



EVALUATIONS CONNECTIONS

EES

EUROPEAN
EVALUATION
SOCIETY

SEPTEMBER 2013

ARTICLES – NEWS – EVENTS

CONTENT

A Message from the President	1
Editorial: Evaluation in the 'real world'	2
Why do we need an Evaluation Capabilities Framework?	3
Evaluators as agents in development change	5
Evaluation for equitable development results	6
The importance of negative evidence in evaluation	8
The Theory of No Change	9
The lost prophets: a practitioner's view of M&E	10
A Portfolio Approach to Evaluation: Critical Aspects and Lessons	12
Developing Capacity for International Development Evaluation	14
The authors	15

A Message from the President

Evalpartners has provided funding for EES and the UK Evaluation Society to sponsor a joint workshop designed to debate the potential for a Voluntary Evaluator Peer Review (VEPR) system. The VEPR proposal was designed under the aegis of the EES Professionalization Thematic Working Group (TWG). The EES Board unanimously agreed that the TWG proposal deserves exposure to a broad evaluation audience as the first step in a proactive consultative process.

Professionalization has long been on the policy agenda of the evaluation community. Canada has already implemented a credentialing scheme and the International Development Evaluation Association (IDEAS) has initiated its own review. While the issues involved are controversial and sensitive I believe that the time has come for the European Evaluation Society (EES) to engage in a fulsome public debate about the pros and cons of a peer review system geared to professional development and designation.

There is little doubt that the demand for a legitimate and coherent approach to evaluator credentialing is rising. Some organisations (e.g. the UK Department for International Development) have gone ahead and set up their own evaluation designation systems. I personally think that it is up to the evaluation community to engage with the complex issues involved taking full account of the diverse and nuanced perspectives that

emerged when EES and UKES developed their capabilities frameworks.

Only evaluators should settle such questions as what evaluation is and what is not; what can be considered rigorous or not; what quality evaluation consists of; who can claim to be a fully fledged member of the 'evaluation community', etc.

Some evaluators argue that designation would help raise the bar of evaluation quality through verification of the basic attributes (knowledge, experience, and dispositions) required for sound evaluation practice. Others see merit in a peer review mechanism that would help new evaluators develop their capacity to deliver good work. Many evaluation commissioners would welcome the advent of a well run designation scheme that would attest to the capabilities of evaluation teams bidding for evaluation contracts.

On the other hand many evaluators hold the view that designation would have many unintended and deleterious consequences and that there are other more effective ways of enhancing the quality of evaluation practice.

The 'evaluation community' is very broad, heterogeneous and diverse. It is served by evaluation societies mandated to bring together people who, one way or the other, are devoted professionally or academically

to the evaluation endeavor or who as producers or users of evaluation services value the contribution of our discipline to society.

At this critical juncture in evaluation history we should not hesitate to be a little intrepid. We should ask ourselves new questions. We should search for new and innovative ways to advance our profession. I have learnt a great deal from evaluation students and graduates of the Master program that I direct. They are, along with many others in Europe and beyond, new generations who have been trained to be good evaluators, commissioners, evidence based policy makers and managers, or people in charge of the evaluation policy in their organizations. Let us listen to they have to say.

Is a system of designation a way to enhance the quality of evaluation services? Do we need one? Are there alternatives? Is a peer-review system the way to go? Are we able to foresee its positive effects? What about unintended consequences? If we are aware of them, would we be able to minimize them? If we decide to go ahead with a VERP pilot what should it look like?

I visualize a fulsome and fiery debate. For me it is more important at this moment to launch a principled discussion and a participatory process in which every EES member has a say than to settle on the 'product' and its design. If as a community we decide to 'produce' and implement a VERP, it will have to be piloted first and carefully and systematically evaluated. It cannot be another way: we are evaluators!

If you are an EES member in good standing you have already received a questionnaire that we have sent to all members on behalf of the TWG on Professionalization. Please do participate in this process and let us know what you think. If you are not an EES member or you have not renewed your membership lately, it is an occasion for (re)joining, participating and exercising your voice option in the community. You can get in touch with me or any other EES Board member. We are looking forward to know you and listen to you.

María Bustelo,
President, European Evaluation Society

EDITORIAL: EVALUATION IN THE 'REAL WORLD'

Robert Picciotto

This issue of *Connections* addresses evaluation in the 'real world'. What does it take to produce high quality evaluation work? What is the value of independence in evaluation governance? How can evaluation contribute to equitable social transformation? What counts as evidence in evaluation? How does monitoring connect to evaluation? Is it possible to draw valid evaluative conclusions and to share evaluation experience across a wide variety of social contexts?

First, Helen Simons addresses the recent advent of evaluation capabilities frameworks. She relates their emergence to internal and external influences. Specification of the knowledge, skills and attitudes needed to promote a professional evaluation culture is a natural follow up to the development of ethical guidelines and standards. But the construction of competency and capabilities frameworks also responds to evaluators' professional development aspirations and it originates in evaluation commissioners' eroding tolerance for mediocre evaluations.

Are the learning and accountability dimensions of evaluation incompatible? Not according to Indran Naidoo who argues cogently that these two goals are complementary and synergistic in actual evaluation practice: independent evaluation strengthens learning as well as accountability by bringing credible evidence to bear on decision making. Independent evaluation is especially needed in governance environments where those in authority do not hold themselves accountable to the people they are supposed to serve.

Marco Segone's article stresses the rising prominence of equity in contemporary development policy thinking. Growing inequality undermines human rights, hampers human freedom and increases social instability. Accordingly, evaluation of public interventions is incomplete without an assessment of their equity impacts. This raises methodological challenges for the evaluation function. In particular, the merit and worth ascribed to pub-

lic interventions can be severely distorted if the value frameworks used to assess public policies are inadequately informed by equity considerations.

In a vigorous opinion article that challenges the dominance of randomized controlled trials and quasi-experimental evaluations Rob van den Berg points out that such 'black box' impact assessments fail to account for the implementation failures that inevitably plague interventions undertaken in complex and volatile operating environments. The methodological fundamentalism that underlies the sudden popularity of experimental methods irrespective of the context risks inducing premature abandonment of worthy initiatives. It also denies decision makers the option of learning from experience.

In her companion piece, Christine Woerlen identifies a series of factors underlying implementation failures in climate change mitigation projects. Identifying the obstacles that stand in the way of successful implementation through meta-evaluation can lead to a better understanding of "why" interventions fail by directing attention towards promising remedial actions. This iterative developmental approach runs counter to classical theories of change that view causal linkages and assumptions as static. It suggests that theories of 'no change' may hold the key to improved public sector performance in contexts that call for adaptation and learning.

A similar concern with the uncertainties that plague public sector interventions in the zones of turmoil and transition of the developing world underlies Daniel Tichehurst's contribution. He argues that in development "*the most important things are unknowable: institutional and personal behaviours defy prediction*". Since it is impossible to predict how things will play out and why flexibility and responsiveness to the ultimate beneficiaries of development interventions (i.e. the poor and disadvantaged) is critical to success. This means that judicious monitoring, long neglected by evaluators, should

be carefully designed and actively used not just to track management performance but to help improve it towards development impact.

Can valid policy lessons be drawn from tailor made monitoring and evaluation approaches adapted to the peculiarities of individual interventions? This is the question that Rosalia Rodriguez-Garcia tackles in her article. It describes the successful 'portfolio approach' used in a meta-evaluation deliberately designed to identify what worked and why in community responses to the HIV/AIDS epidemic plaguing Burkina Faso, India, Kenya, Lesotho, Nigeria, Senegal, South Africa, and Zimbabwe. To address this 'real world' complexity, the meta-evaluation emphasized selectivity, a mixed-methods approach, multidisciplinary teams, phase-in of studies and a consultative process.

Finally, Stewart Donaldson and Marco Segone show how today's evaluation practice is being transformed by the new information technologies. Their article describes multiple initiatives (webinars, knowledge management systems, e-learning applications, massive open online courses, social media instruments) that have enhanced national and international conferences, and university-based evaluation capacity development efforts.

Thus, this Connections issue weaves together potential solutions to 'real world' evaluation problems in contexts characterized by complexity, rapid change and growing interconnectedness. They are permeated by a common desire to face up to the social and governance dysfunctions of an increasingly inequitable world. All the articles are motivated by professionalism and ethical con-

cerns and reflect a determination to share pragmatic lessons of evaluation practice across borders.

Note: Most of the articles included in this issue of Connections summarize presentations made at the Global Assembly of the International Development Evaluation Society (IDEAS) held in Bridgetown (Barbados) May 6–9 2013. This event focused on evaluation and inequality. It brought together evaluators from around the world and sought to address the 'real world' problems that evaluators must contend with in their day to day practice as they struggle to serve the global public interest

See: <http://conference.ideas-int.org/content/index.cfm?navID=1&itemID=348>.

