



Evaluation Revisited

Improving the Quality of Evaluative Practice by Embracing Complexity

Irene Guijt
Jan Brouwers
Cecile Kusters
Ester Prins
Bayaz Zeynalova

Conference Report



ALKE



Learning by Design



WAGENINGENUR
For quality of life

Context,
international cooperation



Wageningen UR Centre for Development Innovation (CDI) works on processes of innovation and change in the areas of secure and healthy food, adaptive agriculture, sustainable markets and ecosystem governance. It is an interdisciplinary and internationally focused unit of Wageningen University & Research centre within the Social Sciences Group.

Through facilitating innovation, brokering knowledge and supporting capacity development, our group of 60 staff help to link Wageningen UR's expertise to the global challenges of sustainable and equitable development. CDI works to inspire new forms of learning and collaboration between citizens, governments, businesses, NGOs and the scientific community.

More information: www.cdi.wur.nl



Innovation & Change



Ecosystem Governance



Adaptive Agriculture



Sustainable Markets



Secure & Healthy Food

Context, international cooperation (Utrecht, the Netherlands) is a development organisation understanding itself as a social enterprise. We are innovators who use business principles that aim to achieve fundamental social change and generate revenue. Currently Context has 14 staff members, in addition to a number of external Programme Associates. We position ourselves between the development and academic communities, and try to facilitate dialogue between policymakers, academics and development practitioners. Context staff members and Associates have a sound academic background as well as working experience in development practice.

More information: www.developmenttraining.org

Learning by Design is the company name under which Dr. Irene Guijt operates. She works independently on assessing results in social change, such as through the [BetterEvaluation](#) initiative that is being beta-tested and the [Big Push Forward](#). She advises and researches on, and facilitates learning processes and systems in rural development and natural resource management, particularly where this involves collective action. Work over the past twenty years with multilateral organisations and many development NGOs has focused largely on strengthening critical reflective thinking to strengthen pro-poor development.

Evaluation Revisited

Improving the Quality of Evaluative Practice by Embracing Complexity

Irene Guijt
Jan Brouwers
Cecile Kusters
Ester Prins
Bayaz Zeynalova

Conference Report

March 2011

Evaluation Revisited

Improving the Quality of Evaluative Practice by Embracing Complexity

Guijt, I.
Brouwers, J.H.A.M.
Kusters, C.S.
Prins, E.
Zeynalova, B.

March 2011

Centre for Development Innovation, Wageningen University & Research centre; Context, international cooperation, Learning by Design

This report summarises the discussions and presentations of the Conference ‘Evaluation Revisited: Improving the Quality of Evaluative Practice by Embracing Complexity’, which took place on May 20-21, 2010. It positions these discussions within international debates on measuring development results, through additional insights and observations from several key evaluation events in 2010/2011 during which the May conference discussions were shared.

Cartoons

Auke Herrema

Orders

+ 31 (0) 317 486800
info.cdi@wur.nl

Acknowledgements

This report summarises the outline and outputs of the Conference ‘Evaluation Revisited: Improving the Quality of Evaluative Practice by Embracing Complexity’, which took place on May 20-21, 2010. It also adds additional insights and observations related to the themes of the conference, which emerged in presentations about the conference at specific events.

The May conference was organized by the Wageningen UR Centre for Development Innovation (Cecile Kusters), Learning by Design (Irene Guijt) and Context, international cooperation (Jan Brouwers). The Conference organizers are also deeply grateful for the funding support provided by: Cordaid, Hivos, ICCO, IDRC, IKM, LNV/Wageningen UR, Oxfam-Novib, Plan Netherlands, and Stichting het Groene Woudt. Without their support, the Conference would not have been possible or as successful as it was. Preparations were enthusiastically guided by the Advisory Group members: Ton Dietz, Sarah Earl, Zenda Ofir and Roger Ricafort.

The Conference organizers are very grateful to the participants of the Conference. They showed high interest in the topics raised, which allowed a fruitful discussion, exchange and reflection. Participants who shared cases and methods opened up by sharing the context of their work and seeking critique to improve their practice. The Conference organisers are also indebted towards the provocative and insightful keynote speakers, the superb chair, and the hardworking support staff and volunteers. Many follow-up discussions have since taken place nationally and internationally, demonstrating the great interest in the topics covered by the Conference.

Additional background is available at the Conference Website (<http://evaluationrevisited.wordpress.com>) and in the Conference Flyer.

