

# EVALUATION CONNECTIONS

## CONTENT

Presidential Message	1
Ethics as the last evaluation frontier: an editorial	2
Democratic Evaluation and Equity Focused Evaluation: Cousins but not Siblings	3
The weight of evidence in systematic reviews: appraising research quality	4
Principal Components Analysis of Evaluability Assessments: An Empirical Study	6
The EvalQuality Meta-Evaluation Dashboard	8
Accounting for Women's Empowerment: Meta-Evaluations from a Gender and Equity Lens	11
Paths to effective evaluation for interventions on violence against women and girls (VAWG)	12
Sharing lessons learned in SME Assistance in Africa	13
Evaluation briefs	15
News Flashes	16
Letter to the Guest Editor of the Special Issue on Private Sector Evaluation	17
The authors	19

## Presidential Message

Dear members and colleagues,

I have the pleasure to share with you insights and outcomes from our 11th Biennial Conference in Dublin "Evaluation for an Equitable Society. Independence, Partnership, Participation". The Conference attracted a record number of delegates and it was responsive to a wide range of concerns within the evaluation community as reflected by the diversity and wealth of participants' contributions. The notion of "equity" was paramount and it was debated throughout the event, starting with the excellent opening speech of the Irish Minister of Public Expenditure and Reform followed by inspiring keynote addresses that you can access on the Conference Website (<http://ees2014.eu/keynote-video-presentations.htm>)

Let me cite a few facts and figures: we welcomed 714 delegates from all over the world. We hosted 151 sessions (62 panels, 89 paper presentations) and we displayed 16 posters; The pre-conference consisted of 14 workshops. I wish to thank our abstract reviewers for the excellent work they performed under great time pressure: out of 562 abstracts received, 361 were accepted, and out of 69 proposals for training at the pre-conference, 14 were selected. This points to a rigorous selection and reflects our never-ending pursuit of excellence.

Awards ceremony has now become an EES conference tradition: five awards were bestowed -three for the best evaluation papers, one for the best student paper, and one for the best poster. Thirty-three bursaries were granted to enable evaluators from Africa, Asia and South America to attend the conference; we are most grateful to our generous sponsors (Minister of Foreign Affairs from Finland, DFID, UNDP and the Swiss Agency for Development and Cooperation).

In addition to the paper presentations and panels, special sessions dedicated to our Thematic Working Groups were held. They gathered TWG members and non-members, and led to lively discussions and networking around Professionalization, Gender, Private/Public sector, Sustainable Development. Two new TWGs emerged: Education and Training, and New Evaluators. Should you be interested in participating in any of these TWGs (or in several of them), you are most welcome to become a member by writing to our Secretariat: [secretariat@europeanevaluation.org](mailto:secretariat@europeanevaluation.org) or by using the online Join the TWG form on the EES website.

Numerous were the activities carried out with EES partners such as the National Evaluation Societies, EvalPartner, IOCE and many other organisations, mostly oriented towards the preparation of the Evaluation

Year 2015 thereby setting the stage for an active EES participation in this global evaluation action. Also worth mentioning is the one-day meeting of Parliamentarians, a first time event in an EES conference; their discussions focused on the participation of parliamentarians in evaluation to be celebrated during the International Year of Evaluation 2015 and to culminate in the establishment of a Global Parliamentarians Forum for Development

Evaluation. This is a step forward in the internationalization of democratic evaluation and the advent of more equitable societies.

This newsletter highlights some of the issues fervently debated at the conference such as equity-focused evaluation, democratic evaluation, trans-disciplinary evaluation, formative versus summative evaluation, as well as the rights and obligations of evaluators.

And building on the exchanges and connections forged at the 11<sup>th</sup> Biennial Conference let us now get ready for 2015 Evaluation Year and work together towards further initiatives in Europe and beyond that will make evaluation contribute to a more equitable and sustainable society.

Claudine Voyadzis, EES President

## ETHICS AS THE LAST EVALUATION FRONTIER: AN EDITORIAL

Robert Picciotto

In the leading article of this Newsletter Saville Kushner asks Connections readers: *who are we?* Since “we are what we do” the answer hinges on the meaning of evaluation as practiced in the real world. This issue of Connections confirms that, as a knowledge occupation, evaluation has three identities: it is a trans-discipline as well as a discipline and a specialized practice.

Sebastian Lemire’s article illustrates the trans-disciplinary feature. It shows how evaluative thinking can help improve the merit, worth and value of systematic reviews in social science research. But arguably, evaluation is also a discipline in its own right endowed with its own distinctive language; its own set of tools and its own standards.

Thus Timothy Guetterman and Joseph Hare’s offering demonstrates how, in the right hands, expert use of evaluation tools (such as principal components analysis) can help guide the design and conduct of “evaluability assessments” In the same vein Sara Vaca’s Meta-Evaluation Dashboard captures what makes evaluation what it is by unpacking the quality characteristics of evaluation products.

As a practice, evaluation is an analytical process that generates valid and useful information. It contributes to social learning by linking knowledge creation with decision making. Three articles in this issue demon-

strate how evaluation can be formative as well as summative by illuminating programme performance in a wide variety of settings. From a gender equity perspective, Ratna M. Sudarshan, Ranjani K. Murthy and Shradda Chigateri used meta-evaluation to assess the effectiveness of Government of India’s far flung poverty reduction and social inclusion flagship programmes.

Similarly a deliberate focus on results characterizes Michaela Raab and Wolfgang Stuppert’s qualitative comparative analysis of evaluation approaches and methods for interventions designed to reduce violence against women and girls. Finally Mohamed Manai, Elsa de Moraes Sarmento and Khaled Hussein Samir draw rich operational lessons from experience gained in providing assistance to small and medium industries in Africa.

The authors of all three articles and of most evaluations listed in the Evaluation Briefs section conceive of evaluation not simply as a way of enhancing understanding but as an intervention that ought to influence outcomes. They do not wish to be restricted to the narrow confines of *clientism*. Evidently they do not subscribe to the notion that evaluation is simply a service industry dedicated to the satisfaction of individual clients akin to management consulting. Hence the “Kushner question” (what ethical stance evaluators should adopt and what evaluation model they should embrace) is highly relevant.

Michael Scriven’s tackled this very question in his 11<sup>th</sup> EES Biennial Conference keynote address. He made clear that evaluators have selected an occupation that cannot escape the responsibility of taking an ethical stand and defending it based on scientifically acceptable evidence. By dint of impeccable logic and through compelling examples he demonstrated that “sitting on the fence” with no preference for any outcome or any goal is not consistent with evaluation professionalism.

To be sure the interface between advocacy and evaluation remains highly contested. But a deliberate orientation towards the public interest is a necessary prerequisite of professionalism. Without it evaluators do not deserve the franchise to practice in the public sphere. Accordingly altruistic principles are embedded in guidelines endorsed by all evaluation associations. Furthermore, without evaluation moral philosophy cannot qualify as an applied science.

From this perspective ethics and evaluation are siblings rather than distant cousins. Equally the right ethical dispositions (including independence of mind and appearance) form an integral component of evaluator competency frameworks – along with knowledge and skills. It follows that ethics, the last frontier of the evaluation world, should be carefully mapped and jealously protected. ■

## DEMOCRATIC EVALUATION AND EQUITY FOCUSED EVALUATION: COUSINS BUT NOT SIBLINGS

Saville Kushner

Given the constraints of working under contracts evaluators are, perhaps, a surprisingly aspirational group of practitioners. That small sliver of evaluation opportunity that is left to us after the behemoths of performance management, service inspection, low-trust accountability and results/impact measurement have passed through, is the space where we focus our desires to be democratic, affirmative, responsive, personalised, dialogic and all the rest. That space accounts for a fraction of the social resource spent on evaluation. What of it we can dedicate to promoting democracy, equity, fairness is just a fraction even of that. Our aspirations are built on the assumption of warrant arising from social and moral obligations, whereas reality for most is that our warrant derives from narrower contractual obligations. We define ourselves as a service, though many of these aspirations define us as moral interventionists. When Bob Stake, at the Dublin EES conference reminded us that ‘we do not know who we are’ – the most significant contribution to the conference – part, at least, of what he was referring to was the opaqueness of how we each strike a balance between service and intervention.

Perhaps the clearest expressions of the service-dominated approach come from Michael Scriven and Michael Patton, both of whom define the obligations of evaluators by the immediate needs of ‘end-users’, ‘consumers’ and ‘clients’. We see the lines more blurred in democratic and equity-focused approaches to evaluation which acknowledge ethical obligations from the political contexts within which the service is provided – often beyond the contract. The democratic evaluator seeks to provide a contractual service, but in a way that plays to broader politics of inclusivity and information rights.

Where do we sit in that confined space between our contractual obligations and our moral aspirations? I focus on three approaches, much discussed at the Dublin conference:

Democratic Evaluation (DE), Deliberative Democratic Evaluation (DDE), and Equity Focused Evaluation (EQ). Each seeks to make evaluation what it mostly is not and can only rarely be. But the airing of such evaluation dimensions is timely. Both democracy in the many forms we aspire to, and evaluation, as a discipline and practice that relies on it, are in decline. Following Stake’s unsettling observation, we need to know where we align our aspirations, for it turns out that these three approaches are quite distinct.