WHY DO WE NEED AN EVALUATION CAPABILITIES FRAMEWORK?

Helen Simons

Introduction

In 2013 the United Kingdom (UKES) published a Framework of Evaluation Capabilities.¹ Two years earlier the European Evaluation Society (EES) had posted its own capabilities framework on the EES website.² Both initiatives shared the same overarching aim – to help further the professionalization of evaluation. This article explores the reasons behind the recent emergence of these frameworks and their potential use in different contexts.

How the UKES Capabilities Framework emerged

The Framework is the result of widespread consultation with the UKES membership and, like the UKES Guidelines for Good Practice in Evaluation which preceded it over a decade ago, it took three years to reach a consensus. The Framework was inspired by

a broader debate about competencies within several evaluation societies, notably the Canadian Evaluation Society (CES)³, the International Development Evaluation Association (IDEAS)⁴ and the German Evaluation Society DeGEval.⁵ The UKES initiative was especially influenced by the parallel work conducted by the EES in this area. The outcomes of both processes proved quite similar.

Why now?

There are two particular reasons, the first professional, the second political. First, the initiative is an extension of the trend during the past two decades for evaluation societies to produce ethical guidelines and standards. These helped to build a professional evaluation culture and promote good practice. However they did not specify the attributes that evaluators should have. Secondly, it is a response to the increasing awareness of the potential use and misuse

of evaluation and commissioners' growing concerns about its quality. In the absence of formal qualifications for evaluators, an Evaluation Capabilities Framework can offer a useful check on the quality of evaluation and those who conduct it.

Scope of the Framework

The UKES Capabilities Framework has three main sections: Evaluation Knowledge, Professional Practice and Qualities and Dispositions. Evaluation knowledge refers both to the knowledge of methods drawn from the social sciences but also the specific knowledge of evaluation theories and practices generated within the field of evaluation itself. Professional Practice refers to the competencies/capabilities evaluators need to have in the field to conduct a credible evaluation and the political and interpersonal skills to manage the process. Qualities and Dispositions refer to those

personal characteristics that enable evaluators to operate in difficult socio/political circumstances.

Choice of the term capability

Various people have asked why UKES has chosen to name their initiative a Capabilities Framework rather than a Competency Framework, which other associations have adopted. Here I outline some of the reasoning which led to this decision.

Several of the competency frameworks which have developed during the past five years by different societies and organizations focus on specific skills, knowledge and dispositions that evaluators ought to have. These have to be there of course: we want evaluators to have such skills and knowledge. But the process of evaluation is so much more than can be codified and represented in a list of skills and knowledge. It is what happens in the field, in the rough and tumble of the socio/political reality that determines whether a quality evaluation has been conducted. The word capability (which the EES had also adopted) captures that broader conception of what it means to conduct evaluation in the field, i.e. not just the prior skills and knowledge but the ability to, for example negotiate difficult contracts, demonstrate cultural and gender awareness, maintain the integrity of evaluation when challenged and so on.

The wider connotation of capability takes us beyond the specific focus on the individual having to meet all the skills required for any one evaluation to consider what combination of skills and knowledge is required to conduct a quality evaluation and in different contexts. In other words, the focus is on quality evaluation and how this is accomplished rather than on individuals. Many evaluations are conducted by teams, so are not depend-

ent entirely on the specific skills of any one evaluator. The word capability is more open in that sense to encompass the whole enterprise of evaluation.

Uses of the Framework

It has primarily an educative function – to raise awareness amongst all those concerned with evaluation – evaluators, commissioners, participants – of the capabilities required to conduct a quality evaluation. But within that, there are five potential uses. One is to serve as a self-evaluative aide-memoire for professional development for individual evaluators to think about what further skills and capabilities they may require to be an effective evaluator. A second is to offer a formal framework for evaluation training. A third is to provide guidance for commissioners of the different capabilities and skills needed in an evaluation team to effectively evaluate different programmes and policies. A fourth is to offer quality assurance to the external world – stakeholders and our audiences who receive evaluations – to assure them that the evaluation was conducted by evaluators who have these capabilities. And finally it can act as a framework for UKES designation of evaluators if we choose to go down that route.

Contribution to professionalism in evaluation

The main advantage of the Framework is that it is not prescriptive. It gives agency to different groups to consider how they might improve their professional evaluation capacity: to *individuals* to think about what additional skills they could acquire; to *organizations* what training programmes they might need to set up; to *commissioners* what skill sets and qualities they require in teams of evaluators to ensure they gain a ‘good evaluation’ of programmes or policies; to *societies and*

other groups the opportunity to reflect upon whether these capabilities are embedded in their procedures or what further might be needed to advance quality in the practice of evaluation

Next steps

In one sense the published Framework is just the beginning. The next step is to make it a ‘living document’. Hopefully individual evaluators and commissioners will use it, as indicated above, to guide their commissioning and practice of evaluation. But we will also schedule discussions in our annual conference and Regional networks and run workshops to test out how the capabilities might be demonstrated in practice.

I also expect that the UKES and the EES Capabilities Frameworks will be useful in the development of a voluntary peer review and designation initiative that is currently being explored jointly by both societies with support from Evalpartners.

-
- 1 UKES Evaluation Capabilities Framework. Obtainable from www.evaluation.org.uk/ or ukes@profbriefings.co.uk
 - 2 http://www.europeanevaluation.org/files/EES_Capabilities_Survey_Evaluation.pdf
 - 3 <http://bit.ly/1WtiC6h>
 - 4 <http://bit.ly/1VMq6m7>
 - 5 See www.degeval.de, de. DeGEval, Recommendations on Education and Training in Evaluation, Requirement Profiles for Evaluators, Mainz, Germany, 2008

EVALUATORS AS AGENTS IN DEVELOPMENT CHANGE¹

Indran Naidoo

Independent evaluators are in a privileged position to effect a multitude of changes through their work. This is because of their ability to influence the performance agenda by bringing in evidence and perspectives that rightfully challenge misleading perceptions of performance based on self-reporting. Independent evaluation forces a candid reflection on the connection between stated goals and actual outcomes. The sharp juxtaposition between policy and strategy on the one hand and impacts on the ground on the other provides a timely reality check, and helps achieve course corrections.

The pursuit of independent evaluation inevitably makes evaluations unpopular due to the starkness of evaluative language. Whilst providing contextual detail and acknowledging effort independent evaluation ends with an evaluative judgment of results and performance. Conversely evaluation as a value-ascribing activity requires the protection of independence. Accordingly this article argues that independence is fundamental to methodological soundness and evaluation credibility.

The distinction between an independent evaluation and other assessments that inform programme management is critical. The notion that one should privilege learning and by implication take a process orientated approach that steers away from making judgments misses the point. If it is not independent it cannot be called an evaluation and it should not be presented as such. It is only part of normal management and cannot replace a credible accountability oriented evaluation function. Learning should be on-going but it should also embody accountability, a critical dimension of independent evaluations. Independent evaluation must transcend personalities and the fashions of the time. It must pitch toward objective assessment of organizational performance. It is obliged to reach beyond short-term interests and must

contribute to course correction. It must be supported by senior leadership and oversight bodies, who have an ethical responsibility to protect the function and ensure that higher level probity is supported as there shall always be the predictable pushbacks from various sections of an organisation that tend to suppress critical performance information. A mature and responsible organisation that respects its accountability function shall ensure that its independent evaluation is structurally and behaviorally pitched to provide critical perspectives, and as part of its ongoing quality assurance regular reviews of the policy effectiveness of such units must be undertaken. The Peer Reviews by the OECD/UNEG/DAC Network have proved useful, in addition to the more comprehensive policy reviews that take place as per programme of work. There are several examples where multilateral and bilateral organizations have constituted independent evaluation offices for periodic evidence-based programme assessments. The UNDP has been a leader in the UN system by approving and respecting an Evaluation Policy that guarantees independence of the evaluation function. This attests to a solemn commitment to provide the level of assurance required by member states represented by the Executive Board to which the Evaluation Office reports directly.

This is in line with the United Nations Evaluation Group (UNEG) norms and standards for evaluation, and mirrors the practice of evaluation units in the International Financial Institutions and the Organization for Economic Cooperation and Development and its Development Cooperation Directorate (OECD/DAC). Hence UNDP's evaluation policy and practice are consistent with what the global evaluation community advocates. Furthermore, independence needs to be seen as a credibility issue relating directly to the criterion of utility: a poor quality evaluation is potentially damaging.

Evaluation should be unbiased. It needs to assess the intervention from the perspective of the beneficiaries or recipients of the programme. Evaluation work has the potential to shift power relations towards the poor and disempowered, giving them voice, and helping to ensure that their experiences are more fully considered by decision-makers. Accordingly development evaluators have the potential to promote "performance democracy."