Wageningen / Randwijk / Utrecht

March 2011

The Conference organizers

Cecile Kusters, Wageningen UR Centre for Development Innovation

Irene Guijt, Learning by Design

Jan Brouwers, Context, international cooperation

Table of Contents

Acknowledgements.....	iii
List of abbreviations and acronyms	vi
1 What is Contested and What is at Stake	1
1.1 Trends at Loggerheads.....	1
1.2 What is at Stake?.....	3
1.3 About the May Conference	3
1.4 About the Report	4
2 Four Concepts Central to the Conference	6
2.1 Rigour	6
2.2 Values	8
2.3 Standards.....	11
2.4 Complexity.....	13
3 Three Questions and Three Strategies for Change.....	19
3.1 What does ‘evaluative practice that embraces complexity’ mean in practice?.....	19
3.2 Trade-offs and their Consequences	22
3.3 (Re)legitimise Choice for Complexity.....	22
4 The Conference Process in a Nutshell.....	28
Appendix 1 – Programme	31
Appendix 2 – List of Participants	32
Appendix 3 – Summaries of Case Clinics	36
Appendix 4 – Summaries from the ‘Methods & More’ Market	74
Appendix 5 – References and Further Reading	94
Appendix 6 – Evaluation of the Conference	97

List of Abbreviations and Acronyms

CDI	Wageningen UR Centre for Development Innovation
Wageningen UR	Wageningen University & Research centre
AEA	American Evaluation Association
BRAC	Bangladesh Rural Advancement Committee
CBO	Community Based Organisation
CD	Capacity Development
CORT	Collaborative Outcome Reporting Technique
DE	Developmental Evaluation
DFID	Department For International Development (UK)
DGIS	Directorate General for International Cooperation (Netherlands)
EES	European Evaluation Association
ICCO	Interchurch Organisation for Development Cooperation
IDRC	International Development Research Centre
IKM	Information and Knowledge Management (IKM)
LFA	Logical Framework Approach
LIC	Low Income Countries
LNV	Ministry of Agriculture, Nature and Fisheries (Netherlands)
Logframe	Logical Framework
MIC	Middle Income Countries
MSC	Most Significant Change
NGO	Non Governmental Organisation
NONIE	Network of Networks on Impact Evaluation
OM	Outcome Mapping
PADEV	Participatory Assessment of Development
RCT	Randomised Control Trial
REPEC	Rigorous evaluative practice that embraces complexity (shorthand term used during the conference)
RMIT	Royal Melbourne Institute of Technology
SPARC	Promotion of Area Resource Centres (Mumbai, India)
SROI	Social Return On Investment
TB	Pulmonary tuberculosis
3ie	International Initiative for Impact Evaluation

1 What is Contested and What is at Stake

1.1 Trends at Loggerheads

How do we know what works? What kinds of information are convincing and why? And when are we satisfied that we know enough to decide and act? These questions lie at the heart of intense debates in recent years, as development effectiveness has come under ever closer scrutiny. Words like ‘impact’ and ‘accountability’ have emerged as powerful mantras (Ebrahim and Rangan 2010). One response has been methodological, with an enormous surge of interest in (quasi)

experimental methods that has led to many applications and policies showing preference for this methodological option. In the process, issues around rigour of evaluation practice have come to the forefront. And they are much needed, given the ongoing need for quality improvement in evaluation.

Though the mainstay for disciplines like medicine and education, (quasi) experimental methods¹ represent an innovation for the development sector (see Box 1). Examples of the proliferation of interest in this option are the NONIE² Guidance Notes on Impact Evaluation with its almost exclusive focus on this suite of methodological options, 3ie and its reference to this methodological option as best practice, and the growing pressure in bilateral aid agencies to use randomized evaluation to show what works in development. In contrast, many (local) development non-governmental organisations (NGOs) and community-based organisations (CBOs) still focus on non-experimental evaluation approaches – by default or design.

How can revisiting evaluation – given these sectoral innovations – help us be more rigorous, while acknowledging the complexities of change? And within that domain they have innovated considerably as well.

Understanding the impact of development efforts and how change happens has never been higher on the development agenda. Partly triggered by a critical global questioning of the relevance of development efforts, the interest in measuring development effectiveness is also embedded in a more paradigmatic tug-of-war.

Development as a process of social transformation is increasingly commonly accepted. Interventions that involve multiple nested processes, involving many players, with no guarantees of results are more norm than aberration. After decades of development as technical or economic change, the transformative nature of much of development is recognised. Effective development is being understood in many quarters as revolving around people-centeredness and institutional transformation. This makes it inherently complex, i.e. non-linear, emergent, unpredictable (see Box 2).

Box 1. Rising interest in RCTs

“Creating a culture in which randomized evaluations are promoted, encouraged and financed has the potential to revolutionize social policy during the 21st Century, just as randomized trials revolutionized medicine during the 20th.” Esther Duflo, Lancet Editorial, “The World Bank is finally embracing Science” (2004)

¹ In shorthand, these are often incorrectly referred to as RCTs (randomized control trials), which are only one methodological approach.

² NONIE – Network of Networks of Impact Evaluation.

