thinking: How reflection aids performance. *Harvard Business School NOM Unit Working Paper*, (14–093), 14–093.

Kahneman, D. (2011). Thinking fast and thinking slow. *Farrar, Strauss and Giroux, New York, NY.*

The Accountability training company, (2009). Achieving Results through Greater Accountability. *Partners in Leadership.*

Indran Naidoo

In this short piece I wish to state upfront that I write from the perspective of an independent evaluation office function, which reports directly to United Nations member states, and where the focus is at the corporate governance level. Most UN evaluation offices do not have this arrangement and are embedded within the management function of the organization, with very limited operational independence and without a final say on issuance of evaluations to the public.

The relatively recent progress in entrenching independent evaluation in the United Nations Development Program (UNDP) reflects a growing member states' demand for more voice and whilst there is inevitable contestation around findings and recommendations the work produced by the Independent Evaluation Office (IEO) enjoys high credibility.

What is perhaps even more important is that IEO's independent status allows it access to diverse stakeholders. It also enhances the credibility associated with presenting evidence based results in a transparent manner. Giving a voice to actual and potential beneficiaries at ground level means that the function helps to connect legitimate voices to policy making.

Independence does not mean isolation or ignorance of subject matter. Rather it conveys constructive and principled engagement with emphasis on impartiality and rigor in methodology. This helps learning. So does full disclosure within the ethical constraints imposed by generally accepted evaluation guidelines and parameters e.g. with respect to confidentiality.

Before striking a balance between accountability and learning one needs to answer the question of who/whom or what body the evaluation function is answerable to. Agency heads invariably prefer it to be answerable to themselves since an evaluation function that is outside their control may threaten or undermine their control over corporate messaging and agenda setting. In particular corporate policy formation frequently elicits management pushback and calls on the independent evaluation function to demonstrate institutional loyalty. This is frequently combined with the argument that overemphasis on accountability inhibits learning, i.e. that a reporting relationship to senior management is the most effective way for evaluation to facilitate internal adjustments.

There is no common view on evaluation for accountability purposes. Everybody sees accountability as applicable to others than themselves. Likewise, among the evaluation models used by agencies – whether handled by independent offices or outsourced for credibility – the appetite for addressing accountability for results varies. Equally in the global South where there is strong growth in evaluation the emphasis lies in transparency that makes public performance visible but often with insufficiently robust evidence. But progress is in the right direction and there is greater understanding across the globe today about the role that evaluation serves in democratic settings that demand accountability and give voice to citizens.

This is confirmed by a rising demand for evaluation policies that contribute to accountable institutions and thus help to build trust and credibility among citizens. The Bangkok Conference co-hosted with the International Development Evaluation Association (IDEAS) in October, and leading to the Bangkok Declaration, makes clear the critical importance of independence, a prerequisite of credible evaluative judgment.

Throughout the developing world the "value for money" proposition is widely held and assessing whether citizens receive appropriate services, within an ethos of democracy and accountability, is the mainstream perspective for evaluation professionals. The change in the geographical and sectoral representation for such premier training programs as the International Program for Development Evaluation Training (IPDET) highlights this evolution in demand. The overarching perspective is a clear-cut notion that evaluation is about program improvement with citizens in mind. Recognizing that there are vested interests, evaluators must be protected from pressure and retaliation so that they can talk truth to power.

Much discourse on "what evaluation is" has been framed by management. For most managers evaluation is a management tool used by evaluators compliant to the leadership. When an internal evaluation unit challenges management the frequent response is to curtail the scope of the evaluation function, to control evaluation budgets, to undermine evaluation processes and/or to denigrate the methodologies of critical evaluations. Other means of control include selection and tenure of the evaluation head, a pervasive threat to evaluation independence.

I have heard about the "what about learning" argument, as if learning is more important than (or conflicts with) accountability. This suggests an internal focus that lies beyond public scrutiny. Some evaluation units go along with this narrative and see themselves as management consultants. They may respond to legitimate organizational needs, but such units should then be clearly labeled self evaluation units geared to facilitate internal learning with the implication that accountability is of lesser significance. The reporting line in such cases is to management, with evaluators having no final say on content or methodology.