MacDonald laid out the rubric for Democratic Evaluation (DE) in 1974 (MacDonald, 1987), and House & Howe (1999) made a proposal for ‘deliberative democratic evaluation’ (DDE). In 2011 Segone laid out some indications of what he called an Equity Focused Evaluation – *“a judgement made of... policies, programs and projects concerned with achieving equitable development results...It provides assessments of what works and what does not work to reduce inequality...”*. This raises a number of important questions, such as: *what if our sponsors promote inequality? Since wealth and social advantage are relative, as we promote the interests of some in society (the vulnerable and excluded) whose interests are we threatening? Since inequality is an essential driving element of even liberal capitalist theory (i.e. meritocracy, utilitarianism) are we promoting anti-capitalism?*

These raise dilemmas that attend any advocacy evaluation and will be levelled at any evaluation that seeks to intervene *substantively*. Are we ready to pin our ideological colours to the mast? But how does this set of considerations relate to DE and DDE? Who are we?

DE, DDE and EQ assume an ideological comradeship that may suggest common political cause, but which is imprecise in representing how each plays out differently in practice. Here is a thumbnail of each to highlight the distinctions:

DE: The evaluator has no interest in program outcomes and is exclusively interested in the quality of understanding of the program, revealing the diverse aims and values embodied by it. The DE evaluator aims for impartiality (no preference for any idea or goal) and neutrality (all evaluands treated identically with no privileges). The evaluator has *“no concept of evaluation misuse”*, leaving the report and its implications for the democratic process outside the evaluation. The evaluator makes no judgements or recommendations which might favour one or another party and disfavour others. The DE evaluator would accept all contracts for evaluation on the grounds that the citizenry has the right to have subjected to their equal scrutiny whichever program or experiment it hosts. The aim of DE is to foster debate and argument as a sign of political wellbeing.

DDE: The evaluator takes a similar information brokering role as in DE but allows her impartiality and neutrality to erode in recognition of the special needs of disadvantaged or excluded groups, who might well receive privileges. In fact, the evaluator seeks positive discrimination to counter unfairness, giving privileged access to some. The evaluator, too, has a strong concept of evaluation ‘misuse’, since evaluation reporting is merely the precursor to a managed process of deliberation over program implications in which all are expected to play their part. Nonetheless, the evaluator retains her neutrality, so long as program outcomes do not prejudice the welfare of the worst off. The DDE evaluator has the aim of facilitating a process that leads to consensus.

EQ: Equity-oriented evaluators see evaluation as an intervention advancing the cause of social equity. Their driving vision of a more equitable society and the role evaluation might play in achieving that imply a great interest in program outcomes – and a discomfort with outcomes that do not move society in that direction. The EQ evaluator

would also privilege knowledge claims of the vulnerable and excluded. These evaluators would find it intolerable to evaluate a program that explicitly promoted inequity and would not accept such a contract. The aim of the EQ evaluator is to advocate greater justice and equity.

The primary difference between these methodological standpoints is the degree to which they are interventionist. Of course, they are all ethical interventions in that they all insist on a particular moral stance towards knowledge rights, often against the grain. *Procedurally*, they are all political interventions insofar as they insist upon equal or else equitable treatment for all against the normal rules of hierarchy. But *substantively* they vary from non-interventive (DE) to passive intervention (DDE) to purposeful intervention (EQ) in relation to social policy. DE has no interest in social outcomes;

DDE has no interest in outcomes other than that they must not diminish the position of the disadvantaged; EQ has specific outcomes in mind which entail favouring whatever definition of equity prevails in that setting.

Part of our context is an intensification of evaluation activity, a proliferation of evaluation associations and the emergence of an international evaluation bureaucracy (Eval-Partners) which now publishes global declarations and rules. For the most part these movements follow the liberal aspirational leanings that have characterised evaluation over the past 50 years – still ill-defined – but it remains the case that the weight of evaluation practices lie in illiberal and sometimes authoritarian approaches. We should remember that the program to dismantle welfare states would not be possible without the analyses and legitimations of evaluators.

2015 has been declared by some as ‘The Year of Evaluation’ – with no signal as to ‘the year of evaluation...what?’.

## References

- House, E., & Howe, K. R. (1999). *Values In Evaluation and Social Research*. Thousand Oaks: Sage
- MacDonald, B. (1987) Evaluation and the control of education. Reprinted from 1974 original in Murphy R, Torrance H (eds.) (1987). *Evaluating education: issues and methods*, London: Paul Chapman, pp. 36–48.
- Segone, M. (2011) ‘Evaluation to accelerate progress towards equity, social justice and human rights’ in Segone, M. (2011). *Evaluation for Equitable Development Results, Evaluation Working Papers #7*, NY: UNICEF, pp. 2–12.

## THE WEIGHT OF EVIDENCE IN SYSTEMATIC REVIEWS: APPRAISING RESEARCH QUALITY

Sebastian Lemire

The interest in systematic reviews is strong and still growing. This article examines a fundamental premise underlying systematic reviews, namely that the primary studies that they cover are of adequate quality (however defined). The premise seems self-evident: “if the ‘raw’ material is flawed, then the conclusions of systematic reviews cannot be trusted” (Jüni et al., 2001). However, a number of fundamental questions arise: How do we define research quality? How do we appraise the quality of research? And what do we do with the results? These are the questions addressed by this article.

The first question – how to define research quality – is both persistent and pesky. The only point of agreement among evaluators and social scientists is that research quality is multidimensional. However and despite fervent exchanges there is no shared understanding regarding what these multiple dimensions are, to what extent or how they are connected or if the dimensions are the same for different types of research. In this context the usual suspects of internal and

external validity (and occasionally construct and statistical validity) make an appearance in the quality “line up” but even these familiar faces are not identified in the same way by different reviewers.

The lack of an agreed-upon definition has not stopped social researchers and evaluators from appraising research quality. The number of frameworks and rating scales for doing so are in the hundreds (Moher and colleagues, 1995). They reflect differences in scope and content, complexity and degree of quantification, among other things (Jüni et al., 1999). Sadly, common denominators include the absence of strong theoretical grounding and thorough validations (Wells and Littell, 2009). As things stand, the use of a specific framework merely reflects the predilections of the individual researcher.

The lack of a shared understanding of research quality combined with a vast body of appraisal frameworks has implications for systematic reviews. To understand why that is we must consider the three most common

approaches for the use of quality appraisals in systematic reviews: the gatekeeper approach, the weighting approach, and the statistical analytical approach.

The *gatekeeper approach* involves the estimation of a single quality score and the exclusion of studies scoring lower than a pre-specified cut-off. The underlying rationale is captured by the old adage: “garbage in, garbage out”. While the approach seems intuitively reasonable it is problematic since the concept of quality merely reflects a researcher’s individual preferences and inclinations: what is one researcher’s ‘prize pig’ might just be another researcher’s ‘pulled pork’. To make matters worse the binary choice neglects the fact that studies are rarely good or bad: even excluded studies may have good and sometimes unique features. Thus the single-score approach may hide more than it reveals.

The *weighting approach* estimates a single quality score for each study and then uses the scores as weights when estimating the combined effect size. This is perhaps best

viewed as a nuanced version of the gatekeeper approach – especially since some studies are typically assigned such small weights that for all practical purposes they are excluded from the review. The rub here is that “the incorporation of quality scores as weights lacks statistical or empirical justification” (Jüni et al., 2001, p.45). More than that, ratings from different frameworks can produce dramatically different (even discordant) results when applied as weights (Jüni et al., 2001). As such, the estimated combined effect size is confounded by the researcher’s idiosyncratic selection of a framework.

Third is the *statistical analytical approach*, in which the estimated quality scores are used for testing the relationship between study quality and estimated effect sizes. This typically involves sensitivity and/or moderator analyses (Well & Littell, 2009). This approach is appealing because it treats the variation in research quality as an analytical opportunity – the aim of which is to understand how the variation influences the results of the review. More than that, the approach lends itself well to the use of sub-dimension scores (i.e., scores representing different dimensions of quality). Unfortunately, and despite the many merits of this approach, many systematic reviews have too few studies to support a valid statistical analysis.

This is admittedly a sobering state of play. What is to be done? First, we might allow the quality appraisals a less definitive role. This could involve moving away from the use of single scores towards the use of more nuanced descriptions of quality, for example “quality profiles” as recommended by Valentine and Cooper (2008). These profiles would provide a finer-grained understanding

of the strengths and weaknesses of primary studies.

Second, we might move away from the use of quantified appraisals altogether, focusing instead on developing qualitative statements about the diverse dimensions of studies. These statements could still be structured around an explicit framework. This would allow for a broader range of analytical opportunities. As just one example, we might rely on qualitative analytical techniques, such as Qualitative Comparative Analysis, for understanding how and why variation in research quality affects the results.

Thirdly, we should remind ourselves that validity relates to our inferences rather than to the property of specific design features (Shadish, Cook, and Campbell, 2002). Consequently, we should seek to link quality appraisals with specific inferences, identifying how and in what way specific quality dimensions impede the ability to answer certain questions or make inferences. For example, the observation of uneven attrition rates might impede our ability to draw causal conclusions about a program and a set of desired outcomes, but matter little in relation to claims about program implementation. By tailoring the appraisal in this way, we would be more specific about the ways in which different dimensions of research quality connect with our conclusions.

Fourth and finally, we might reconsider the very purpose of appraising research quality in the context of systematic reviews. What is the promise of measuring these variations? At root, we tend to think of variability as something to be controlled for – or weeded out. This is not surprising given the fervent

push for unfounded net-effect estimates. However, there is lot to learn from variations in quality and we should do more to engage and learn from this variation, rather than just weeding it out.

## References

- Jüni, P., Altman, D.G., & Egger, M. (2001). Assessing the quality of controlled clinical trials. *BMJ*, 323, p.42–46.
- Jüni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*, 282, p.1054–60.
- Moher, D., Jadad, A.R., Nichol, G., Pennman, M., Tugwell, P. & Walsh, S. (1995). Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials*, 16, p.62–73.
- Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Valentine, J.C. & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The study design and implementation assessment device (study DIAD). *Psychological Methods*, 13, p.130–149.
- Wells, K. & Littell, J.H. (2009). Study quality assessment in systematic reviews of research on intervention effects. *Research on Social Work Practice*, 19, p.52–62.