Box 2. Knowing an iceberg, evaluating its tip (Patel 2007)

“Between 1998 and 2005 in Mumbai, 18,000 households voluntarily demolished their shacks along the railroad tracks so that the city could make long-overdue improvements to its vital railroad system which provides public transport for over 70% of all trips taken in the city. There were no confrontations, no coercion, no bulldozers, teargas or police. In a time of unprecedented numbers of forced, and even violent, evictions in the world, this was an unusual event. ... In terms of assessment standards, this project was clearly successful and its impact as a precedent clear. There were tangible, quantifiable outcomes.... There were partnerships involved, good governance, gender equity, and civil society participation – all the buzzwords we use to accompany numbers in describing ‘a successful project’. ... And yet ... this kind of assessment is unsatisfactory and even misleading. It only describes the tip of an iceberg. ... “We are happy to claim it as an achievement but find it frustrating that the larger part of this iceberg remains invisible. The real story ... is about the dense underpinning of values, principles, processes and relationships built over many years, and how they made it possible to seize the opportunity when it came along. ...”

However, in discussions on measuring development effectiveness, many powerful and resource-rich development agencies have zoomed in on methodologies suited to measure simpler, more linear interventions, as being ‘the gold standard’ for evaluation. Sometimes this is implicit, but increasingly explicit so as in the case of 3ie’s quality standards for inclusion in its impact evaluation database: *“Quality evaluations employ either experimental or quasi-experimental approaches.”* This emphasis has emerged particularly strongly due to the surge of interest in ‘impact’.

A hierarchy of method appears to have emerged, placing quasi-experimental options high on the ladder. Placing method above question runs counter to long agreed good evaluation practice, the key question of ‘what do we need to know and why?’ – would logically be followed by ‘and how can we best measure this under the conditions we have?’. Thus we face an emerging duality of evaluative practice: *“one the one hand, the methodologies of the economist with RCTs (randomised control trials) and counterfactual analysis (a comparison of what has happened with what would have happened if the intervention had never taken place) and, on the other hand, a group of alternative methodologies with their roots in the social sciences”* (Cummings 2010).

This methodological narrowing is occurring in a context marked by five trends (building on Brouwer’s keynote at the May Conference):

- *interconnectivity* beyond national, sectoral and professional boundaries through new social media but also through strengthening interdependencies on cultures, economies and scarce resources and a ‘globalisation of vulnerability’ (Ramalingam 2011);
- *uncertainty* with previously distant problems and events now touching on everyone’s lives, touching on them with increasing frequency and speed, and requiring action;
- an *emerging 2.0 society* that privileges results over hierarchies, enables multiple directions of information flow, and real time sharing and responding;
- fewer resources for international development through traditional channels that seek more security and certainty about results up front; and
- many more resources through new players and their paradigms (foundations, China, business) that is making aid a relative sideshow.

These trends represent deep institutional changes – and are themselves triggering further far-reaching shifts. ‘Development’ with its poverty focus over the past 30 years or so has changed, a point highlighted

by [Summers](#) in his comment on an evaluation of DfID's aid to India: it is not about poor countries but about equity and shared prosperity. The poor are not where we think they traditionally are but are mainly in middle-income countries: in 1990 93% were estimated to come from low income countries while now 75% come from middle income countries. This data is highly significant as it pulls the carpet out from the core logic of targeting development resources to the poor based on national statistics.

The trends are also feeding the parallel search for methodologies that can help ascertain 'certainties', as well as methodologies that deal with the 'emerging complexities'. It is in this dynamic that knowing what is appropriate when becomes crucially important, as well as knowing what we can ascertain with the evidence we have in hand. Lehrer (2010) comments on emerging discussions on empiricism: 'it is as if our facts were losing their truths.'

1.2 What is at Stake?

Much is potentially at stake:

'Those development programs that are most precisely and easily measured are the least transformational, and those programs that are most transformational are the least measurable.' (Natsios, ex USAID 2010)

The risk of a narrow menu of methodological options is the reduction of development to simple interventions, in order to facilitate its measurement. Quasi-experimental methods are well suited to study single-strand objects of study that occur in stable environments. These two mutually influential trends sit side-by-side in increasing discomfort for those who understand development as societal transformation. Easterly (2010) takes a practical stance on what is at stake at the ground, compared to what can be known from randomized evaluations:

"So it comes to how likely it is that different official aid agencies and NGOs are to make the money reach the end of the road. This is a bit different than whether different aid interventions "work" according to randomized evaluation (RE). Even if the interventions pass the RE test, how do you know that one hundred additional dollars given to one particular agency will translate into additional interventions?"

Given the complexity of context and of intervention, and the limitations of 'randomized evaluations', what are ***the options for evaluative practice that respect the complexity of societal transformation while fulfilling quality standards?*** And what can these methods do – and not do? Examples certainly exist of good evaluation practices that try to bridge these tensions. These options start by recognising that societal change is transformational and people-centred and, therefore, complex. This requires evaluative practice that is based on long-term thinking, accepts the need for adaptation en route. It sees change as non-linear, multi-actor and unpredictable.