This is a collective responsibility. Development evaluations of high quality, if strategically led and managed, can help generate verifiable development results and touch the lives of impoverished citizens throughout the world. The UNDP has poverty eradication as a key priority area. It recognizes that "empowering lives and resilient nations" implies that inequality should be reduced to promote peace, security and development.

We can promote transparency by bringing performance results to the surface, increase critical dialogue and reflection and break down the many walls that often exist between policy makers, implementers and citizens. What counts is what happens on the ground, and the unit of analysis for measuring change should be the intended beneficiaries whether citizens, groups or institutions. It cannot be amorphous. Too often we measure what is easy, although it may be irrelevant and it is thus not surprising that after much evaluative activity we do not effect change by saying the same things or by failing to produce critical diagnostics lest we offend.

Global efforts to accelerate progress towards the Millennium Development Goals (MDGs) are underway. The share of persons living in poverty has improved but the gap between rich and poor has been widening. From an independent evaluation perspective more diagnostic questions must be asked

where levels of inequality are rising and the poor and disempowered remain below the radar screen. The journey is not over.

Development evaluation is all encompassing, rooted in real contexts and development questions, and most importantly focuses attention on the fact that evaluation should improve the quality of lives of the world's citizens. It heralds an implicit agenda for change. This means that the pursuit of development evaluation should recognize that key questions about inequality need to be addressed; making the evaluator not just a technical expert but also one who understands the social and political context, and recognizes the difficulties in producing results. Development evaluators must keep in mind that their work is designed to bring about social transformation. There is thus a clear purpose that drives development evaluators that must not be lost, as ultimately the purpose is to make sure that positive change is actually taking place.

Our job is to help answer the “so what questions”. What happened to the development investments, the promises of delivery made, the transformations promised? More mentoring is required, as well as the development of competencies and evaluator certification. The evaluation community must continue to look for ways to support independent evaluation in difficult places where inequality is a pervasive problem and where governments do not hold themselves accountable to the people they supposedly serve.

I This article is based on keynote address presented at the IDEAS Global Assembly, Barbados in June, 2013.

EVALUATION FOR EQUITABLE DEVELOPMENT RESULTS

Marco Segone

The context

According to Bob Picciotto's keynote address at the Global IDEAS Conference in Barbados in April 2013, 40 percent of the world's wealth is owned by the richest one percent of the population while the poorest half own only one percent; average per capita income in developed countries is seven times that prevalent in the average developing country; child mortality is 18 times higher in Mali than in the USA and the average OECD person has 13 years of schooling compared to less than 6 in Sub-Saharan Africa.

These facts help explain why equity has become central to development policy. It is now widely recognized that inequity hampers human development:

- *Inequity constitutes a violation of human rights.* Based on the principle of universality and non-discrimination, all human beings, from birth to childhood and adolescence, boys and girls, of whatever colour, race, language or religion and wherever they may live, are entitled to the same human rights.
- *Inequity is a major obstacle in taking advantage of diversity.* Unequal societies do not give individuals and groups equal opportunity to contribute to development. Focusing on a single “cultural model” precludes taking advantage of diverse “cultural models” and the societal innovation and creativity that they foster.
- *Equity has a significant positive impact in reducing income poverty.* As demonstrated by Martin Ravallion, a 1% increase in incomes cut income poverty by 0.6% in the most unequal countries and by 4.3% in the most equal ones.
- *Equity contributes to democratization:* it facilitates citizen participation in political and civic life.
- *Inequity may lead to political conflict and instability.* Unequal opportunities for so-

cial groups – and inequities as perceived by them – are a significant factor behind social unrest.

The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda presented at the UN Secretary General in May 2013, and which will inform the next round of Sustainable Development Goals, calls for ensuring that no person – regardless of ethnicity, gender, geography, disability, race or other status – is denied universal human rights and basic economic opportunities, and request the international community to design goals that focus on reaching excluded groups.

What are the implications for the evaluation function?

The new focus on equity in development poses important challenges to – and offers opportunities for – the evaluation function: What are the methodological implications in designing, conducting, managing and using equity-focused evaluations? What are the questions should equity-focused evaluation address? What are the potential challenges in managing equity-focused evaluations? This article highlights the key issues. A fuller treatment is found in the references listed at the end of this article.

What is an equity-focused evaluation?

An equity-focused evaluation is a judgment made of the relevance, effectiveness, efficiency, impact and sustainability – and, in humanitarian settings, coverage, connectedness and coherence - of policies, programmes and projects concerned with achieving equitable development results. It provides reliable assessments of what works and what does not work in the reduction of inequities and it highlights intended and unintended results

for worse-off groups as well as the gap between best-off and worse-off groups.

In the example from Nepal depicted above, if the national policy to reduce malnutrition was evaluated using “traditional” evaluation approaches, the main finding would have been positive, as the national average of malnourished children decreased. However, if the same programme was evaluated using an “equity-focus”, the main finding would have been negative, as malnutrition among children from the worse off groups not only did not decrease, but actually increased significantly as did the gap between the best-off and worst-off groups. Therefore, evaluation of the same policy generates different findings, and consequently different recommendations, if carried out with or without an equity-focus.

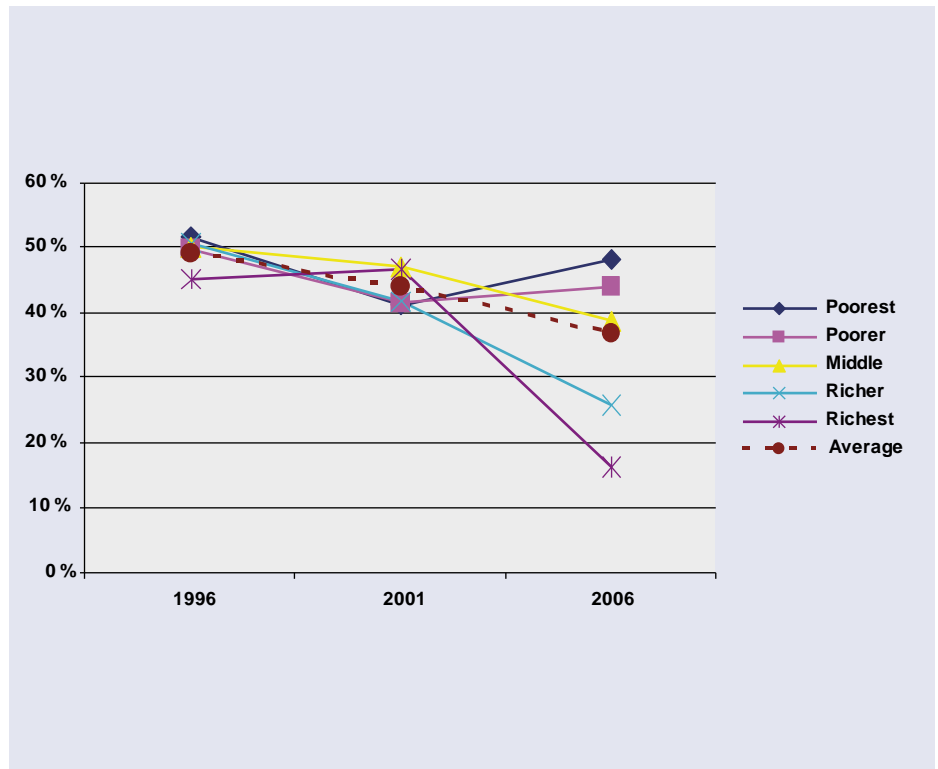
From a methodological point of view, equity-focused evaluations also go beyond conventional quantitative data to the analysis of behavioural change and complex social processes and attitudes, using such approaches as “bottleneck analysis” (Bamberger and Segone, 2011).

Last but not least, equity-focused evaluation processes are shaped so as to empower worst-off groups to the full extent possible, including making sure that evaluation questions are relevant to these groups. This means that equity-focused evaluation should be culturally sensitive and comply with strict ethical standards. In particular, equity-focused evaluation should use participatory and/or empowerment evaluation processes to ensure that disadvantaged groups are meaningfully involved and part of the decision making process starting at the design phase. Selection of stakeholders in equity-focused evaluation processes and methods ought to correct, and not reinforce, prevailing patterns of inequity and exclusion.

Potential challenges

An indicative list of challenges faced when undertaking equity-focused evaluations is given below. It aims to help evaluation manager and stakeholders identify challenges early on in the evaluation process and develop relevant strategies to overcome them:

- *Potential challenges in promoting equity-focused evaluations:* reluctance to accept



Trend for children (6–23 months) by Wealth Status.

disaggregated indicators, which can show country performance in a poor light; political and social resistance to addressing the causes of exclusion and vulnerability; resistance to empowerment of worst-off groups; lack of interest/incentives and reluctance to invest resources in the worst-off groups; weak decentralized governance; and the legal status of worst-off groups.