## PRINCIPAL COMPONENTS ANALYSIS OF EVALUABILITY ASSESSMENTS: AN EMPIRICAL STUDY

Timothy Guetterman and Joseph Hare

In the 1970s, evaluability assessment (EA) emerged as a pre-evaluation process to determine whether a program is ready for evaluation. EA's primary purpose was to ascertain the presence of measurable program objectives (Trevisan, 2007). These early frameworks provide a strong foundation but their further development is warranted for three reasons. First, evaluators conducting EA have to struggle with unclear, ambiguous methods (Smith, 2005). Second, the EA literature is largely conceptual given the paucity of empirical studies of evaluability. Third, existing EA frameworks focus mainly on reviews of program theory with little emphasis on two potentially important domains: organization support and data availability.

Our previous study of evaluability (Hare & Guetterman, 2014) suggested that these areas are important, interrelated components of EA. The premise for further exploring these domains is that organizational stakeholders must support the evaluation project to ensure it is pursued to completion and used and that sufficient information is available for

analysis. Accordingly we statistically tested the components of organizational support and data availability that affect program evaluability so that they may be included in EA.

Building on the results of our qualitative study and relevant literature (e.g., Wholey, 2011) we developed and administered a survey of Likert-type items related to evaluability to evaluators and program stakeholders involved in evaluation activities ( $n = 108$ ). After examining descriptive statistics we conducted a principal components analysis (PCA) of the respondent data to test the hypothesis that organizational support and data availability are primary considerations in evaluability assessments of programs.

The results revealed the shared items within each component that remained relatively independent from one another (Tabachnick & Fidell, 2013). These final items were then used to calculate a scale score using the mean across items, for each component. We first assessed the internal consistency reliability of the 33-item Organizational Support and

Data Availability Survey. Seven demographic items were excluded from the reliability analysis of four scales. The coefficient alpha of the four scales ranged from .78 to .81.

We then conducted PCA with direct oblimin oblique rotation on 26-items using listwise deletion (i.e. excluding from the analysis any item where a single value is missing) and allowing all remaining factors to be correlated. The instrument was constructed so that these items formed scales intended to measure different aspects of evaluability. Items focused on evaluation activities, evaluability considerations, the usefulness of evaluation, and the spontaneous use of data when making decisions.

The instrument yielded evidence of simple structure among four components: resources for evaluation, attention to the evaluation, use of evaluation findings, and impromptu use of data. The final structure was unexpected. Although we hypothesized items would relate to organizational support and data availability, the PCA results revealed a different structure. Items related to data availability

### Extended Evaluability Assessment Category

### Description

#### Program Considerations

- Program logic model

The structure of the program, the degree to which the program is likely to achieve clear goals, and the extent to which the goals might be credibly examined.

#### Program Context

- Past use of findings
- Frequent use of data

The organization's past experiences with evaluation and how useful evaluation findings were. The spontaneous use of data for decision-making.

#### Attention to Evaluation

- Leadership support
- Plans to use evaluation results

Engagement of champion with inter-departmental influence. Involvement of stakeholders from the appropriate functional areas with interest and influence. Plans to use evaluation findings.

#### Resources for Evaluation

- Time
- Material
- Funds
- Human resources
- Intrinsic quality of data
- Contextual data considerations

The funds, supplies, time, people, and space are adequate support the evaluation. Program staff and program participants are available. Material resources are allocated for evaluation. Intrinsic data quality includes accuracy, completeness, and consistency. Contextual data considerations include accessibility, relevance, and systems integration. The authority and ability data managers have to query and release organizational data. Degree to which culture supports data-driven decision making.

Table 1. Empirically Based Evaluability Assessment Considerations.



(e.g. “The data available in the organization is related to the evaluation questions”) correlated with the subset of items related to resources for evaluation.

The *attention to evaluation* component contained items related to leaders, and interestingly, plans to use evaluation results. This correlation of evaluation plans with leadership support suggests that plans to use results are related to organizational support for evaluation. In fact, the two most strongly endorsed items were related to leader support ( $M = 4.43$ ) and plans to use results ( $M = 4.11$ ). The importance of plans to use results is consistent with the existing literature concerning evaluability conditions.

Usefulness of evaluation findings and impromptu use of data also yielded a clear two component structure. With the *use of evaluation findings* component, we sought to clarify the forms of use that might influence the evaluability of a program. Respondents indicated evaluation results were useful or very useful to determine the merit of a program, improve a program or practice, influence thinking, and mobilize support ( $M = 3.91$ ). The *impromptu use of data* reflected a commitment to evidence-based decision making so that it may indirectly affect evaluability. Respondents indicated the *impromptu use of data* every few months to once per month

( $M = 3.67$ ) to determine the merit of a program, improve a program or practice, influence thinking, and mobilize support.

Integrating these results with the existing EA literature, Table I presents our proposed set of EA considerations. The program logic model and outcome measures remain critical to the assessment of evaluability. Program context and attention to evaluation provide specific criteria to determine if the evaluation milieu is supportive. Past stakeholders’ experience with evaluation may enable or inhibit evaluation. Assessment of the perceived usefulness of past evaluation findings, engagement of influential champions in the prospective evaluation, and plans to use findings reveal relevant elements of the environment in which the evaluation will occur. Adequate resources including funds, data, and time are practical necessities.

The results of this study have significant implications for evaluators. Programs and their evaluations can be expensive. Stakeholders desire timely information about the programs and their potential for evaluation. We offer empirically based considerations to determine the evaluability of a program. The specificity of resources, attention, and contextual considerations can assist evaluators in determining the potential for further more extensive evaluation. Finally, the nature

of these considerations may help an evaluator meet the information requirements of program stakeholders.

## References

- Hare, J., and Guetterman, T. Evaluability assessment: Clarifying organizational support and data availability. *Journal of MultiDisciplinary Evaluation*, 10, 9–25, 2014.
- Smith, M. F. Evaluability assessment. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 136–139). Thousand Oaks, CA: Sage. 2005.
- Tabachnick, B.G and Fidell, L.S. *Using multivariate statistics* (6<sup>th</sup> ed.). Boston, MA: Pearson, 2013.
- Trevisan, M.S. Evaluability assessment from 1986 to 2006. *American Journal of Evaluation*, 28, 290–303, 2007
- Wholey, J.S. Exploratory evaluation. In J.S. Wholey, H.P. Hatry, & K.E. Newcomer (Eds.), *Handbook of practical program evaluation* (3<sup>rd</sup> ed., pp. 81–99). San Francisco, CA: Jossey-Bass, 2011.

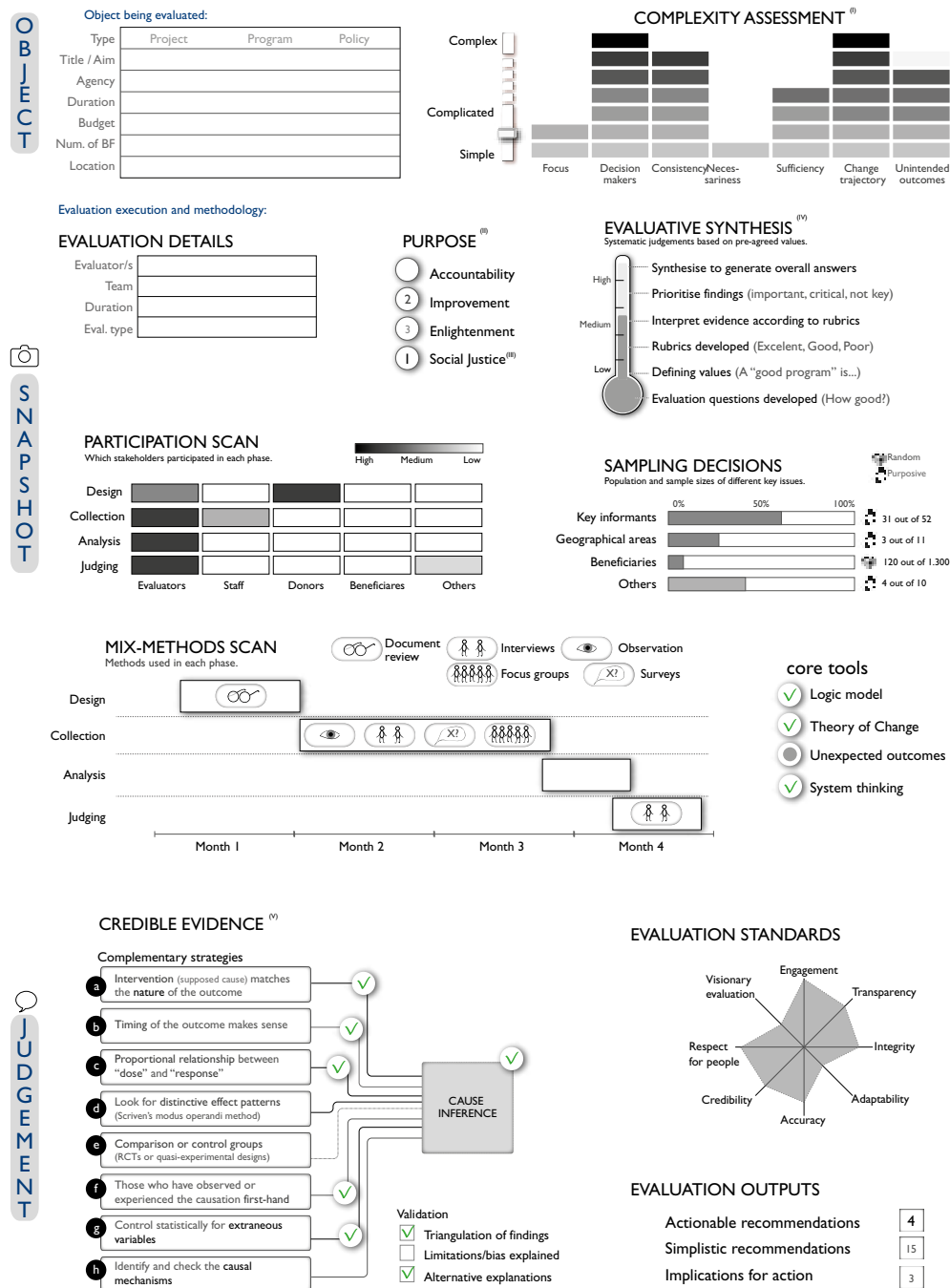


# THE EVALQUALITY META-EVALUATION DASHBOARD

Sara Vaca

The EvalQuality Dashboard summarises the most relevant aspects of an evaluation for quality assurance. It x-rays what happened during an evaluation, using the evaluation report as source. After a thorough analysis and a subsequent synthesis the most important elements are selected and visualised.