1.3 About the May Conference

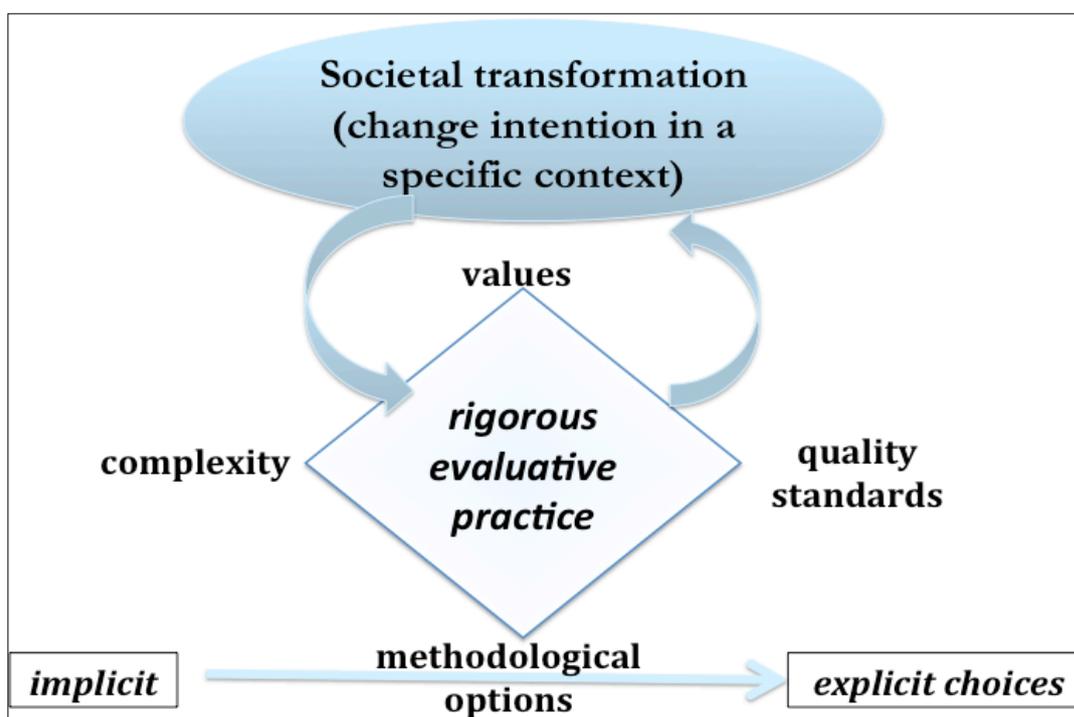
In May 2010, a conference was held in Utrecht (The Netherlands) that tackled the question set out above. Set up around critical reflection on case studies and specific methods, debates were held on rigour, quality standards, values, and complexity. The 150+ participants were driven by a common need to help evaluation better play its key role – that of offering well-informed insight for decision making for change.

Figure 1 illustrates the logic on which the conference was based. Given that much of development seeks *societal transformation*, we are dealing with a certain kind of change intention that is highly context and

time specific. *Values* inform the change intention and the choice in change strategies. These require *rigorous forms of evaluative practice*. Ideas of rigour need to recognise the implications of *complexity* of context, and sometimes complexity of intervention. Existing *quality standards* help guide such practice. Only then do we need to scrutinise the full range of *methodological options* in order to come to contextually-sensitive choices. Rather than an *implicit* methodological preference guiding evaluative preference, our choices need to be *explicit*. And for that we need to articulate our values – in the context of complexity, quality and rigour.

In the conference, the term ‘evaluative practice’ was a conscious choice. Understanding change, the route towards impact, and impact itself requires not just a one-off evaluation, or results-oriented monitoring, or adaptive innovation, or impact evaluation. It is the suite of evaluative practices that are needed to value what is happening – and learn from it and use this to improve our actions, our behaviour and our thinking for social transformation.

Figure 1. Rigorous evaluative practice



1.4 About the Report

This document is a ‘conference and more’ report. It has suffered from a multi-month delay in being finalised. This was a period during which the topic fuelled many heated debates. Conference organisers were present at the European Evaluation Society (EES³, Prague, October 2010), American Evaluation Association (AEA⁴, San Antonio, Texas, November 2010), and the GiZ conference on ‘Systemic Approaches in Evaluation’ (Frankfurt, January 2011⁵). In this same period, some bilateral aid organisations

³ <http://www.europeanevaluation.org/>

⁴ <http://www.eval.org/>

⁵ <http://www.gtz.de/en/unternehmen/32453.htm>

sometimes dramatically shifted their evaluation policies and norms. We have included additional discussions and key documents that occurred after the May 2010 event in this report to enrich and contextualise the importance and multiple dimensions of the debate.