- *Potential challenges in implementing equity-focused evaluations:* methodological challenges in the evaluation of complex interventions; lack of disaggregated data or data collection capacity, and reluctance to change existing methodologies; and additional cost and complexity.

Conclusion

The development community is more and more promoting social inclusion and equity. Equity-focused evaluations are gaining traction as they address the key question increasingly posed by policy makers: are interventions producing equitable development results? Equity-focused evaluations while more complex can provide relevant answers to this question.

References

Pickett, R. & Wilkinson, K. (2009). *The Spirit Level: Why equality is better for everyone*. Penguin Books.

Ravallion M. (2011). *How long will it take to lift one billion people out of poverty?*, World Bank.

Segone, M. (2003). *Equity and diversity in Brazilian childhood and adolescence*. UNICEF Brazil.

Segone, M. (2004). *How to achieve the Millennium Development Goals? Reducing inequality through the celebration of diversity*, UNICEF Global Policy Section.

Segone, M. and Bamberger, M. (2011). *How to design and manage Equity-focused evaluations*. UNICEF. Available at <http://www.mymande.org/selected-books>.

Segone M. et al (2012), *Evaluation for equitable development results*, UNICEF in partnership with CONEVAL, IDEAS, IDRC, ILO, IOCE, UNDP and UNWomen. Available at <http://www.mymande.org/selected-books>

UNEG (2011). *Integrating human rights and gender equality in evaluation*.

UN (2013), *Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda*.

THE IMPORTANCE OF NEGATIVE EVIDENCE IN EVALUATION

Rob D. van den Berg

Over the past decade, evaluations have been influenced by the scientific method: to show what works and what doesn't through 'rigorous' methods. The focus of discussion has often been on proving that an intervention actually works. Relatively less attention has been paid to why a specific intervention or policy does not work. The stock answer from impact evaluators has been that failure indicates that a supposed causal mechanism does not exist. But other evaluators have been quick to point out that the causal mechanism may in fact work except that the intervention did not succeed due to "implementation failure". The reasons for failure can be manifold, and evaluations often cannot clearly identify why something does not work.

Many impact evaluation tools such as randomized controlled trials or quasi-experimental evaluations use a 'black box' approach. They are not geared to identify implementation failure, do not differentiate it from design failure or seek an explanation in unexpected changes of external circumstances. Data is gathered but only its relevance to the proof or disproof of the causality hypothesis is considered. Evidence not relevant to the attribution question is discarded as irrelevant. Those in favor of this approach tend to dismiss such knowledge: if something does not work, one should obviously stop doing it, they say. However, an important objective of evaluations is to learn from mistakes and the current popularity of one particular form of impact evaluation is not helping.

The disregard for learning from mistakes may be related to the assumption that we need large n to provide positive proof that something works, and that in fact only large n can have scientific meaning. In a presentation to the International Development Evaluation Association Global Assembly in Barbados in May 2013 I pointed to the fact that the philosophy of science recognizes more than one way to skin a cat or to use rigorous methodology to explain what is happening in interventions. In fact, a full range of n (from very

small to very large) is in use in scientific research. For example, in astronomy the adoption of new or rival scientific theories is not decided by experiments or by differences in n , but by the differences in explanatory power between theories.

A guiding scientific principle has been Occam's razor; i.e. when there are conflicting theories based on the same data, scientists should adopt the theory that is simpler. Both the Ptolemaic and the Copernican theories of the solar system use the same n : data on which planet is where in the sky at what moment in time. The Copernican system provides a simpler explanation and calculation of the movements of the planets than the Ptolemy version, which had convoluted circles within circles to explain why a planet was observed at a certain point in space and time. Still simpler explanations were provided by Kepler, showing that the planets do not run in circles around the sun, but ellipses. Again, his laws of planetary motion were taken up in astronomy not because new n emerged, or experiments were conducted in the planetary system but because they provided a better explanation than Copernicus' proposals that the planets ran in circles around the sun.

Astronomy also provides us with an example of one n that tipped the balance between the Newtonian view of gravity and the Einsteinian view. Einstein's general theory of relativity posed that light rays would be bent by gravity; Newton's theory does not include this. At an eclipse of the sun on May 29, 1919, one star was observed that was actually behind the sun; its light was bent by the gravity of the sun so that it could be seen. This one n proved to the world that Einstein was right and Newton was wrong. The search for such a crucial and decisive n is often elusive; it is difficult to find one n that would prove or disprove a theory. The best known example of negative evidence through one n in the philosophy of science has been the "all swans are white" theory that philosopher Karl Popper used to demonstrate that if one

black swan is observed, the theory would be disproven.

Popper proposed to shift looking for evidence from positive that something works to negative that something does not work. His claim was that enormous numbers of positive data in the world that swans are white will not survive the sole fact of one black swan. Popper posited that it is better to search for black swans, i.e. to disprove theories, than to gather ever larger n that still would not prove that the theory is the case. This is easier said than done, as illustrated in the contributions of Nassim Taleb on "black swan events" in hedge fund and investment calculations, based on very large n , which led to the 2008 international financial crisis. Taleb argues that statistical systems of large n are fundamentally incomplete as they cannot deal with unexpected events.

In the current emphasis in the evaluation community on large n , the lessons from the natural and physical sciences are ignored and evaluations are increasingly turning into scientific research on whether a causality can be demonstrated to exist through statistical analysis of large n . This is a search for positive evidence that disregards the induction problem and disregards small n – and especially unexpected "black swan" n – as irrelevant for scientific discovery. The current mainstream claims that only large n – through counterfactual analysis based on control groups – will demonstrate whether something works. Facts, from this perspective, are only powerful if you have a lot of them. There is nothing wrong with a lot of facts – the more the better – but if we only select them for the positive conclusions we can draw from them, we fail to learn from failure.

Even single facts may have the power to prove or disprove theories. This requires analyzing facts that experimental methods do not normally capture. By gathering evidence on implementation failure, barriers to progress and unexpected developments one may learn from failure as well as success.

THE THEORY OF NO CHANGE

Christine Woerlen

Evaluating projects, programmes and policies is typically based on information supplied from within the narrow domain of the intervention – implementers, designers, recipients, affected stakeholders. Often, the program logic is taken from project design documents, and not questioned further. If a project fails the programme logic might be discarded as faulty without further ado. But without going beyond the assumptions and logic that underlie the evaluandum it is often hard to understand why an intervention might not have delivered the intended results. In such cases learning from the evaluation is necessarily limited.

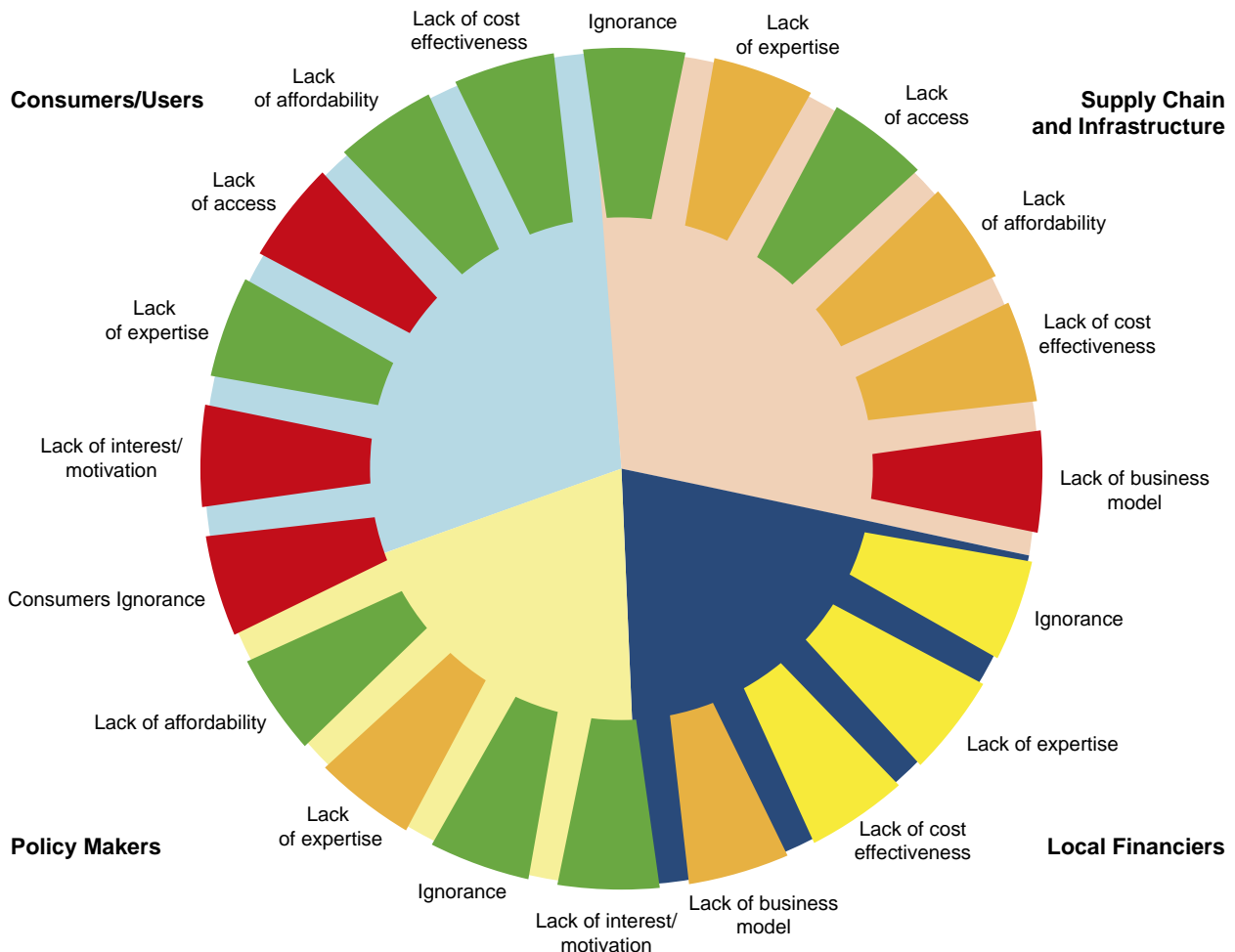
A meta-evaluation of climate change mitigation evaluations supported by a community