## META-EVALUATION DASHBOARD®



Sara Vaca (October 2014).



The essential features of the Dashboard are as follows:

### a) Information about the Object

In a succinct manner, assuming that the reader is (or has the means to be) familiar with the object, this index card is used to capture the object of the evaluation with cross-references to the terms of reference or the evaluation report.

**Object being evaluated:**

Type	Project	Program	Policy
Title / Aim			
Agency			
Duration			
Budget			
Num. of BF			
Location			

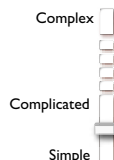
### b) Complexity Assessment

The complexity –or simplicity– of an intervention and its context is portrayed with seven variables, according to Patricia Rogers' *complexity breakdown* (2011):

- Focus
- Involvement of decision-makers
- Consistency of delivery
- Necessity of the intervention for achieving the intended impacts
- Sufficiency of the intervention for achieving the intended impacts
- Change trajectory
- Feasibility of identifying the unintended outcomes in advance or not.



By developing these seven criteria into rubrics one can readily depict the levels of each one visually, as in a frequency equaliser.



After this, the overall complexity level of the intervention and its context is “rated”; however, complicated and complex levels should not be included on the same axis, as they characterize different dimensions of the programme.

### c) Evaluation Purpose

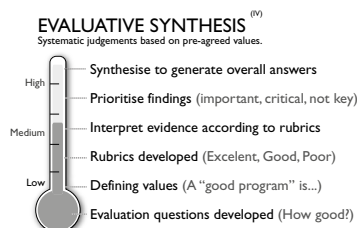
This reflects the ranking of purposes that motivated the evaluation, always in accordance with the evaluation report (or Terms of Reference). More than one purpose can be intended or declared. Methodological strategies and designs should be driven by the aims of the evaluation and the evaluation questions. This criterion enables an assessment of whether the proposed methodological strategy is coherent with the declared purpose/s of the evaluation.



### d) Evaluative Synthesis

A “unique” feature that differentiates an evaluation from a research study is that of Evaluative Synthesis (Davidson, 2014).

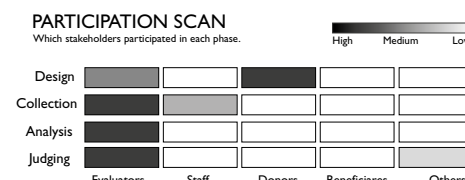
Many evaluations do this to some extent – by defining the evaluation criteria and questions – but higher levels of evaluative synthesis require defining values – what is “good”, “excellent” and “poor” – in each particular context, along with the evidence that demonstrates each element.



### e) Participation Scan

The Participation Scan maps the involvement of each of the evaluation's main stakeholders in each evaluation phase. Darker shades indicate higher levels of involvement and responsibility.

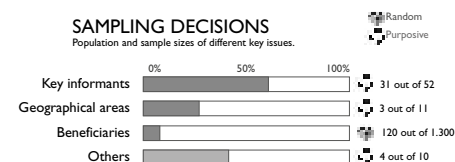
Mapping the main *phases* of any evaluation on the left axis and the most relevant *stakeholders* on the other results in a matrix through which participation can be easily examined. The result is a transparent way to map participation: it shows which stakeholders participated in each of the evaluation stages and to what extent.



### f) Sampling Decisions

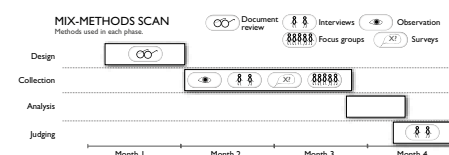
In most evaluation studies, decisions are made with regard to the sampling of potential informants. In estimating the number of potential sources, the number of each type who have been consulted by the evaluators is reflected. This is then represented as a %.

The Dashboard also shows whether the sampling was purposive or random.



### g) Mix-methods scan

The Mix-methods scan shows the techniques and methods used in each phase to assess how these techniques complement each other.

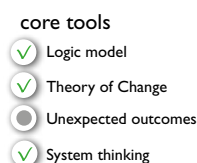


Phases are displayed in proportion to the entire timeframe of the evaluation.

## h) Core tools

In addition, the Dashboard shows whether the evaluators have included a logic model and/or a theory of change.

It also incorporates a verification of whether unintended outcomes were explored.



## k) Evaluation Outputs

Finally, most evaluations conclude with recommendations. These, however, vary enormously in number and quality. The Dashboard summarises how many recommendations (if any) were included and specifies whether they are actionable, elaborated, useful, insightful and inspirational or merely simplistic and symbolic.

### EVALUATION OUTPUTS

Actionable recommendations	4
Simplistic recommendations	15
Implications for action	3

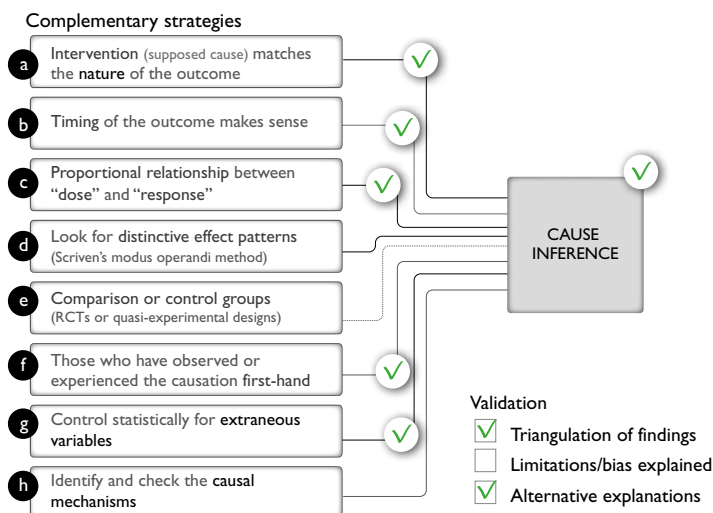
## Applications and potential uses

The Dashboard can be used to visualise the evaluation methodology of an evaluation report after its completion. It can also be used by evaluators to explain the rationale of the evaluation and its methodology. It is a tool for meta-evaluating and quality assurance, but it can also be used to visualise an evaluation design prior to its realisation. It can help demonstrate that the design proposed complies with the terms of reference. Alternatively it can be used to guide discussions of a proposed evaluation design with an evaluation commissioner and help explore various options that might be pursued. Finally it can help teach evaluation to beginners.

## i) Credible Evidence

For assessing the credibility of the evidence a mix of alternative strategies should be considered so as to be reasonably sure that the findings reflect reality (Davidson & Rogers, 2010). Not all are needed, but the mix should be complementary and should demonstrate causality or causal inference.

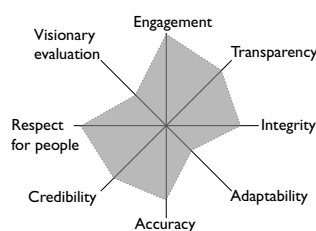
### CREDIBLE EVIDENCE <sup>(v)</sup>



## j) Evaluation Standards

A mix of standards defines key aspects that should be taken into account and, at the same time, can be easily checked as specific behaviours (in rubrics).

### EVALUATION STANDARDS



## References

Rogers, P. (2011) Program theory and logic models for systemic evaluation. International Conference on Systemic Approaches in Evaluation.

[http://www.evaluation-conference.de/downloads/63\\_CI\\_presentation.pdf](http://www.evaluation-conference.de/downloads/63_CI_presentation.pdf)

D.L. Stufflebeam, Anthony J. Shinkfield (1984). Systematic Evaluation: A Self-Instructional Guide to Theory and Practice (Evaluation in Education and Human Services).

Mertens, D. (2007) Transformative Paradigm: Mixed Methods and Social Justice. Journal of Mixed Methods Research 2007; 1; 212.

Davidson, J. (2014) It's the very core of evaluation and makes or breaks our work: so why is it in hardly anyone's toolkit? The EES Newsletter "Evaluation connections" – March 2014.