Section 2 summarises the key concepts that formed the backbone of the conference: rigour, values, quality standards and complexity. Section 3 reflects on three crosscutting questions that shaped the debates and four strategies for change. At the request of conference participants, Section 4 is included, which summarises the conference process. The Appendices are an essential part of the report, as they contain summaries of the cases and methods that participants generously shared. These participants opened themselves and their experiences and methods up to critique, greatly enriching the discussions by providing essential context.

2 Four Concepts Central to the Conference

Four concepts formed the backbone of discussions during the May 2010 Conference: rigour, values, quality standards and complexity. All these concepts are contested and diversely understood. Their interrelationship, as described above in Figure 1, is crucial. *Values* inform all choices, objectivity is relational and relative but not the absolute quality behind which some evaluators hide. *Quality standards* define the professional code. *Rigour* is code for 'high quality' within the profession but diversely defined. And *complexity* refers to both science and discourse; relates to both context and intervention. It is very diversely understood. These concepts will be discussed in some detail here, with reference to conference discussions, cases and methods, plus other key references. In this section, two of the conference questions will be addressed:

- (1) What values and quality standards are needed within evaluative practice in order to understand and influence such change processes?
- (2) What are the strengths and weaknesses of innovative examples of evaluative practice that do justice to the complexity of change?

2.1 Rigour

During the EES discussions in October 2010, a director of evaluation of a European bilateral aid agency said that '*rigorous evaluation of the field of conflict prevention and peace building was not possible*', referring to the double-in-difference approach that had just been presented in a case study on vocational training where it had been used. Why was rigour so narrowly understood? What about other aspects of rigour like utility, external validity, method mastery, sense-making, or indeed substantiated methodological choice?

The heart of the matter with the methodological tensions is 'rigour'. Much of the spin in the often-heated debates relates to whose version of rigour dominates or that some version of 'rigour' is best. This debate is particularly strong in relation to impact evaluation, where 'rigour' of impact evaluation is increasingly narrowly defined in terms of a certain kind of statistically significant indication of difference with and without an intervention. This differs from the broader DAC definition of impact that seeks the "*positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended*" (OECD/DAC 2002).

By focusing on (quasi) experimental methods and measurable before/after and with/without difference, there is a privileging of the notion of internal validity to define rigour. Other aspects of rigour seem to have disappeared, such as external validity, reliability, and credibility. Terms like 'rigorous', 'hard data', 'evidence-based', 'scientific', 'technically more sophisticated', 'smarter', 'revolutionary' all imply an hierarchy in what is valid and permissible and legitimate. The rest is, by default it seems, inadequate. Alternative evaluation designs are delegitimized, and ignored. Worse still, those involved are accused of not being interested in 'the truth'. Words that imply a certain kind of rigour elevate one set of methods above another.

In this debate, methods have become more important than questions in determining quality of evaluation. Yet any method can only answer to certain kinds of questions. Many different types of questions need to be answered, and therefore many kinds of evaluation exist. For some questions and many contexts, quasi-experimental methods would not only be rigorous but quite useless. Rigour depends on context and question; in the words of Schwandt (1981) it is about relevance of method first and then rigour in its use.

The standards for rigour for one form of evaluation, such as those for certain kinds of impact evaluation, cannot be assumed to hold for all. We must separate out generic issues of rigour in evaluative practice from those that pertain specifically to impact evaluation.

The notion of rigour needs to be brought back to other key considerations of quality that have become undervalued in current debates. Discussions on rigour from a narrow methodological focus revolve around defining 'good design' in terms of a statistically valid logic of attribution. But beyond good design there are many aspects of rigour that need attention (cf Perrin 2010). And for this, Schwandt's classic 1981 article advocates that rigour requires 'intellectual honesty and applying standards consistently'.

In the May 2010 discussions, rigour was approached from the notion of principles rather than method. This implies a focus on clarifying, from one's value system, which principles matter most and guiding methodological design and implementation from that perspective (see 'Values' below and Box 3).

But any evaluation should meet certain core standards, including: be transparent methodologically, have warranted arguments and follow solid reasoning. Patricia Rogers referred during the conference to the four other aspects of rigour for all evaluations: systematic, empirical, sceptical, and technically competent. This leaves open, however, the question of what is 'technically competent'. Does it mean that all evaluators should know the entire toolkit of evaluation options – from regression analysis to participatory mapping – and then choose depending on the situation, as some have argued? Or is being good at some and knowing when these are appropriate also allowable? If it is the former, where does this leave those working in low resource contexts with limited access to in-depth evaluation training, and yet needing and wanting evaluative practice?

Box 3. Rigour or vigour?