of practice hosted by the Evaluation Office of the Global Environment Facility (GEF EO) identified a series of factors underlying failures. This led to a better understanding of “why” interventions failed and directed attention towards policies that might have more impact on climate change trends. Rather than a classical theory of change, which postulates that certain causal linkages and assumptions make an intervention “work” a theory of no change puts forward hypotheses regarding why certain causal linkages are in fact broken or why interventions cannot (yet) work in identified circumstances.

The meta-evaluation led to the formulation of a framework that identified explicit barriers to change – in this case intended market

changes – that had prevented up-scaling of desired practices – i.e. energy efficiency measures. A case study of ten evaluations on energy efficiency projects, policies and programs in Thailand was undertaken to test whether twenty identified barriers helped explain market dynamics. A second case study in Poland was used for further testing. The case studies helped reduce the “Theory of no change” (or TONC) framework to seventeen crucial barriers.

The idea behind the barrier framework is that it is not always the behavior of the target group of an intervention that makes an intervention fail. Specifically the analysis identified – for the case of energy efficiency – four (macro-) groups of actors whose behaviors



also shape outcomes, i.e. significant barriers to change are found among users of energy, suppliers of energy-using equipment, financiers and policy makers. In turn, each of these actors faces one or more of six generic types of barriers: lack of motivation, lack of awareness, lack of access to the “better” technology, lack of technical expertise, lack of affordability, lack of cost effectiveness. In some cases, they may be part of the intervention and the original program already. In most cases, where projects failed, these were actually not part of the original considerations but constituted “contextual challenges” to project success. A mapping tool was developed to illustrate the relevance of these barriers to the achievement of intended outcomes (cf. figure). The traffic light scheme informed by evaluation data allows for instant analysis of a given situation. Specifically the barriers that have proven to be effectively limiting change are marked in red. Those that exist but are not decisive bear orange and yellow colors while barrier-free dimensions are displayed in green.

The generic challenge ahead is to integrate “theories of change” with “theories of no change”. Following Ray Pawson, we believe that it is useful and possible to generate useful theories through “realist synthesis” as in this

case – by providing a legitimate role for negative proof to help improve our understanding of what is happening compared to what was supposed to happen. In fact, a composite model of why things were supposed to happen combined with a systematic assessment of why they did not happen enables evaluators to look beyond the aspects that programme sponsors have considered already when designing an intervention. It provides the evaluator with a structure for a systematic and empirical analysis of possible root causes of implementation failure thus helping him better understand situation and context of the evaluandum and reducing the risk of wrongly ascribing failure to design faults or of getting lost in a plethora of alternative explanations for the observed outcomes.

The power of this analytical tool has led to further identification of the importance of negative evidence in the work of the GEF Evaluation Office through case studies. This framework for understanding “no change” is now being developed into a tool that can be used not only for *ex post* evaluations, but also for diagnostic purposes when new interventions are designed. Currently it is being applied to the identification of new areas of intervention for the German Ministry’s National Climate Initiative.

The climate and development evaluation community of practice (www.climate-eval.org) which developed this stakeholder-based conceptualization of the Theory of No Change has presented it at evaluation events hosted by the International Programme Evaluation Network (IPEN) and the International Development Evaluation Association (IDEAS). Preliminary reactions indicate that this model or at least the fundamental thinking behind it might be useful in many sectors and evaluation fields. The model itself or the basic approach can potentially be helpful in such domains as health care or education where it is clear that changed behavior by various actors is in the interest of all either because it is more cost effective than the current behavior, or because it bears additional and originally unintended long-term benefits. In such cases, there were non-obvious barriers to desired behaviors either by the target group of an intervention or by other actors that operate at levels that were disregarded at the design stage. If you are looking for more information on this topic you may wish to join Climate-Eval.org. The studies and a webinar can be downloaded at (<http://www.climate-eval.org/?q=node/28>), to explore the issue further please join the climate-eval community or contact me (woerlen@arepo-consult.com).

THE LOST PROPHETS: A PRACTITIONER’S VIEW OF M&E

Daniel Ticehurst

The purpose of development aid is to induce transformation processes within complex systems in operating environments characterized by ‘jinks’ and ‘sways’. Often the most important things are unknowable: institutional and personal behaviors defy prediction. It is impossible to predict how things will play out and why. Flexibility is therefore important. Most politicians and civil servants in donor countries struggle with the resulting uncertainty.

Many of us are enthralled by evaluation; less so by monitoring. It has become neglected due to the hegemony of evaluation, particu-

larly impact evaluation. Of course, evaluation is important. It attracts more intellectual and institutional support. By contrast, monitoring is almost taken for granted and widely perceived as subordinate to evaluation. The differences are disobligingly explained: monitoring assesses whether aid is “doing things right” while evaluation is about whether it is “doing the right things”. The popular use of pre-fixes confounds: people-centered, participatory, process, and outcome mapping etc...

Monitoring, the focus of this article, is a practical antidote to complexity. It is or rather

should be integral, not a contrived parallel, to management if it is to meet its purpose: to improve, not just comment and report on, performance

To achieve its purpose monitoring needs to set in motion processes that answer:

1. *On Effort* – What and who are we spending money and effort on?
2. *On Outreach* – How popular is our support, who is responding and how and why does this vary by the support made available, place and people?
3. *On Change* – How and in what ways are ‘beneficiaries’ (households, institutions

etc..) – and others – changing their behaviours and how does this vary among places and people?

No doing so begs a fundamental question: what results matter most to beneficiaries?

The restrictive culture of the public sector

M&E is embedded in cultures that comply with politically-led illusions. The problem is not that some donors are allocating insufficient resources; rather they are often used in the ‘wrong’ ways and for the ‘wrong’ people. This is as true today as it was in 1983:

“The procedures adopted for designing and implementing aid interventions often become ever more rigid and detailed at the same time as recognizing that development problems are more uncertain and less amenable to systematic design, analysis and monitoring.”

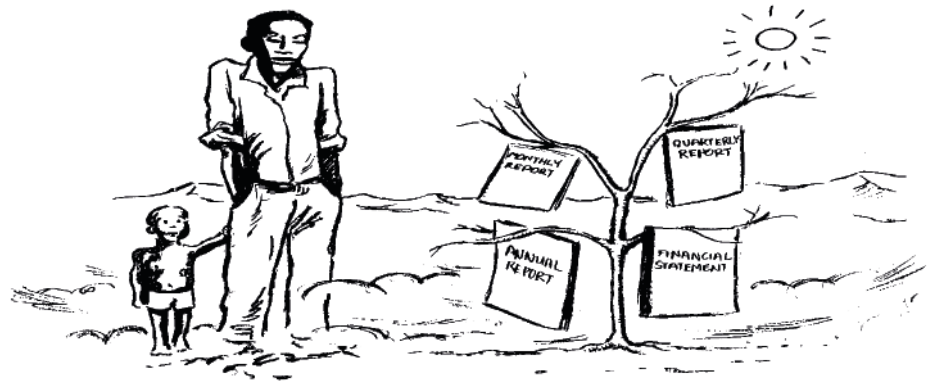
...or in 1994 as confirmed by Dennis Casley’s comment on training materials developed by the author:

“References to baseline and terminal surveys reflect the need, but during the course emphasis will be needed on the difficulty of measuring change in socio-economic variables in a short span of years in establishing the significance of change and its attribution to aid interventions. Such surveys will not establish trends in statistically ‘noisy’ systems but their imposition leads to expenditure of the major proportion of M&E effort on these surveys and unrealistic expectations from donor economists that a ‘rate of return’ will be calculable from these surveys.”