This distinction between internal evaluation offices and independent ones is therefore fundamental and it is not surprising that those offices that are in fact independent have had to append to their names the term "independent" – in patent recognition that all evaluation is not so and cannot be assumed to be. The argument that work on methodology will resolve the issue of credibility masks the fundamental question of what purpose evaluation serves and glosses over the power dynamics embedded in all evaluation processes.

■

Jacques Toulemonde

Recently I was invited to speak in a conference on evaluation use[1] and I took that opportunity to further my reflections on the two main intended uses of evaluation: accountability and learning. I focused on development assistance, a domain where evaluation is a 50 year old industry. I also focused on evaluation use at the higher level, i.e. donor agencies' top management and financial constituencies.

Many players in the development evaluation industry share the view that accountability and learning are two sides of the same coin (ADB, 2014). Indeed there are nice examples of that two-sided coin. For instance a series of evaluations of the Budget Support instrument in 2010–2012 (IOB, 2012) answered key efficiency questions that had been repeatedly asked in the European Parliament (accountability). At the same time these works enabled donor agencies to understand where the instrument works and does not work and to act upon that lesson (learning).

What do we mean by accountability? Traditionally, donor agencies have been held accountable for compliance to budgets and rules. This means that funding constituencies look into management weaknesses with the support of auditors and question top managers about their responsibilities in case of dysfunction . Over the last decades, agencies have been held accountable also for their performance, performance being increasingly understood as the achievement of intended development outcomes. In that context, funding constituencies are now provided with annual performance reports and dozens of evaluations. However I have seen very little evidence of funding constituencies relying on evaluators for questioning top managers about their responsibilities in case of weak outcomes.

In practice, it is quite difficult for evaluators to deliver sharp assessments and to point out responsibilities for poor outcomes as auditors do in case of weak management. This is due to the fact that result information often stems from disputable cause-and-effect claims and responsibility for results is always shared among several institutions and stakeholders. Consequently, evaluators supply result information to financial constituencies but not in the form which would enable them to use it in a real accountability process.

In my professional experience, I have seen that most evaluation activities are shaped as if accountability was a priority, even where the commissioning body is primarily willing to learn. How can we recognize accountability oriented evaluations? They tend to focus on major expenditures, follow programming cycles, cover a wide scope, include the whole range of evaluation criteria, favor internal validity, focus on potential risks and problems, and highlight independence. Most often, such choices are made in an implicit manner in order to comply with professional standards that are accountability oriented, especially in the area of development assistance. Hence, my view of the broad picture is not that of a two-sided coin. It is a picture of accountability oriented evaluations that are not really keeping their promises in terms of accountability. Are they at least good at learning?

Let us first clarify what we mean by learning. The term refers to acquiring and using knowledge. In an evaluation context, we speak of providing decision makers with lessons on how to better achieve the desired outcomes, to avoid undesirable effects, and to set relevant priorities. Newly acquired pieces of knowledge may be called 'lessons' but lessons are seldom highlighted as such in

evaluation reports. They may be retrieved by reading recommendations although most of current evaluators' recommendations do not build on fresh knowledge. Moreover lessons need to be absorbed and acted upon before we can speak of learning.

Of course, we should not be naïve about policy-makers' willingness to learn lessons from evaluations. Uptake is not a linear process. It requires a long chain of events of which one or another is often missing, something which makes the success of learning unpredictable (Bossuyt & al., 2014). Precisely because of that risky game, evaluations should be optimized for learning if they are meant to feed knowledge back in the policy-making process.

Ten years ago, a Washington based working group[2] argued that large resources were spent on development assistance evaluations but not in the kind of evaluations that could help learning fast on big questions. They added that even evaluations addressing big questions were not delivering valid findings because they were not relying on a rigorous counterfactual. I disagree with the latter point because evaluators have a wide choice of rigorous causality analysis approaches, of which counterfactual analysis is not always the best choice and methodological rigor is just one among many ingredients in the learning recipes.