Cummings in her [post-conference blog](#) raises an important question: should we beat economists at their own game of 'rigorous' evaluation or adhere to a quality that is more closely aligned to values about development: '*Vigour is to life what rigour is to death*'. Cummings' notion of vigorous '*has implications of strength, active in mind and body, energy and determination. It also reflects more the power and dynamism of social engagement and the importance of volition – roughly meaning intentions – which is key to many development interventions*'.

Other aspects of rigour that were debated in May included the following.

- Credibility for whom. Rigour revolves around ensuring credibility. The question then rises of credible for whom and what is valued by those groups or individuals. Those using the evaluations and those whose experiences inform the evaluations may adhere to different views of what is credible. For example, SROI (Social Return on Investment; see Case 4) enables people to articulate what they find important in each context, while other methods follow specific discipline-bound standards that make findings credible for that profession.
- Quality of sense-making process. An example of rigorous sense-making is evident in the CORT approach was developed by Jess Dart, with her colleagues at Clear Horizon. It is based on making the most of 'multiple lines of evidence' and subjecting these to serious review by a wide array of stakeholders – both those living with the changes and subject matter experts. The sense-making process includes two phases that ensure a double perspective analysis and public crosschecking. In this manner, many different views are incorporated and legitimacy of findings is widely shared.

- Quality of application, the mastery. No matter how solid an evaluation design is, the implementation can be flawed. Each method has its particular set of non-negotiable steps and necessary skills. A household survey can be done poorly, and has been often documented. But so can focus groups. Dart made the important observation that we all need to be serious about improving the quality of participatory evaluations. Rigour there is needed as much as we need to guard against a narrow description. And rigour includes honesty about application; accurately portraying what an approach can and cannot tell and where choices were made that compromised methodological potential.



Irrespective of why evaluations are undertaken, and who and how and when these occur, transparency is crucial. With evaluations being usually contested interpretations of various forms of evidence, making sure the reasoning is clear and evidence is accessible is crucial. This ensures accessibility of evaluations, increases their likely utility but also opens the door for improving their quality⁶.

Terminology clarification is important if we are to reclaim space for original, broader understandings of rigour. This includes understanding what robustness is and how this relates to rigour. Rogers shared this view during her keynote address: *“For me, robustness means one can withstand a few knocks. Resilient evaluators and evaluations can be challenged. Rigorous evaluative practice stands scrutiny – so when people say there’s a problem with methods, these methods can be defended. Evaluative practice is helping organisations to be rigorous about what and how they look at things.”* For some participants in May, rigour therefore includes being flexible and appropriately adapting the initial evaluation design to emerging circumstances.

2.2 Values⁷

Values are crucial for two reasons in evaluation. First, of course, the very mandate of evaluation is to value the object of study. As Davidson says ‘It isn’t called eVALUation for nothing’ (see Box 4). The point of evaluation is to ask about quality, importance, and value so that any conclusions can help the intended audience use it. Second, however, evaluative practice involves *the values that inform choice*, not only in design but in

Box 4. Jane Davidson on e-VALU-ation

One thing that makes something genuinely an e-**valu**-ation is that it involves asking and answering questions about quality or value. For example ...

- It’s not just measuring outcomes; it’s saying how substantial, how valuable, how equitable those outcomes are.
- It’s not just reporting on implementation fidelity (did it follow the plan?); it’s saying how well, how effectively, how appropriate the implementation was.
- It’s not just reporting whether the project was delivered within budget; it’s asking how reasonable the cost was, how cost-effective it was, and so forth.

⁶ A humorous example of this can be found at <http://genuineevaluation.com/the-friday-funny-is-hell-endothermic-or-exothermic/>

⁷ The conference discussion on values is based on those emerging from conference participants’ diverse experiences across the world. The AEA issued a set of principles in 1994 that guide evaluator’s behaviour in meeting the professional standards (see section 2.3). These are based on North American values. <http://www.eval.org/Publications/GuidingPrinciples.asp>

applying methods and in making sense of the findings. Choices are made about what to include and exclude, who to speak to and who not, what is considered acceptable quality of outcomes, how to wield a certain methodological tool or not, what is left in and out of analysis, and much more. Each of the cases spoke of the values that were informing their choices (see Table 1).

Values about whose voice count were particularly prominent during the May discussions. *CORT* is based on the value of engaging those people in evaluating the programs that directly touch their lives (see Case 7/Method 6). *Developmental evaluation* is based on the value of fostering innovation through a coaching type role for the evaluators (see Method 1). Whitmore mentioned the importance of sharing power, as a key design variable (see Case 1). PADEV is driven by the value of local voices and long term perspectives on development as integrated change (see Case 9).