Davidson, J. & Rogers, P. (2010) Causal inference for program theory evaluation <http://genuineevaluation.com/causal-inference-for-program-theory-evaluation/>

## ACCOUNTING FOR WOMEN'S EMPOWERMENT: META-EVALUATIONS FROM A GENDER AND EQUITY LENS

Ratna M. Sudarshan, Ranjani K. Murthy, Shraddha Chigateri

In recent years, a large number of flagship programmes directed at reducing poverty and exclusion, and enhancing education, health and livelihood outcomes have been introduced by the Government of India. They are intended reach out to poor men and women. In most cases, quotas for women and designated social groups historically excluded from mainstream growth processes must be fulfilled and the programmes address issues faced by both genders although they also seek to tackle gender-specific constraints. The programmes are designed centrally and implemented through individual projects spread out across the country. They are evaluated on a regular basis by independent agencies. The evaluations focus on *implementation* and *effectiveness*.

No meta-evaluations seem to have been conducted and certainly none with a gender and equity lens. Little is known about the extent to which common themes or concerns have been probed. Nor have formal assessments of their scope and quality been carried out. To fill this gap a project on *'Engendering Policy through Evaluation'* supported by IDRC, Canada and the Ford Foundation, New Delhi and managed by the Institute of Social Studies Trust, New Delhi, has supported three meta evaluations of flagship programmes from a gender and equity lens.

At a panel session in Dublin at the 11<sup>th</sup> Biennial EES Conference we discussed, the frameworks developed for the meta-review of projects included in (i) a non-gender specific programme and (ii) a national women-focused programme. The idea of 'substantive equality' was incorporated into these frameworks an approach that yielded rich insights. The substantive equality goal recognizes that purposeful interventions are needed to tackle unequal initial conditions and create a level

playing field. Formal equality of access is inadequate to yield 'fair' outcomes. Further, it emerged upon interrogation that the term 'women's empowerment' frequently used in describing women-focused programme objectives had limited content in it. Specifically the level of women's participation and the possibility of generating additional income was often equated with 'empowerment'.

Ranjani Murthy evaluated 22 evaluations of the flagship Mahatma Gandhi National Rural Employment Guarantee Scheme, started in 2005. It guarantees 100 days of employment per rural household per year, with 33 percent of the work days generated reserved for women. She used the UN-System Wide Approach to Evaluation to assess the evaluation preparation, its methodology, its quality and its use. Another question about the relevance of evaluation findings was added.

The gender and equity analysis focused on the evaluation team, the identification of stakeholders, the evaluation criteria, the evaluation approach, questions, methods, indicators, data analyses, data validation, management responses and dissemination. Findings on impact were also reviewed from a substantive equality and empowerment lens. Of the 22 MGNREGA evaluations reviewed, 64 percent 'approached requirements' and 36 percent 'met requirements' as per the modified UN SWAP Evaluation Indicator. None of the evaluations 'exceeded' or 'missed' requirements. The availability of gender-expertise within the team was identified as a key factor in the meta-score awarded to the evaluations.

Shraddha Chigateri and Tanisha Jugran evaluated 20 evaluations of a large government programme (*Support for Training and Employment Programme or STEP*) that pro-

vides training targeted to poor, marginalized women in order to generate higher incomes as well as enhancing other dimensions of 'empowerment'. Attempting what is described as a 'formative meta evaluation' or one that would yield a framework for future evaluations, they used the OECD-DAC framework of relevance, effectiveness, efficiency, sustainability and impact. Questions on methodology and utilization of evaluation findings were added. Together, this enabled an assessment of whether the evaluations themselves contributed to the empowerment of women, and the indicators and criteria used to manage the programme. The meta-evaluation concluded that deeper engagement with the concept of 'empowerment' is critically important for the transformation of power relations, e.g. women's participation in decision making at home and in public fora.

A question that came up in discussion was whether it is 'fair' to assess evaluations from a gender perspective when the programmes themselves were only gender-responsive in a limited way. In response it was argued that doing so draws attention to issues that are not yet in the limelight of the policy discourse so that – as the project title suggests – it helps to engender policy through evaluation. Other methodological questions were raised. Would changing the weight given to different criteria alter the evaluation results substantially? Are some criteria for meta-evaluation more important than others, and if so should they be accorded greater weight? If the intention is to highlight gendered outcomes, should greater weight be given to the integration of gender into evaluation criteria? Further, should national level evaluations be accorded greater weight than provincial or district level evaluations? These issues deserve further analysis.

# PATHS TO EFFECTIVE EVALUATION FOR INTERVENTIONS ON VIOLENCE AGAINST WOMEN AND GIRLS (VAWG)<sup>1</sup>

Michaela Raab and Wolfgang Stuppert

In 2013–2014, the authors carried out a review of evaluation approaches and methods for interventions related to violence against women and girls (VAWG). A distinctive feature of the review was its assessment of the effects of the evaluations and to identify the configurations of factors that had made some of them effective. That is, rather than selecting a handful of evaluations meeting certain pre-determined criteria and analysing those in detail, our review examined a highly diverse set of 39 VAWG-related evaluations to identify drivers of evaluation effectiveness.

The review encompassed 74 English language evaluation reports of interventions on VAWG in development, humanitarian and (post-) conflict contexts. The evaluations, commissioned by a wide range of agencies, had been completed in 2008–2012. After a first analysis of the features of those reports, we ran a survey with evaluation stakeholders. We obtained responses from sufficiently diverse stakeholders for 39 evaluations. Data extracted from the reports and survey responses were used for Qualitative Comparative Analysis (QCA). Subsequently, five evaluations exemplifying different paths to effective evaluation were studied using Process Tracing<sup>2</sup>.

Our survey measured four types of evaluation effects, listed below, by triangulating the responses of different stakeholders.

- *Action effects*: the evaluation has helped strengthening or correcting the course of an intervention.
- *Persuasion effect*: the evaluation has influenced policy makers or donors.
- *Learning effect*: the evaluation has generated learning beyond those involved in the intervention.

- *Empowerment effect*: the evaluation has empowered the intended beneficiaries of the intervention. We considered any evaluation that caused harm to those beneficiaries to be ineffective.

Our initial literature review and dialogue with VAWG evaluation specialists has identified the following likely conditions for effective evaluation.

- *Evaluation context*, favourable if (i) the intervention is evaluable, and the evaluation is endowed with (ii) a stable political environment (in and beyond the intervention), (iii) a simple evaluation task, as well as (iv) a powerful mandate for the evaluation team (appropriate resources, timeliness, skills sets).
- *Evaluation approach*, i.e. quantitative, qualitative or a strong mix of both.
- *Evaluation quality*, which encompasses:
  - *Compelling evidence*, based on robust data and transparent documentation.
  - *Sensitivity to the gender-based violence (GBV) context*, in particular gender sensitivity and sensitivity to evaluation-related risks.
  - *Participatory design*, i.e. consultation with key stakeholders in the evaluated intervention (implementers, donors and intended beneficiaries) during the evaluation design phase and when findings are analysed.
  - *Good communication*, a result of accessible presentation and wide distribution of findings.

Our QCA identified eight different configurations of conditions that brought about four types of evaluation effects. The diagram below is a decision tree based on the *configurations of conditions or paths* that we have

found to lead to effective evaluation. Both *qualitative and quantitative* approaches, as well as mixed methods, have yielded effective evaluations.

All paths in the diagram work in a favourable *evaluation context*. In an unfavourable evaluation context, only the four paths in the lower half of the diagram have produced effective evaluations. This means that evaluation commissioners can choose from more options for effective evaluation by picking the right moment for the evaluation, asking straightforward evaluation questions and appointing a competent, well-resourced evaluation team.

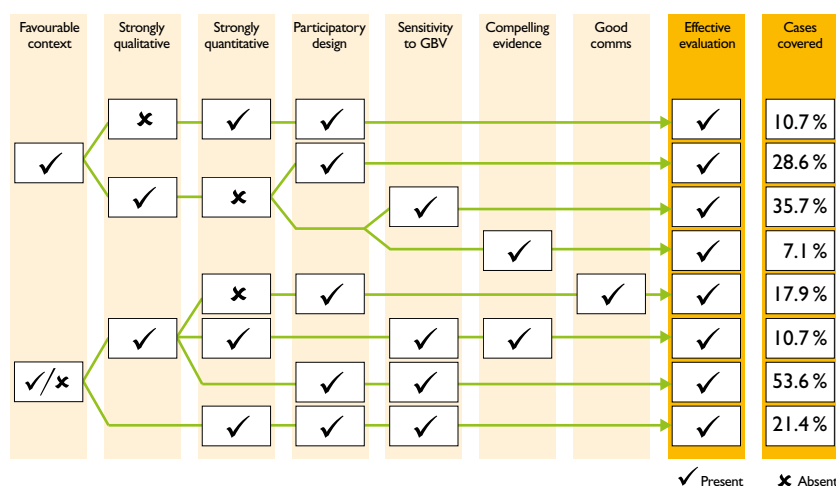
Every path includes elements of *evaluation quality* as defined above. That is, evaluations that do not fulfil any of those quality standards are not effective, regardless of their context. However, evaluations do not have to fulfil all quality standards to be effective.

*Participatory design* features in most paths: consultation with stakeholders enables evaluators to obtain the right data, interpret it correctly, produce appropriate recommendations and generate 'buy-in' among evaluation stakeholders. *Sensitivity to the Gender Based Violence (GBV) context* is a necessary condition in 4 out of 8 paths. Evaluation teams should be familiar with gender research and VAWG issues, and must observe ethical guidelines to avoid violating the rights of those potentially affected by the evaluation.

We found that evaluations can be influential even if data collection and sampling do not fulfil established standards of social research. That is acceptable in some cases, for instance if an evaluation serves to improve an intervention

<sup>1</sup> This article is based on the poster that attracted the "award for best poster" at the 11<sup>th</sup> Biennial EES Conference in Dublin.