If I quote the working group, this is because its members initiated a stream of evaluations designed for learning. They wanted to implement their favorite impact analysis approach but unintentionally, they also revealed some of the features of learning-oriented evaluations. How can we recognize such evaluations? They tend to focus on challenging knowledge gaps, target policy-making win-

---

1 The conference was jointly organized by the Organization for Economic Cooperation and Development (OECD), the United Nations Educational, Scientific and Cultural Organization (UNESCO), France's evaluation society (SFE) and the European Evaluation Society (EES). It took place in Paris on September 30, 2015.
2 Savedoff, W.D et al., 2006.

dows, cover a narrow scope, use a few evaluation criteria or just one, seek generalisable lessons, explain the reasons of successes as well as failures, and establish close links with field level agents. Then we realize that an evaluation optimized for learning is not optimized for accountability and conversely[3].

I recognize that accountability oriented evaluations can teach useful lessons but, due to their very design, they tend to deliver a larger number of weaker lessons than learning oriented evaluations would do. Moreover lessons are delivered in a less timely manner, something which makes the overall learning process slower.

At this point, we could suggest that agencies' evaluation work plans include some accountability oriented works besides other ones geared at learning. My preferred option is another one. It would be to reconcile both types of use by referring to the promising idea of accountability for learning (Gray, C. et al., 2014). This would be a third approach to the concept after "accountability for compliance" and "accountability for results" (Perrin et al., 2007). Following that approach, donor agencies would be called to report every year on their learning accomplishments, i.e. whether they have identified challenging knowledge gaps, teamed up with partners in order to bridge these gaps, asked and answered specific evaluation questions related to these gaps, and acted upon the acquired knowledge. This could be done by both reporting on appropriate indicators and writ-

> „At this point, we could suggest that agencies' evaluation work plans include some accountability oriented works besides other ones geared at learning. My preferred option is another one. It would be to reconcile both types of use by referring to the promising idea of accountability for learning."

ing narratives of the most significant learning stories.

I have said that it is difficult to hold donor agencies responsible for their outcomes since causal claims and the sharing of responsibilities are often disputable. On the contrary, the good or poor learning records of an agency would be easy to ascertain. Properly trained auditors could check the learning claims and agency managers could be held responsible for their learning accomplishments, and this would create a very strong incentive for learning, something which is too often missing at present. In that new context, evaluation might serve both

learning and accountability (please consider the reverse order of the terms!) more successfully than it does at present.

References

Bossuyt, J., Shaxson, L., Datta, A. (2014). Study on the uptake of learning from Europe-Aid's strategic evaluations into development policy and practice. In *Final Report – June 2014*. Development and Cooperation EuropeAid – European Commission.

Dijkstra, G., Oukemp, A., Bergkamp, D. (2012). *Budget support: Conditional Results – Review of an instrument.* IOB Evaluation, Ministry of Foreign Affairs, Netherlands.

Gray, C. et al. (2014). Accountability and Learning: Two Sides of the Same Coin. In *Evaluation for Better Results* (pp. 47–69). Asian Development Bank.

Perrin, B., Bemelmans-Videc, M.L., Lonsdale, J. (2007). How Evaluation and Auditing can Help Bring Accountability into the Twenty-First Century. In *Making Accountability Work: Dilemmas for Evaluation and for Audit*, (pp. 239–264).

Savedoff, W.D. et al., (2006). When Will We Ever Learn? Improving Lives Through Impact Evaluation. In *Report of the Evaluation Gap Working Group*, Centre for Global Development.

■

---

3 Gray, C. et al., 2014.

THE AUTHORS

Geeta Batra

Geeta Batra is Chief Evaluation Officer and Deputy Director at the Independent Evaluation Office of the Global Environment Facility, where she provides direction and supervision to IEO's evaluations. Prior to joining the GEF IEO in June 2015, she was Manager of Country and Corporate Evaluations in the Independent Evaluation Group (IEG) of the World Bank Group. Prior to joining IEG, Geeta spent several years with the International Finance Corporation (IFC), notably as the Global Head of Results Management for IFC's Advisory Services. In all, she has been with the World Bank Group for about 17 years. Geeta holds a PhD in economics and has a strong background in international economics, econometrics, and development. In addition to evaluations, she has operational experience both in the Bank and in the private sector.