One lesson from the 1980’s was the futility of setting up baselines to measure attributable outcomes and especially impacts based on the mistaken belief that before-after comparisons are valid pursuits. These cockroach policies emasculate those responsible for managing change. They ignore their decision uncertainties. To paraphrase Paul Krugman: *People who can do Monitoring do; those who can’t try and write about how to.* Often the latter group impedes the former.

The centrality of creativity

One task of monitoring should be to show, as Hayek pointed out, how ignorant we are



in imagining what we can design and implement to bring about change.²

“If you are not prepared to make mistakes you will never come up with anything original”³

Donors vary in their capacity to allow for creativity in understanding results. Often outcomes and impacts are treated as predictions rather than aspirations. Successful strategies emerge from a process of trial and error. Monitoring should provide discretion, space and voice to those intended to benefit from aid interventions. Benefits do not result from implementing a fixed intention and mobilising around it – nailed down by a Log-frame (or a theory of change or a results chain if the Logframe’s simplicity offends some practitioners).

Acknowledging other lessons from the past

Much current M&E are not adequately informed by past lessons. Lawrence Salmen’s work on Beneficiary Assessments (BA) and listening has over twenty years of practice and lessons that precede recent efforts by DFID, Keystone Accountability and the Humanitarian Sector, for example.⁴⁵⁶⁷ Useful lessons have also emerged from the use of the Balanced Scorecard, specifically in relation to the client perspective that followed Casley and Kumar’s work on understanding farmers’ production behaviours in the late 1980’s.⁸

Currently relevant is his work of the early 1990s’ about institutional perspectives on poverty reduction that underlie sustainability:

“Growth and prosperity are unlikely to be maintained if the institutions which guide them are dysfunctional. And if development is to be accompanied by poverty reduction, those organisational competencies and values must also be attuned to the needs of the poor.”⁹

BAs are used to adapt aid programmes to changes in the operating environment by focussing on people, their perceptions and behaviours. This learning process reinforces interaction between the programme and the institutions that require change. Meaningful and practical contributions to understanding how and with what effect systemic approaches catalyse institutional changes are proving elusive among those who advocate and write up monitoring standards.¹⁰

The case for monitoring

Good monitoring is pursued by developing communication processes that help understand beneficiaries so as to make aid accountable to them. It does not require rigorous methods geared to academic concerns and obsessive pursuits of attribution of intervention effects. Monitoring has different requirements reaching beyond mechanistic ‘feedback loops’ to enhance efficiency. It is about developing ways of being accountable to poor people, neglected communities and struggling institutions through listening, gauging how responses vary, feeding back interpretations and taking appropriate action. This does empower.

Managing these processes involves an organic approach. One whereby monitoring focuses on how it can help create the environment (conditions we can control) by which change

can flourish through engaging with and listening and responding to those we support.

As Korten puts it:

*“The key to learning is not analytical method, but organizational process; and the central methodological concern is not with the isolation of variables (as in establishing a counterfactual) or the control of bureaucratic deviations from plans (the pop objective of monitoring), but with effectively engaging with the necessary participation of system members in contributing to the collective knowledge of the system”.*¹¹

- 1 Development Projects as Policy Experiments. An Adaptive Approach to Development Administration. Development and Underdevelopment Series. D. Rondinelli. Methuen. 1983
- 2 *The Fatal Conceit*. Frederick Von Hayek. University of Chicago Press. 1991.
- 3 <http://sirkenrobinson.com/>
- 4 *Listen to the people: participant-observer evaluation of development projects*, Oxford University Press, 1989
- 5 http://www.openaidmap.org/documents/DFID_OAP_presentation_Stockholm_April_2012.pdf
- 6 <http://www.keystoneaccountability.org/analysis/constituency>
- 7 *Time to Listen: Hearing People on the Receiving End of International Assistance* CDA Collaborative Learning Projects. 2012.
- 8 *Project Monitoring and Evaluation in Agriculture* by Dennis J. Casley , Krishna Kumar. Johns Hopkins University Press. 1987.
- 9 *Reducing Poverty: An Institutional Perspective* Poverty and Social Policy Series Paper No. I Lawrence F. Salmen 1992.
- 10 <http://www.enterprise-development.org/page/measuring-and-reporting-results>
- 11 D. Korten (1981). *Management of Social Transformation at National and Sub-national Levels*. Unpublished Paper, Manila. Ford Foundation.

A PORTFOLIO APPROACH TO EVALUATION: CRITICAL ASPECTS AND LESSONS

Rosalía Rodríguez-García

From 2009 to 2012, the World Bank, the UK’s Department for International Development (DFID) and the UK Consortium on AIDS and International Development (UK Consortium)¹ spearheaded an evaluation to determine the results achieved by community responses to HIV/AIDS. Actions taken by communities to prevent or address a problem barely capture the complexity of the phenomenon (Rodríguez-García et al., 2011). To address this complexity, this evaluation emphasized selectivity, a mixed-methods approach, multidisciplinary teams, phase-in of studies, and a consultative process.

The “portfolio” approach was designed to capture a wide diversity of contexts and processes. The evaluation comprised a total of seventeen studies, including country-specific evaluations (Burkina Faso, India, Kenya, Lesotho, Nigeria, Senegal, South Africa, and Zimbabwe); secondary analysis of data and desk studies. The vast scope of the undertaking underscored the need for a coordinated approach to determine the optimal resource mix needed to reach the evaluation goals within the constraints imposed by resources, timeframes and real-world factors. **Three Critical Aspects of a Portfolio Approach to Evaluation.**

Selectivity

It was not possible to evaluate all aspects of community responses in all regions affected by the HIV epidemic. The approach used was to (i) select countries reflecting different forms of community responses, (ii) probe the various dimensions of the response in a few countries, and (iii) evaluate specific community HIV/AIDS interventions in other countries. This was complemented by global surveys and desk analyses.

Two evaluations (Kenya and Nigeria) assessed the effects of the intensity (low versus high) of CBO activity at the community

level². An intervention-specific evaluation (Kenya) assessed the effects of home-based testing and counseling. In Burkina Faso, the evaluation examined the effects of prevention programs implemented by community-based organizations (CBOs), and the protective effects of indigenous community group membership (such as burial societies, women groups, sports clubs, and AIDS-related groups) in Zimbabwe. The empowerment of groups at high risk of infection was prominent in two evaluations in India. The effects of peer mentoring on HIV testing and counseling were evaluated in Senegal, while the effects of peer support on adherence to antiretroviral treatment were studied in South Africa. Finally, the study in Lesotho investigated the socioeconomic determinants of HIV-related stigmatization attitudes. Ethical and quality standards were respected throughout (Rodríguez-García et al. 2013a, 2013b).

Mixed-method approach

Diverse methods compensate for each other’s weaknesses. The evaluation made use of quantitative, qualitative, financial, spatial and other data. More importantly, it applied a variety of evaluation designs, maximized the use of available data for secondary analysis (e.g. Demographic Health Survey data) and generated new data. Three evaluation studies used an experimental design with households, individuals, or community randomization. Six others were quasi-experimental using repeated cross-sectional surveys and propensity score matching or other matching methods.

Additional information was provided by qualitative studies and analysis of CBO budgets. In Zimbabwe, a study of the role of grassroots organizations provided information on the pathways through which behavioral changes takes place. In India, trend analyses illuminated the process by which

sex workers and gay men are empowered. Information on resources mobilized came from a three-country analysis of funding flows in India, Kenya and Nigeria. Taken together, the studies generated a robust body of evidence.

Phased-in approach

A virtual survey of stakeholders was followed by (i) a typology of community responses, (Rodriguez-García et al., 2011) and (ii) an analysis of funding flows from donors through a global survey of more than 140 large and small civil society organizations (CSOs) (Bonnell et al., 2013).

The first study in Kenya assessed the effects of CBOs on a continuum of HIV interventions, from prevention to care, support, and mitigation. This evaluation paved the way for the evaluation in Nigeria. These two evaluations yielded important findings on the overall effects of community responses, but they also showed gaps, e.g. with respect to the effects of home-based counseling and testing. This was later assessed in Kenya, and the empowerment effects of high-risk groups were studied in India. The diversity of CBOs prompted a further investigation into the indigenous community groups of Zimbabwe.