<sup>2</sup> Detailed information on our methodology, including the scoping, inception and final review reports, as well as our dataset and references, can be downloaded from our dedicated blog [www.evawreview.de](http://www.evawreview.de). As of October 2014, the blog had attracted a world-wide readership (more than 5,500 page views from some 50 countries).



Paths to effective VAWG evaluation.

by facilitating joint reflection. Finally, transparent *documentation* of the evaluation process helps to assess the likely validity of evaluation findings. Whenever compelling evidence is important, evaluators should collect original data from different sources and avoid bias.

The diagram above can be used to verify whether an evaluation design combines conditions that have been found to make

evaluations of VAWG-related interventions effective. Users should start by assessing whether the evaluation context is favourable or not (left); subsequently move to the right to determine the approach (qualitative or quantitative); and then verify which aspects of evaluation quality are present.

The symbol ✓/✗ under “Favourable context” means that the paths which start from that

symbol lead to evaluation effectiveness, i.e. that it does not matter whether the context is favourable or not. “Comms” stands for communication of findings.

The first four rows are linked to paths that have worked in favourable contexts only. Apart from favourable context, only two extra conditions are required for those paths to lead to evaluation effectiveness: strongly quantitative and participatory design (1<sup>st</sup> row), strongly qualitative and participatory design (2<sup>nd</sup> row), strongly qualitative design and sensitivity to the GBV context (3<sup>rd</sup> row), or strongly qualitative design with compelling evidence (4<sup>th</sup> row).

Where the context is not necessarily favourable, more conditions are required for effective evaluation. Rows 5–8 show those options. In addition to examining the configurations shown above, commissioners and evaluators should discuss the quality of evidence that is needed for a particular evaluation. If highly robust data are needed, adequate time and resources should be provided for data collection and analysis in order to fulfil scientific standards.

## SHARING LESSONS LEARNED IN SME ASSISTANCE IN AFRICA<sup>1</sup>

Mohamed Manai, Elsa De Morais Sarmento and Khaled Hussein Samir

Sharing of lessons is intrinsic to the culture of a learning organization. Evaluation helps determine which interventions are most effective in helping SMEs tap the benefits of economic reforms. Evaluation also helps Development Finance Institutions (DFIs) become more effective in identifying Small and Medium Enterprise (SME) projects that generate positive development outcomes.

SMEs are varied and heterogeneous. Targeting the right ones is hard and costly without reliable knowledge about the factors that drive their success and sustainability. Improved access to finance is not enough. Diverse and innovative measures are also needed to

complement financial assistance and address specific constraints. This assessment should be carried out on a case by case basis. Equally, strengthening country and regional strategies implies familiarity with the legal and regulatory environment and calls for an accurate diagnostic of the obstacles and opportunities faced by the private sector. This is because providing the right incentives for SME development hinges on a sound enabling environment that rewards innovation and growth.

In turn, this requires continuous and coordinated efforts from three main sets of actors (i) DFIs, (ii) national and local authorities, and (iii) the international development com-

munity. Both demand and supply side factors have to be taken into consideration along the project cycle, in order to ensure coordination and exploit synergies. It is the optimal combination of these factors that triggers SME development.

On the supply side, evaluation lessons gained through evaluation were concentrated on the creation of the right enabling environment that fosters business creation and growth, through the proactive involvement of national governments and the development community. On the demand side, diverse findings about firm behaviour were collected so as to illuminate the drivers of ownership and par-

<sup>1</sup> This analysis do not necessarily reflect the views of the African Development Bank. The authors take full responsibility for any inaccuracy.

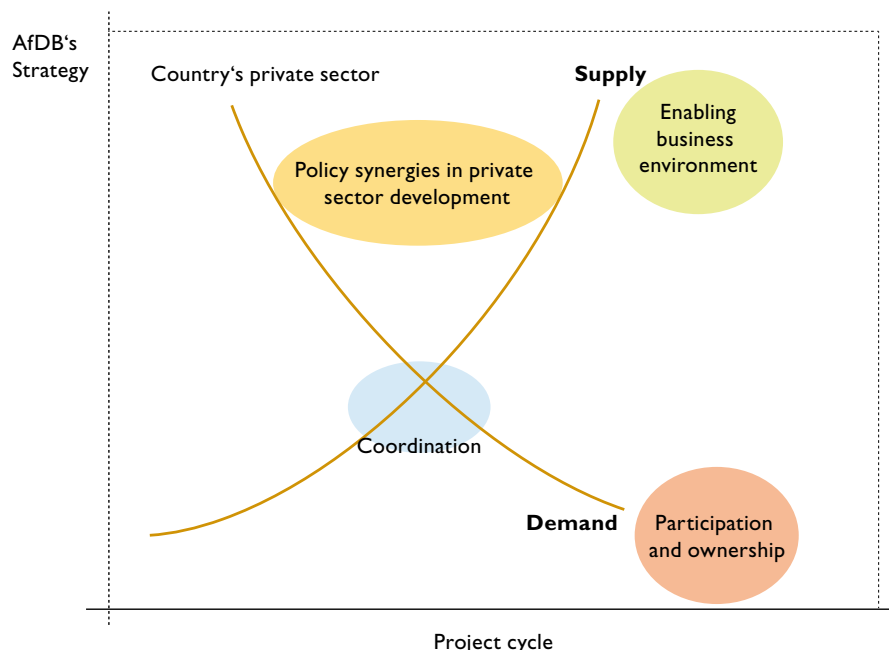


Figure 1: Interactions between identified demand and supply SME drivers.

Source: Manai, M., Sarmento, E. de Moraes and K. Hussein (2014, forthcoming), "Lessons learned: Small and Medium Enterprise Support", Independent Development Evaluation, African Development Bank.

Stakeholders	Axis	Lessons for:	Recommendations
Development community	Demand and supply interaction	Participation and ownership	Engage local actors
Country	Demand and supply interaction	Business environment	Promote a conducive institutional, business and investment environment
Country and development community	Demand and supply interaction	Private sector development Coordination	Coordinate and synergise across different levels of policies affecting SMEs
Project/ Program teams	Supply-driven	Project preparation, implementation, monitoring and evaluation	Strengthen SMEs capacities

Figure 2: Summary of lessons and recommendations.

Source: Manai, M., Sarmento, E. de Moraes and K. Hussein (2014, forthcoming), "Lessons learned: Small and Medium Enterprise Support", Independent Development Evaluation, African Development Bank.

ticipation that proved critical in influencing the development outcomes of SME support.

These lessons were drawn from the African Development Bank's (AfDB) Independent Development Evaluation (IDEV) "Evaluation Results Database", covering the period 1993 to 2013. They were extracted from 23 documents of 7 different typologies (Corporate Evaluations, Country Assistance Evaluations, Process Reviews, Project Completion Reports, Project Completion Report Review Notes, Project Performance Evaluation Reports, Thematic Reviews) featuring 22 countries (Algeria, Benin, Burkina Faso, Cameroon, Centrafrique, Chad, Dem Rep Congo, Gambia, Ghana, Lesotho, Malawi, Mali, Mauritania, Morocco, Mozambique, Nigeria, Sao Tome, Senegal, Tanzania, Tunisia, Uganda) and 13 sectors (Agriculture and Rural Development, Banking, Education, Electricity, Health, Poverty Alleviation, Micro-Finance, Social, Private Sector Management, Railways, Roads and Highways, Telecommunications, Water Supply and Sanitation) drawn at stages of the project cycle: Identification, Design, Stakeholder Consultations, Preparation/Feasibility, Implementation, Supervision, Completion and Post Evaluation.

At the project/program level, addressing the capacities of SMEs and their uptake potential is of vital importance for effective aid absorption. This usually entails training and related capacity-building approaches to fill domestic capacity gaps. Furthermore, capacity-building is also needed in the DFIs themselves as well as in the international agencies that fund them. They too need to commit themselves to learn from experience.

At the country level a set of critical enabling conditions facilitates private SME sector development. These include basic infrastructure provisions (e.g. roads, electricity, water and telecommunications), reducing bureaucratic costs involved in new venture formation and operation along with promoting a favourable business and investment environment climate that induces firms to step into the formal market. Supporting the development of sustainable financial institutions and markets, and improving access to medium and long term finance does help small firms grow in sales, revenues and operations. Fostering an entrepreneurial culture, mentoring of entrepreneurs, vocational



training and basic managerial training are also instrumental in generating results.

For the international development community, effective dialogue with state authorities and local actors is critical to generate participation, ownership and buy-in from an early stage. For effective SME support, key stakeholders need to be involved and engaged throughout the preparation, implementation and monitoring process, for timely feedback. In particular, active involvement is needed with the final beneficiaries.

Figure 2 summarizes lessons according to the stakeholders and axis and draws additional recommendations.

For both the development community and national stakeholders, evaluation lessons relate to the need for good coordination in support of a coherent private sector development strategy. Interventions need to synergise and be complementary to countries' broader national development and poverty reduction programmes. Effective support for all the policies and

programmes that affect the performance of private enterprises, promote access to social services and improve infrastructure is key to development effectiveness in SME support. There is also vast scope for better coordination efforts within the international development community including financial institutions in order to reduce overlap and increase complementarities. The prospects for broad based equitable economic development would be further enhanced if SME development would remain high on the global development agenda.

## EVALUATION BRIEFS

This new 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.

**Herwig A. Viechtbauer**

email: [herwig.viechtbauer@gmail.com](mailto:herwig.viechtbauer@gmail.com)

### **Evaluating Secondary Schooling from a capabilities' perspective: Tanzanian Students' Voices**

What is the value of secondary schooling to those who are supposed to benefit from it? Inspired by Amartya Sen's capability approach the author assesses the functional, intrinsic and social worth of education by listening to the voices of individual students within specific social settings in terms of their perceptions regarding the contribution of learning to their well-being and their agency fulfilment.