James Copestake

James Copestake is Professor of International Development at the University of Bath, and director of studies for the University's professional doctorate in policy research and practice. He has carried out research, supervised doctoral students and published widely on rural development, development finance, the conceptualisation of human wellbeing and political economy of development. His evaluation experience has focused particularly on microfinance.

Ian C. Davies

Ian Davies is a Credentialed Evaluator (CE) and former President of the European Evaluation Society. He has been a practising evaluator for over 30 years with professional experience in Europe, the Americas, Asia and Africa, a member and board member of professional evaluation organisations, i.e. EES, IDEAS, SFE, AEA, CES, and a regular presenter at annual evaluation conferences and professional events.

Ian joined EES in 1996 and presented at the second annual conference in Stockholm in March 1997. He heads up an international consultancy practice in organisational management, corporate governance, audit, review and evaluation.

Caroline Heider

Caroline has more than 25 years of international experience, the majority of which in evaluation. Prior to the World Bank Group, she has worked with five multilateral organizations, including two international finance institutions (the Asian Development Bank and International Fund for Agriculture Development), a technical agency (UNIDO) and two Funds and Programmes of the UN System (UNDP and WFP). She is a life-time member of the International Development Evaluation Association (IDEAS). She has been a member of the American Evaluation Association, the Australasian Evaluation Society and the European Evaluation Society. She served a 2-year term as vice-chair of the UN Evaluation Group.
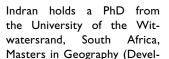
Mark Matthews

Mark Matthews is a strategy advisor to governments, universities and non-profit organisations. His remit spans diagnostic assessments of organisational capability, methodological innovation, executive education and top-level executive search. He is the author of Transformational public policy: a new strategy for coping with uncertainty and risk to be published by Routledge in July 2016.

Indran Naidoo

Indran holds a PhD from the University of the Witwatersrand, South Africa, Masters in Geography (Development Studies), West Virginia University (USA), B.Ed. in Curriculum Development (UNISA) and a B.A. (Honors) in Geography and B. Paedagogics from the University of

KwaZulu Natal. He has received scholarships to participate in Executive Management courses in Education at Oxford University and the Senior Executive Programme at the Harvard and Wits Business Schools. He has published and presented widely on development, public administration and evaluation topics nationally and internationally, and has received several leadership and academic scholarships and awards.

Riitta Oksanen

EES President, is a senior advisor on development evaluation in the Ministry for Foreign Affairs of Finland. She previously held the post of director for development policy in the Ministry and recently chaired the OECD/DAC Evaluation Network task team on evaluation capacity development. She has also acted as advisor on management and effectiveness of development cooperation; served 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 development cooperation planning, management and evaluation. She is a University of Helsinki graduate specialized in marketing, business administration and economics with emphasis on the forestry sector.

Robert Picciotto

Robert ('Bob') Picciotto, AcSS, is Visiting Professor, King's College (London) where he teaches evaluation. He holds graduate degrees in public and international affairs (Princeton University, USA) and aeronautical engineering (ENSA, France). He serves on the boards of the UK Evaluation Society and the European Evaluation Society. At the World Bank he was Director General of the Independent Evaluation Group from 1992 to 2002 after holding a variety of operational and senior management positions including Vice President, Corporate Plan-

ning and Budgeting and Director, Projects in three of the World Bank's Regions. Since his retirement from the World Bank he has provided senior independent evaluation advice to many international organizations, philanthropic foundations and think tanks.

Fiona Remnant

Fiona Remnant works in research, communications and brokering the gap between international development research and practice. She has an MSc in International Policy Analysis from the University of Bath. Before returning to the Centre for Development Studies at the University she spent four years working as a Senior Researcher for the Centre for Poverty Analysis in Sri Lanka, and worked for Oxfam Campaigns. Her current focus is promoting wider use of the QuIP among NGOs and impact investors.