Table 1. Values shaping choices in the cases

Case	Values shaping methodological choice
1. Canada	Address social justice issues Sharing power Openness for peer critique
2. Indonesia	Pragmatic and (cost) effective Engage with data providers and share results Make sense of PM&E processes Focus on both process and results
3. Somalia	Participation Context sensitivity Methodological appropriateness
4. SROI India	Mutual understanding and trust Transparency Empowerment
5. Aflatoun – (global network)	Participation Partnership
6. Cambodia	Quality engagement Flexibility Plurality of methods and data Identify strengths and weaknesses and work on it Encourage embracing mistakes
7. Australia	Values of Indigenous panel members, programme staff and key stakeholders were of highest priority
8. BRAC Uganda and Tanzania	Design RCTs whilst reflecting on ethic conditions Balancing community participation versus random assignment of communities
9. PADEV Ghana and Burkina Faso	Two sets of values are involved. One consists of the values and principles within the team of researchers, basically on what constitutes reliable impact measurement tools. The second and more important set of values is derived from the people involved.

Each of these values has strong methodological implications. Here are a few examples.

- *Local voices, local values.* *CORT*, PADEV, evaluative action research (Case 1) and SROI (Method 2) are all based on valuing local voices and local experiences of change. It is by listening to these voices that the diversity, direction and intensity of changes can be understood. Therefore, these methodologies are built around finding ways to allow diversity of observed changes to be expressed and to allow sharing and feedback through dialogues that actively include local citizens.
- *Partnership and sharing power.* Whitmore (Case 1) and Wilson-Grau (Method 1) describe approaches that explicitly share power to increase utility of the evaluation process. This has

implications for the time frame to suit internal rhythms, but also means that issues of importance are determined internally rather than only by external evaluators.

- *Valuing long-term change beyond attribution of effects to specific programs.* The PADEV methodology took as its time frame more than two decades. PADEV sought to make plausible contribution explanations and link this to the presence of specific actors over time. The method creators choose not to isolate a specific project or program as the starting point in the design but rather to derive understandings of attribution and contribution afterwards.
- *Accommodating changing contexts and needs.* Developmental evaluation intentionally refers to moving questions based on emerging issues, rather than adhering strictly to an initial intent. SROI also accommodates changes as emerged, rather than only as originally stated.

Discussions in May included comments on the value-free claims of empiricism. Some may say that it sounds dramatic and inappropriately subjective to say that rigour embraces values. Yet it is commonsense as people make the choices operating from personal value system (see Box 5): *We are self-interpreting, meaning-making beings, and the task of interpreting the value of our activities and actions is always contingent,*

Box 5. Heider blog (July 19, 2010)

“I came away from the conference with a renewed and deeper sense of the need for awareness: of our own values that influence how we evaluate and of what is unknown; now at the time of evaluation (in hindsight) and at the time people decided on their projects, policies, or whatever social transformation processes they embarked on.”

complex, contested and never finished’ (Schwandt, 2002: 14). This sits uncomfortably for those systems based on a ‘rationalist fallacy’, which Schwandt argued dominates in some versions of evidence-based practice. The “rationalist fallacy - believing that evidence alone is powerful enough to determine policy choices and that policy making is a purely intellectual undertaking” (Schwandt 2010) is strongly present in the values and, therefore, the logic and protocols of many aid bureaucracies. This makes, Schwandt argues, methodological choice a political process.

Added to the rationalist fallacy is tension around different understandings of truth and its knowability. Kell comments on this in her post-conference blog: *“I am currently working on the design of an evaluation of a large-scale, complex initiative ... As I, and my colleagues, think about the learning that we hope will come from this initiative and the nature of the evaluation as a kind of animator of learning, we are grappling with what sometimes seems to be different and mutually exclusive epistemologies, each with their own ‘truth claims’ (or for the more modest - partial or provisional truth claims). Some amongst us would wish to see an evaluation that can make claims about impact through addressing a set of hypotheses using experimental or quasi-experimental methods involving the use of a counterfactual. Others would prefer a more mixed approach that uses quantitative and qualitative methods.”*

What does it mean to work with the invisible in evaluation, concepts that we work with but carry within our work? How can we incorporate the notion of values into practice?

Much discussed was the importance of surfacing and considering values during the process. Indeed, without this, some contended, an evaluative process could not be considered rigorous.

An important observation from Kell concerned the underlying value of the conference itself:

“a tacit agreement was conveyed that all in the room subscribed to a set of ideas best seen in the statements in the programme that “the conference will further methodological democracy” through the “choice of examples and advocating for methodological diversity that do justice to the existing diversity of development modalities and contexts”; and that it will contribute to development by showing how “evaluation can influence societal transformation”. This felt to me like a ‘taken-for-granted’ assumption again; sort of like we are all activists in the same cause. So I felt like I was being recruited to a position rather than being strengthened in my abilities to argue for my position.”