**Wayne Amago Bacale**

email: [wayne\\_bacale@yahoo.com](mailto:wayne_bacale@yahoo.com)

### **Planning, Monitoring and Evaluation and the Capacity Challenges of Nongovernment Organizations**

Despite the growing and impressive contributions that international nongovernment organizations contribute to society capacity constraints have hindered their planning, monitoring and evaluation functions. Overcoming this challenge would enhance their credibility, accountability and performance.

Towards this end, alternative competency assessment tools are proposed to identify competency gaps; design a capacity building curriculum; measure learning outcomes and ascertain impacts on the organization.

**Carlos Rodriguez-Ariza**

email: [crariza@hotmail.com](mailto:crariza@hotmail.com)

**and Rafael Monterde**

email: [rmonterd@gmail.com](mailto:rmonterd@gmail.com)

### **The use of evaluations in the aid sector: the Spanish Cooperation case**

In Spain, particularly in the aid sector, the number of commissioned evaluations has grown rapidly. However, most evaluations to-date have done little more than generate a false perception of accountability. In response the authors argue that actual use should be the standard by which evaluations are judged. Noting that the organizational context matters to utilization they suggest that building utilization focused quality assurance processes at the front of all evaluations would contribute to effective knowledge management and corporate learning.

**Junwen Luo**

email: [j.luo@utwente.nl](mailto:j.luo@utwente.nl)

**Gonzalo Ordóñez-Matamoras**

Associate Professor,

email: [h.g.ordonezmatamoras@utwente.nl](mailto:h.g.ordonezmatamoras@utwente.nl)

**Stefan Kuhlmann**

Professor, email: [s.kuhlmann@utwente.nl](mailto:s.kuhlmann@utwente.nl)

### **The Balancing Role of Evaluation into Organizational Governance:**

### **An International Comparison of Publicly Funded Research Institutions**

Publicly funded research institutions (PRI) are increasingly required to provide reliable evidence of their effectiveness. The evaluation mechanism works at three levels: central supervisory body, specific institutes, and individual researchers and involves internal and external stakeholders (government, scientific community, universities, industries, public etc.). How can diverse and often conflicting stakeholders' interests be taken into account? To address this question, the authors used a conceptual framework grounded in neo-institutionalism principles to illuminate the role of evaluation throughout the policy and implementation cycle in the Max Planck Society (MPG), the Helmholtz Association (HGF) and the Chinese Academy of Sciences (CAS).

**Kaisa Lähteenmäki-Smith**

email: [anna-kaisa.lahtenmaki-smith@vink.fi](mailto:anna-kaisa.lahtenmaki-smith@vink.fi)

**Pekka Pesonen**

email: [pekka.pesonen@tekes.fi](mailto:pekka.pesonen@tekes.fi)

**Timo Kolu**

email: [timo.kolu@aka.fi](mailto:timo.kolu@aka.fi)

**and Kalle A. Piirainen**

email: [kalpii@dtu.dk](mailto:kalpii@dtu.dk)

### **Emerging Dialogue and User-Based Modes of Evaluation for Innovation Policy**

The authors examine the implications of ongoing changes in public policy and evaluation culture associated with evidence-based deci-

sion-making. Based on a specific evaluation of the Strategic Centres for Science, Technology and Innovation undertaken in 2012–2013 they show that dialogue built into the evaluation process at multiple levels strengthened engagement of stakeholders in the evaluation process, raised acceptance of evaluation findings and increased the relevance, usefulness and impact of the evaluation results.

#### **Oluwole Akinagbe**

email: [wolexakins@yahoo.com](mailto:wolexakins@yahoo.com),  
[omakinnagbe@futa.edu.ng](mailto:omakinnagbe@futa.edu.ng)

#### **Evaluation of Constraints to Implementation and Adoption of Cocoa Resuscitation Programmes in Southwest Nigeria**

The production of cocoa, an important cash crop in Nigeria, has suffered a chronic decline. In 1999 the federal government launched a Cocoa Resuscitation Programme. Meanwhile, the increase in cocoa production had not been linear. This led to an evaluation study to help improve programme effectiveness. It focused on the constraints that militated against adoption of improved cocoa technologies by farmers. Interviews with farmers and extension staff as well as focus group discussions generated data that was fed into a principal factor model and helped identify the major obstacles to cocoa resuscitation programme.

#### **Anna Ciraso**

email: [annaciraso@gmail.com](mailto:annaciraso@gmail.com)

#### **An evaluation for the reorientation of a socio-educational programme against child poverty in Spain**

The CaixaProinfancia programme (CPI) provides educational and family support, access to leisure activities and health interventions to address child poverty in Spain. Twelve Spanish universities have teamed up to carry out an action oriented evaluation designed to improve programme design and implementation. The evaluation consists of a multi-site case study that will elicit the perceptions of stakeholders about programme impacts on children, families and communities. In a first phase focus groups and semi-structured interviews have been carried out among parents, NGOs workers and other stakeholders. The second stage will consist of a realist analysis of these data focused on similarities and discrepancies of outcomes across eleven distinct contexts.

## **NEWS FLASHES**

### **NEXT EES CONFERENCE IN MAASTRICHT**

**28–30 September 2016**



Photo: [wikimedia.org](https://commons.wikimedia.org/wiki/File:Maastricht_Binnenhof.jpg) / Author: Michielverbeek / CC BY-SA 3.0

**As announced in Dublin, the City of Maastricht has been selected to host the 2016 EES Conference after a call for candidates last summer, followed by a rigorous selection procedure. The conference will be held in MECC Maastricht ([www.mecc.nl](http://www.mecc.nl)) from 28 to 30 September, and will as usual be preceded by two days of preconference.**

Maastricht is a historical town for the Netherlands and Europe. It hosts around 150 nationalities. It is situated at the crossroads of different countries and its university – which expressed great interest in the conference – at the crossroads of many evaluation-related disciplines. The Province of Limburg is a regional institutional member of EES and very proud that EES has elected its capital city to host the 2016 conference. The Province hosts many different societal experiments involving stakeholders from public and private sectors, and civil society. Evaluation has a prominent role in those. Maastricht therefore provides excellent ground for evaluation and for our conference!

Preparations have started and the Board intends to announce the conference theme shortly. **We hope to welcome you all in 2016!! In the meantime we'll keep you posted!**

## **ROTATIONS ON THE EES BOARD**

Bastiaan de Laat, EES Board member and Secretary General, is leaving the board on January 1<sup>st</sup> 2015 after four years of superb service to your Society. Barbara Befani will step into his shoes as of 1 January 2015 following her unanimous appointment by the Board.

Ole Winckler Andersen (recently elected and endorsed by the AGM on 2 October 2014) will join the Board on January 1<sup>st</sup> 2015. He will be responsible for fund raising and for strengthening the Society's linkages with European institutions as well as with national, regional and global evaluation associations. We are delighted to have him join the team.



Bastiaan de Laat



Barbara Befani



Ole Winckler Andersen

## LETTER TO THE GUEST EDITOR OF THE SPECIAL ISSUE ON PRIVATE SECTOR EVALUATION

Dear Fredrik,

Private Sector Evaluation (PSE) has grown steadily but the *Connections Special Issue* illustrates how embryonic it is in terms of application even for International Financial Institutions using concessional finance. The growing emphasis on the roles of Financial Intermediaries, Small and Medium Enterprises and Corporate Social Responsibility (CSR) have fuelled demand for innovative practice in PSE.

However, practice is lagging behind the discourse so fundamentally that even basic tenets of monitoring and evaluation (e.g. valid counterfactuals; estimates of indirect impacts, activity based costing; non-financial measures of change; distinctive contributions of partners; and providing space to learn about how beneficiaries-cum-clients rate and respond to aid's performance) are dubbed as innovative!

The Special Issue does not support the assertion that there has been a remarkable advance in developing let alone delivering on good practice PSE standards. At the very least, the notion that new performance measurement systems and improved quality of social and environmental impact assessments have emerged – needs qualifying.

The fixation of public sector donors in knowing how many jobs have been 'created' reflects outdated politicians' perceptions of how taxpayers assess the worth of development aid. The Standard for Measuring Results in Private Sector Development put together

by the Donor Committee for Enterprise Development (DCED) sets 'net additional jobs created' as one of three universal indicators adopters of the standard are expected to measure, along with the number of target enterprises who realize a financial benefit and net additional income they accrue.<sup>1</sup>

Job creation is almost invariably associated with job destruction elsewhere in a competitive economy. How is this dimension assessed? The more useful and operative question is: how, how well and in what ways are donor programmes stimulating contributions to inclusive and sustainable economic growth in support of those who create the jobs, the private sector?

It is in this context that the challenges to M&E should focus on understanding how and in what ways can PSE focus on meaningful development challenges. Of particular note is helping private sector and civil society work with government to better diagnose, prioritise and make changes to the business environment and investment climate, and the consequences of these on the behaviours of the private sector and the government in following these changes through.<sup>2</sup> Often these are just related as risks and/or assumptions and are afforded little attention by Monitoring let alone Evaluation.

The relevant questions on accountability are first and foremost: who is demanding accountability; who is accountable to whom; and on what basis? It is certainly true that a large majority of the electorate in post-crisis OECD countries are questioning the

scale of development finance. However this demand is channelled from civil society towards the development sector through donor organisations as the interlocutor. These organisations have themselves, been historically lacking in transparency and capacity to demonstrate results.