Consultative Process

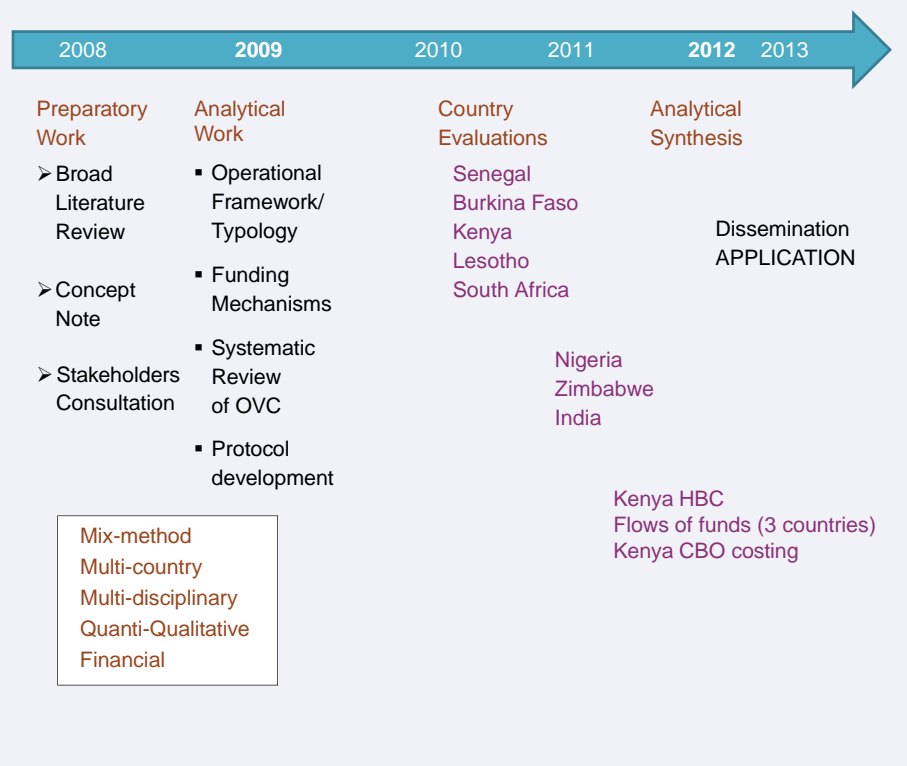
Partnership among the three core agencies (DFID, the UK Consortium and the World Bank) met the UK Audit Commission criteria³. The partners:

- were otherwise independent bodies,
- agreed to co-operate to achieve a common goal,
- created a new process to achieve this goal,
- planned and implemented a joint programme, and
- shared relevant information, risks and rewards.

The consultative process was deliberate, purposeful and sustained. It sought input from a wide pool of specialists and engaged national and international stakeholders in informed discussion about evaluation designs, implementation, and findings dissemination.

Evaluators peer-reviewed each others' studies. Through the UK Consortium's connec-

PHASE-IN APPROACH CRITICAL PATH



Source: Author. Notes: CBO=community-based organization; HBCT=home-based counseling and testing; OVC=orphans and vulnerable children.

tion to civil society networks, the evaluation team engaged with key actors on the frontlines of HIV/AIDS.

One of the major lessons learned relates to what we decided not to do:

- **Not establishing a formal advisory group.** Flexibility was preferred for a deeper engagement of national and international specialists.
- **Not selecting participants based on narrow criteria of experience and skills.** Different specialists, partners, CSOs, and agency representatives were selected to participate in specific consultations according to the themes discussed.
- **Not scheduling regular meetings but communicating around concrete issues.** The use of technologies allow for real-time feedback and enabled a more agile, iterative, and adaptive approach

to the consultation process (Kirkpatrick, 2013).

Conclusions

Randomized control trials alone cannot in isolation illuminate complex social phenomena. This complex evaluation required a mixed methods approach that was applied with creativity and rigor. Collaboration and consultation with specialists and other stakeholders increased evaluation efficiency. Purposeful, flexible and sustained consultation resulted in substantive tangible and intangible improvements. Meaningful engagement with civil society was critical. The evaluation actively sought to include multiple viewpoints at each level of the evaluation process. Civil society inputs shaped the evaluation from the very beginning. This deliberate consultative process was a critical aspect of a successful multi-study, mixed-method approach.

The process was effective and generated value.

References

Bonnel, R. Rodriguez-García, R., J. Oliver and Q. Wodon with S. McPherson, K. Orr and J. Ross. (2013). Funding mechanisms for civil society; The experience of the AIDS Response. Washington DC: The World Bank.

Kirkpatrick, R. (2013). Evaluation for the 21st century. *Evaluation Connections*, March 13, 3.

Rodriguez-García, R., R. Bonnel, N. Njie, J. Oliver, B. Pascual, and Q. Wodon. (2011). Analyzing Community Responses to HIV and AIDS: Operational Framework and Typology. Washington, DC: World Bank Policy Research Working Papers Series No. 5532.

Rodriguez García, R.; Bonnel, R.; Wilson, D.; N'Jie, N. (2013a). Investing in communities achieves results: Findings from an evaluation of community responses to HIV and AIDS. Washington DC: The World Bank.

Rodriguez García, R.; Wilson, D.; York, N.; Law, C.; N'Jie, N.; Bonnel, R. (2013b). Evaluation of the community response to HIV and AIDS: Learning from a Portfolio Approach. *AIDS Care*, vol. 25. No. Supplement 1, 7–19.

- 1 “The UK Consortium is a network of over 80 not-for-profit, faith-based and academic agencies based in the UK, with strong links to governments, international and multilateral agencies and civil society networks. www.aidsconsortium.org.uk.
- 2 Intensity was defined as the number of community-based organizations per 100,000 people.
- 3 <http://www.goodpracticeparticipate.govt.nz/levels-of-participation/collaborative-processes-and-partnerships/index.html>.

DEVELOPING CAPACITY FOR INTERNATIONAL DEVELOPMENT EVALUATION

Stewart I. Donaldson and Marco Segone

At the IDEAS Global Assembly this past May in the beautiful island of Barbados we engaged in discussions about improving international development evaluation capacity. With colleagues Tarek Azzam, Matthew Galen and John LaVelle, we discussed multiple initiatives ranging for the use of webinars, knowledge management systems, e-learning applications, massive open online courses, opportunities to use social media to enhance national and international conferences, and a range of university-based capacity development efforts. In this brief article we describe some of those initiatives.

There is a very strong demand for access to affordable and high quality resources by evaluators in developing countries. MyM&E, a web 2.0 platform, is one answer to that demand (MyM&E Website: <http://www.mymande.org/>). MyM&E offers thousands of resources, the possibility for users to upload their own videos and materials; to interact with world-level speakers through live webinars; and to share ideas with peers all over the world through the LinkedIn-based Community of Practice, as well as via twitter and Facebook.

UNICEF and the Claremont Graduate University with support from the Rockefeller Foundation hosted a series of 33 live webinars with leading evaluation experts from around the world on cutting edge evaluation topics during the past two years. The webinars were attended by thousands of evaluators and researchers from around the world and each webinar was evaluated carefully (recorded versions of the webinars are available on MyM&E). Based on evaluation feedback, some of the highlights of the series were used to build a Massive Open Online Courses (MOOCs) for evaluators.

The MOOCs on Development Evaluation were launched in September 2012. Three courses were developed: Equity-focused

evaluation; National Evaluation Capacity Development for Country-led M&E Systems; and Emerging Practices in Development Evaluations. These courses include online lectures, reading materials, multiple choice tests, and access to a dedicated Community of Practice in which participants can interact among themselves, as well as pose questions to the speakers. At the end of each course, if 80% of the participants' answers to the test questions were correct, they were granted permission to print out a certificate of "virtual" attendance.

As of June 1 2013, the MOOCs reached a cumulative total of 12,725 registered participants from 176 countries. While the MOOCs were designed mainly for evaluators in developing countries, they also attracted evaluators from developed countries: 31% hailed from developed countries. Despite concerns regarding low internet connectivity in Africa, the largest group of participants was from that region: 37% were from Africa, 22% from Asia, 22% from Americas (including Latin America & the Caribbean), 16% from Europe, and 3% from Oceania.

The MOOCs participants worked in a variety of organizational settings: 31% were from Civil Society Organizations, 20% from UN agencies, 16% from private sector, 15% from governments, 9% from universities and another 9% are listed as "others". The educational background of participants was also diverse with 61% having a master degree, 23% having bachelor's degree, and 9% having a Ph.D.

Most of the participants reported that they had an intermediate knowledge level (40%), with others reporting to be beginners (27%), advanced (18%), to have no experience (8%), and 7% were experts. Early estimates are that about 40% of registrants complete the entire course, pass the tests, and get the final certificate. Approximately 72% rate

the course as very good or excellent, and 94% declare they are planning to take future courses.