Lawrence Salmen

Lawrence Salmen has been a manager in a number of development organizations. After receiving his MBA and Ph.D in urban planning at Columbia university, Mr. Salmen served as Director of housing and urban development at ACCION, Director of research and evaluation at the Inter-American Foundation, and Vice president of the Cooperative housing oundation. At the World Bank, Mr. Salmen was Lead social development specialist and was appointed by president James Wolfensohn to lead the task force, "becoming a listening Bank". He is the author of Listen to the People (Oxford, 1987) and, with Eileen Kane, Bridging Diversity (World Bank, 2006).
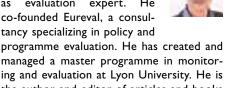
Marco Segone

Marco Segone is Director, Independent Evaluation Office, at UN Women; Chair, United Nations Evaluation Group (UNEG), the network of Evaluation Offices of 46 UN agencies; and co-founder of Eval-

Partners, the global partnership for national evaluation capacities. Previously, he was responsible for the decentralized evaluation function as well as the national evaluation capacity development portfolios at the UNICEF Evaluation Office; Regional Chief, Monitoring and Evaluation in the UNICEF Regional Office for Europe and Central Asia; Regional Office for Latin America and the Caribbean; Brazil Country Office, and Niger Country Office. Marco also worked in international NGOs in Albania, Pakistan, Bangladesh, Thailand and Uganda. He has authored numerous publications including Evaluation for Equitable Development Results and How to Design and Manage Equity-Focused Evaluations.

Jacques Toulemonde

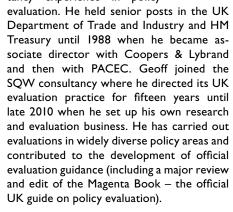Jacques Toulemonde works as evaluation expert. He co-founded Eureval, a consultancy specializing in policy and programme evaluation. He has created and managed a master programme in monitoring and evaluation at Lyon University. He is the author and editor of articles and books dealing with quality and professionalism in the field of evaluation, evaluation techniques and causality analysis, performance monitoring, evaluation in partnership, and evaluation cultures.

Geoff White

Geoff White is a professional economist with substantial practical policy and consultancy experience in policy evaluation. He held senior posts in the UK Department of Trade and Industry and HM Treasury until 1988 when he became associate director with Coopers & Lybrand and then with PACEC. Geoff joined the SQW consultancy where he directed its UK evaluation practice for fifteen years until late 2010 when he set up his own research and evaluation business. He has carried out evaluations in widely diverse policy areas and contributed to the development of official evaluation guidance (including a major review and edit of the Magenta Book – the official UK guide on policy evaluation).

Nick York

Mr. Nick York is Director of the Evaluation Department for Human Development and Economic Management at The Independent Evaluation Group of the World Bank since 2013. He oversees independent evaluations of country programs and economic management as well as corporate evaluations and global partnerships. Prior to this shift to Washington DC, Mr. York was Chief Professional Officer for evaluation in DFID in the UK. He joined DFID in 2004 working on country program evaluations. His team was responsible for independent evaluation of DFID's major policies and programs and developing the evaluation function in DFID including support on capacity and quality. In 2006/7 Nick chaired the steering group for the multi-donor evaluation of budget support. He also chaired the DAC evaluation network and helped to set up 3ie. Nick's background and training in economics includes advising on health systems and quality for 10 years for the UK government and he has also worked as a macroeconomist in the UK.

GUIDANCE
TO CONTRIBUTORS

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

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

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

To facilitate copy editing, authors are encouraged to use end notes rather than footnotes and to use the APA style guide for references. Here are some examples:

For books: Bergmann, I. (1997). Attention deficit disorder. In *The new Encyclopedia Britannica* (Vol. 26, pp. 501–508). Chicago, IL: Encyclopedia Britannica.

For journal articles: Rindermann, H., & Ceci, S. J. (2009). Educational policy and country outcomes in international cognitive competence studies. *Perspectives on Psychological Science*, 4(6), 551–568. doi:10.111/j1745-6924. 2009.01165.x

Website: United States Environmental Protection Agency. (2007, May 4). Climate Change. Retrieved 12 June 2014 from the Environmental Protection Agency website: http://www.epa.gov/climatechange.

In-text reference: (United States Environmental, 2007).

www.ees2016.eu