The goal should be improved transparency through the entire development value chain, from the donors to the service providers and to the ultimate beneficiaries, be they poor people or those who are tasked to serve them. It is on this basis that politicians in northern countries should hold themselves to account through oversight and scrutiny of aid budgets, expenditures and results. All too often such processes play out in reverse with the shareholders (the donors) rather than the clients or the ultimate beneficiaries determining and measuring 'success'.

The fundamental assumption underlying universal (DCED) and global (IRIS) indicators is that the lack of metrics is a binding constraint.<sup>3</sup> This misses a basic point: context matters and indicators need to be specific to a given decision situation. Considerations of scale, aggregation, critical limits, and thresholds, etc. are situation dependent and the operating environment is complex and constantly changing. The current PSE indicator 'industry', is located in the 'north' and it is sustained by donors and academics whose interests it serves. It is time to ask who benefits from it and how.

**Daniel Ticehurst and Jonathan Mitchell**

<sup>1</sup> The DCED Standard for Measuring Results in Private Sector Development. Control and Compliance Points Version VI, January 2013

<sup>2</sup> <http://www.oecd.org/daf/inv/investment-policy/countryreviews.htm> for examples of country based reviews of government commitment to following through on various reforms aligned to the World Bank's Doing Business Indicators.

<sup>3</sup> <http://iris.thegiin.org/>

Fredrik Korfker *replies:*

The Connections special issue did not aim to present a full spectrum of private sector evaluation issues and views. It was a compilation of articles contributed by members of the Thematic Working Group (TWG) on Private Sector Evaluation (PSE) of the EES. They are experienced in this evaluation domain. Most of them are working or have worked in International Financial Institutions.

Considering that the Evaluation Cooperation Group (ECG) of the Multilateral Development Banks was established in the mid-nineties and taking account of the fact that it has issued good practice standards for private sector evaluation, one cannot call PSE embryonic. On the other hand, PSE within the private sector itself is just developing. The broader evaluative concerns associated with Corporate Social Responsibility, the links between social effects and share value, etc. are relatively new and still evolving. And while this movement is exciting and we know that the private sector is sophisticated in measuring what it cares for, one should recognize that the priority of private sector corporations in a market economy is and is likely to continue to focus on profitability and shareholder value.

The good practice standard guidance on private sector evaluation of the ECG, which can be down-loaded from the ECG website (ECGnet.org) presents a comprehensive evaluation framework in which estimates of direct and indirect impacts play an important

role. For example social and transition impacts form the core of non-financial measures of change. We fully agree that measuring such effects remains a challenge.

The good practice framework also deals with roles and contributions of all stakeholders and pays ample attention to the perspectives of ultimate beneficiaries. Such evaluative dimensions cannot be considered innovative; they are common practices in multilateral development banks. The websites of IFC and EBRD give ample evidence of these evaluation approaches in action.

We fully support the view that monitoring and evaluation (M&E) should take on broader development challenges, help all stakeholders understand results and contribute to the business environment and investment climate. In practice, measuring net additional jobs created can be an important performance criterion in PSE but institutions such as the IFC and the EBRD use indicators that go well beyond this, as well as beyond financial and business dimensions. Individual projects have to comply with a number of developmental and transition indicators. In principle all MDBs evaluators are aware of these requirements but the quality of evaluation practice varies. The ECG exchanges experiences among its members to make sure that one learns from the other.

The questions the authors of the letter pose on accountability and need for oversight

are spot-on. Understanding results on the ultimate beneficiaries is fundamental. In this sense, the project approach used more often by private sector-specialised MDBs puts the ultimate beneficiary at the heart of evaluation practice. However, estimating effects on the ground is not an easy task. For instance, there are areas such as SME interventions, where grasping the effects on the ultimate beneficiary is a major challenge, as indirect financing approaches through intermediary financial institutions are dominant. Scrutiny of aid budgets is important too but MDBs do not provide concessional financing to the private sector.

Finally we agree that context matters and that indicators need to be tailor made and specific. ECG has come a long way in addressing these issues. However, compliance with the ECG good practice standards varies among the MDBs so that PSE while not embryonic is a work-in-progress. The TWG on private sector evaluation of the EES seeks to advance understanding of the challenges faced in this field, and it was the purpose of the special issue to stimulate discussion. We look forward to engaging the authors of the letter in debate as future members of this Group.

**Fredrik Korfker**

TWG Head and Guest Editor of the Special Issue

## THE AUTHORS

### Tim Guetterman

is an applied research methodologist in the University of Nebraska-Lincoln's College of Education and Human Sciences Mixed Methods Academy. His professional interests, teaching, and research writings are in research methodology, namely mixed methods and general research design, particularly as applied to evaluation and assessment.



### Shraddha Chigateri

works as Research Fellow at the Institute of Social Studies Trust and leads the project on Engendering Policy through Evaluation. She completed her doctoral research on Dalit feminist politics in Bangalore in 2004 from the University of Warwick, UK after which she taught for a few years at the Universities of Warwick and Keele in the UK before joining ISST.



### Saville Kushner

is Professor of Public Evaluation at the University of Auckland. Throughout the 1970s to the 1990s he worked alongside Barry MacDonald at the University of East Anglia in the development and practice of Democratic Evaluation. He has conducted and directed a number of evaluation commissions within that rubric in many areas of social and public action. His contribution to Democratic Evaluation came in his (Sage, 2000) book, *Personalising Evaluation*.



### Sebastian Lemire

is a doctoral candidate at the University of California, Los



Angeles. His area of interest revolves around causal modeling and explanation, research quality appraisal, research synthesis, and evaluation capacity building. Sebastian has published on these topics in the American Journal of Evaluation, Evaluation, and the Canadian Journal of Program Evaluation.

### Mohamed Hedi Manai

received his MBA in 1976 from the Institut des Hautes Etudes Commerciales (IHEC) in Tunisia and later served as a manager in PriceWaterhouse in Kuwait, Tunisia and Morocco. He has retired from the Africa Development Bank (AfDB) where he served for 23 years as Chief Evaluator and Division Manager for project and programme evaluations. Mr. Manai was member of the Multilateral Development Banks Evaluation Cooperation Group (ECG) working on the harmonization of evaluation processes and development of good practice standards of private sector operations evaluation.



### Ranjani K. Murthy

is a researcher and consultant on gender, poverty and evaluation based in Chennai. She has been engaged in evaluations for several years, including two meta-evaluations, one in India and one at a global level, mainly focusing on gender and social equity. She is an Adviser to the project Engendering Policy through Evaluation.



### Michaela Raab

is an independent researcher, evaluator and facilitator with more than 25 years of experience in international development and human rights. Currently based in Berlin, she has worked in a wide range of contexts across



Africa, Asia and Europe. Michaela is particularly interested in methodological innovation and making research findings accessible to a wide audience. She regularly shares musings about evaluation and development on her blog [www.developblog.org](http://www.developblog.org)

### Dr. Joseph Hare

is the Assistant Dean of the Centers for Academic Excellence and Learning Innovation at Bellevue University, Nebraska. He served as the Director of Bellevue University's Human Capital Lab from 2008 to 2013. Prior to that, his career included a variety of leadership positions designing, delivering, and evaluating training and education initiatives. He has presented at national and international conferences and has authored several papers on evaluation theory and practice.



### Khaled Samir Hussein

is a Principal Evaluation Officer with at the Operations Evaluation Department of the African Development Bank (AfDB). Mr. Hussein has worked previously with The Islamic Development Bank and the Egyptian Ministry of Foreign Trade. Mr. Hussein's evaluation work focuses on AfDB's private sector interventions, particularly the thematic areas of Small and Medium Enterprise (SME) and Public Private Partnership (PPP). Mr. Hussein holds a B.A. in Economics and Computing and an M.S. in Finance and Risk Management from the ICMA Centre, University of Reading in the United Kingdom.



### 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



is Principal Evaluation Officer at the Operations 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.

#### Wolfgang Stuppert



is an evaluation consultant and social scientist. His areas of expertise are the design, implementation and further development of surveys and qualitative evaluation methods, in particular in the fields of gender-based violence and youth education. Wolfgang's academic research focuses on protest, civil society development and democratization. He is a founding member of the Berlin Institute for Protest and Social Movement Research and a PhD candidate at Humboldt University of Berlin.

#### Ratna M. Sudarshan



is currently a Fellow at the National University of Educational Planning and Administration, New Delhi, and a member of the Board of Trustees of the Institute of Social Studies Trust. She has been part of the ISST feminist evaluation group since its inception.

#### Sara Vaca



is an independent consultant ([www.EvalQuality.com](http://www.EvalQuality.com)) expert in evaluation of international development programs, specialised in resilience and livelihoods interventions. She is also fascinated by Data visualization and she has started exploring its huge potential synergies with Evaluation. Part of her job regarding this can be seen in her blog [www.VisualBrains.info](http://www.VisualBrains.info).

Her background is Business Management and she holds a Master's in Non-Profit management and another in Evaluation (by different universities in Madrid). She often works with Red Cross and Red Crescents societies as with their International Federation (Geneva). She is also professor of M&E in a master online ([www.UOC.edu](http://www.UOC.edu)).

#### Claudine Voyadzis



has spent some 20 years working in the evaluation field. Her work consisted mainly of appraisal and monitoring of projects and programmes, impact studies and ex post evaluations in West and East Europe and in Africa. She earned her Ph.D. in the United States in Demography and Sociology and has a dual experience of Anglo-Saxon and European approaches to evaluation. After working in the Independent Evaluation Group of the World Bank in Washington, she joined the Council of Europe Development Bank in Paris as

technical adviser for project appraisal and monitoring. Subsequently she was appointed to set up the Ex Post Evaluation Department and remained for six years as Director of the Department. Claudine Voyadzis works now as an independent consultant, and leads evaluations for multilateral organisations and consulting firms.