To give IDEAS Global Assembly's participants a sense of the content of the MOOCs, we briefly described a new book "Emerging Practices in International Development Evaluation" (Donaldson, Azzam, & Conner, 2013) that draws on many years of experience, successes and failures in development evaluation in Asia and Africa, and reflects recent work supported by the Rockefeller Foundation on Rethinking, Reshaping, and Reforming Evaluation. The concepts, frameworks and ideas presented in this volume contribute to ongoing efforts at rethinking, reforming and reshaping international development evaluation. Leading thinkers and practitioners in development evaluation, research and academia have recognized that development evaluation must evolve if it is to respond to the challenges of the 21st Century and play a meaningful role in social and economic transformation.

Many of the Global Assembly participants told us they were regular participants in our webinars and were fully engaged in the MOOCs. These discussions yielded many insights about how to improve the webinars, MOOCs, and MyM&E resources as we expand our efforts to develop evaluation capacity worldwide. These new initiatives include a new scholarship program for evaluators from developing countries to attend online evaluation workshops free of charge at the Claremont Evaluation Center (<http://www.cgu.edu/pages/465.asp>) and a range of initiatives to build National Evaluation Capacity. To address the National Capacity Development challenge, UNICEF and IOCE founded EvalPartners, the global partnership for National Evaluation Capacity Development (NECD).

So far 34 major organizations, including all the regional Voluntary Organizations for Professional Evaluation (VOPEs), multilateral organizations such as UNEG, UNICEF, UNWomen, UNDP and African Develop-

ment Bank, as well as donor countries (OECD/DAC EvalNet Task Force on NECD, Finland, Spain and USAID) and other major stakeholders, such as the Rockefeller Foundation and Claremont Graduate University, have joined the initiative. Stay tuned and learn more about EvalPartners, NECD, and 2015 as the International Year of Evaluation by visiting: <http://www.mymande.org/evalpartners>

References

Donaldson, S. I., Azzam, T. A., & Conner, R. (2013). *Emerging practices in international development evaluation*. Greenwich, CT: Information Age.

EvalPartners website, available at <http://www.mymande.org/evalpartners>

MyM&E website, available at <http://www.mymande.org/>

THE AUTHORS

Rob D. van den Berg

is the director of Global Environmental Facility (GEF) Independent Evaluation Office. He previously directed the independent Policy and Operations Review Department of the Dutch Ministry of Foreign Affairs and held various positions in Dutch development cooperation, including head of Dutch Development Cooperation at the Netherlands Embassy in Suriname and advisor to the European Commission in Brussels on development research. He has also chaired the Evaluation Network of the Development Assistance Committee (DAC) of the Organisation for Economic Cooperation and Development (OECD).



María Bustelo

María Bustelo, PhD in Political Science, is Associate Professor in the Department of Political Science and Public Administration at the Complutense University, Madrid (UCM), Spain. She has worked as an indepen-



dent consultant for Spanish public administrations and agencies, the European Commission, and NGOs in the field of programme evaluation, and in the areas of community development, health promotion, drug dependency prevention, and gender policies. She is also the director of the Master on Evaluation of Programmes and Public Policies (UCM), and member of the Board of Director Committee of the National Agency for the Evaluation of Public Policies and the Quality of Public Services. She leads the European QUING project team at the UCM (QUING: Quality in Gender+ Equality Policies -2006–2011-www.quing.eu; a research project supported by the VI European Framework Programme). She has a number of publications on analysis and evaluation of public policies and gender equality policies.

Stewart I. Donaldson

is Professor and Dean of the School of Social Science, Policy & Evaluation, and Director



of the Claremont Evaluation Center at Claremont Graduate University, USA.

His recent evaluation books include *The Future of Evaluation in Society: A Tribute to Michael Scriven* (2013); *Emerging Practices in International Development Evaluation* (2013); *Social Psychology and Evaluation* (2011); *Advancing Validity in Outcome Evaluation: Theory and Practice* (2011); *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* (2008); and *Program Theory-Driven Evaluation Science: Strategies and Applications* (2007).

Dr. Indran A. Naidoo

Director, Evaluation Office, United Nations Development Programme. Indran Naidoo is the Director, of the Evaluation Office, of the United Nations Development Programme (UNDP). The Office is responsible for the independent evaluation function of the UNDP, and produces global, thematic and country level evaluations across the UNDP, reporting to the Executive Board, management



at all levels, and stakeholders at the country and global level. Prior to joining the UNDP, Indran worked as the Deputy Director-General: Monitoring and Evaluation at the Public Service Commission (PSC) of South Africa, as well as Deputy Director-General: Leadership and Management Practices and Chief Director at the same institution. He was part of a leadership team that helped develop the oversight systems for this independent constitutional body, which was reconfigured during the post-apartheid and democratic era. Whilst at the PSC he helped advance the discipline and served on elected Board positions on the International Development Evaluation Association (IDEAS) and the South African Monitoring and Evaluation Association (SAMEA). He was the PSC conference leader for the Third PSC/African Evaluation Association (AFREA) Conference held in 2004 and a founding member of the SAMEA, which led to three SAMEA Biennial Conferences under his leadership (2007, 2009 and 2011). He has been a visiting faculty at the International Programme for Development Evaluation (IPDET) on evaluating governance since 2008. Prior to joining the PSC he was the Director: Monitoring and Evaluation, Department of Land Affairs, where he set up the country's first M&E system between 1995 and 2000. He holds graduate and post-graduate qualifications in English and Geography respectively from the University of KwaZulu Natal, post-graduate in Education from the University of South Africa, a Master's degree in Geography from West Virginia University USA, and a Doctorate in Evaluation from the University of Witwatersrand, South Africa. His doctoral thesis examined the role of M&E in promoting good governance in South Africa. He also holds executive management certificates from Oxford University and Harvard (USA) and Wits University Business Schools. He has presented several papers on development and evaluation, has written on the subject, and is the recipient of several awards.

Robert Picciotto

Robert ('Bob') Picciotto, (UK) Professor, Kings College (London) was Director General of the World Bank's Independent Evaluation Group from 1992 to 2002. He previously served as Vice President, Corporate Planning and Budgeting and Director, Projects in three of the World Bank's Regions. He currently sits on the United Kingdom Evaluation



Society Council and the European Evaluation Society's board. He serves as senior evaluation adviser to the International Fund for Agricultural Development and the Global Environment Fund. He is also a member of the International Advisory Committee on Development Impact which reports to the Secretary of State for International Development of the United Kingdom.

Rosalía Rodríguez-García

Dr. Rosalía Rodríguez-García has more than 25 years of experience in development and academia. Prior to her World Bank assignment as Team Leader in the World Bank Human Development Network, she led UNFPA's evaluation of a PAHO/WHO family health program, and USAID's evaluation of a health program in Central Europe (Poland, Hungary, Slovakia and the Czech Republic).



Marco Segone

is responsible for the Decentralized evaluation function as well as the National evaluation capacity development portfolios at UNICEF Evaluation Office. He is the Co-Chair of the EvalPartners Initiative and the UNEG Task Force on National evaluation capacities since 2009. During his 21 years in international development, he worked in Albania, Bangladesh, Brazil, Colombia, Niger, Pakistan, Switzerland, Thailand, Uganda and USA. In 2003 he was elected Vice-President of IOCE. Segone authored and/or edited several books and articles. In recognition of his contributions to the evaluation profession he was awarded the 2012 Alva and Gunnar Myrdal Practice Award by AEA.



Helen Simons

Helen Simons is Professor of Evaluation and Education, University of Southampton, a past president of the UKES and current member of Council, an academician of the Academy of Social Sciences (UK) and a Fellow of the Royal Society of Arts. Specializing in programme and policy evaluation, case study evaluation and the ethics of evaluation from a democratic and humanistic perspective, she has conducted numerous external evaluations and evaluation training internationally incorporating these values. She has also published widely in the field. Recent publications include Case Study Research in Practice (2009) Sage.



Daniel Ticehurst

is Project Director for Performance Management and Evaluation at HTSPE Ltd, a UK based Management Consulting Firm. He has vast experience in agricultural development, food security and the private sector (including 14 years of field assignments mainly in Africa).



Christine Wörlen, Ph.D.,

is a consultant on monitoring and evaluation of sustainable energy programs in developed and developing countries. She is working for governments, associations, foundations as well as international organizations. She is based in Germany. Before starting her consultancy she was the Head of the Renewable Energy Division of the Germany Energy Agency and before that a Programme Manager on Climate Change at the Global Environment Facility. She holds degrees in environmental science, economics and geography from Universität of Bayreuth and Boston University.