An interesting value not much discussed in May is that of sincerity. [Rogers’ blog](#) on values related to the idea of ‘Genuine Evaluation’ identified being sincere as one among five important values (the others being value-based, empirical, usable and humble). Sincerity concerns a commitment by those commissioning evaluation to respond to information about both success and failure (those doing evaluation can influence this but not control it). Rogers continues: *“Sincerity of evaluation is something that is often not talked about in evaluation reports, scholarly papers, or formal presentations, only discussed in the corridors and bars afterwards. And yet it poses perhaps the greatest threat to the success of individual evaluations and to the whole enterprise of evaluation.”*

A powerful challenge to evaluators based on inherently different values came from keynote speaker Sheela Patel. She argued for the need to step away from a project lens when evaluating social change, given that transformation takes many years. Patel expressed her wish for learning to be tracked throughout SDI’s multi-decade history: “Northern organisations have very short memory spans” and “I’m often the historian of many Northern NGOs’ initiatives with us”.

“After 25 years we can talk of scale. In the first 10-15 years, we couldn’t. Scale was conjecture. But, Northern organisations have a short time span and sense of history. Everybody wants here and now. We work with 28.000 pavement dwellers. It took 20 years before the first pavement dweller got a house. ... We have to pretend that the process is a project – we don’t get funding, otherwise. If there is no honesty, the ‘blah, blah, blah’ keeps happening. How can we create consensus between our reality and what you see as Northern evaluators?” (Patel)

Patel’s experiences present an argument for including ‘expectations of ‘scale’ as a value that drives methodological choice.

2.3 Standards

The evaluation profession is clear about standards that need to be met. Evaluation associations, such as [AEA](#), require adherence to four standards: utility, accuracy, feasibility and propriety. These standards get us a long way but how known are they and are they enough for evaluative practice that embraces complexity?

Box 6. Examples of values in existing evaluation standards

U4 Explicit Values: Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.

F3 Contextual Viability Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.

Many May conference participants were not familiar with existing professional standards, a situation that appears to be quite common in development contexts. Clearly, familiarity with existing standards is an

important starting point. Knowing and using standards is akin to armour, said participants – they make us credible and defensible.

But even if they are known, do the notions of complexity, rigour and values require changes to existing standards? Some participants in May contested that existing standards accommodate all aspects of complexity, rigour and values. Box 6 gives two examples of values embedded in the 'Utility' standard and the 'Feasibility' standard.

Schwandt and Rogers, at the AEA panel "Rigor at Stake: Quality Evaluation of Change and Impact in a Complex World" in 2010 disagreed. Schwandt for example said that the term 'accuracy' was awkward, with a more useful standard being 'warrantability' – that is, claims being supported by evidence. The discussion that ensued at AEA highlighted the need to extend quality to different facets of evaluative practice: research design, data sources, and analysis (including looking for alternative explanations).

In May, discussions on standards included four topics – situational responsiveness, systemic-ness, complexity and types of standards.

1. Standards are situational and normative so whose standards are upheld and whose values count?

The existing standards are normative, e.g. one of the accuracy related standards: **'A2 Valid Information** Evaluation information should serve the intended purposes and support valid interpretations." In case of disagreement about core purpose, not uncommon, and validity – whose norms will count the most? Hence in each context, each evaluator, each commissioner of an evaluation still needs to define what 'appropriate' or 'sufficient' or 'useful' means. Participants in May said that standards as aspirations are important for quality guidance but need to be balanced with situational interpretations of standards, by asking whose standards are being used.

2. Standards on 'systemic-ness' of evaluation. Some participants argued that 'systemicness' is not present in the standards. How systemic an evaluation is will be important for utility and accuracy. One aspect of this is when an evaluation reduces its focus to the detriment of external validity and accuracy but tries to make a judgement call about a larger issue than the smaller, dislocated focus justifies. In January 2011, GiZ (the German development agency) held a conference on system approaches in evaluation where the question was discussed on what constitutes 'systemic-ness' of evaluation practice.



A powerful metaphor for this aspect of 'systemicness' is the distinction between the golden eggs – the interim outputs of development – and the goose – the longer term transformations being sought. By valuing only the 'eggs', the 'goose' is in danger of serious neglect (see cartoon). Patel cautions: *'With few insights about how to understand it and measure its [development] level of maturity and sustainability, external assessment processes are too rigid to understand these dynamics. Sadly, the goose is often killed due to lack of understanding'* (2007).

3. Standards on 'complexity' (also see 2.4 below).

Existing standards do not differentiate with respect to the nature of the intervention or context in which evaluations take place. Instead, terms like 'appropriate' and 'relevant' are expected to deal with this issue. But who says what is appropriate? This will vary depending, for example, on whether

you are evaluating complex change processes in complex contexts or more standardised and proven interventions in complex contexts. One idea might be inclusion of a standard that asks for clarity on the ontological premises and epistemological principles that drive the evaluation. This means that evaluations need to make explicit how reality is perceived (its knowability) and what this means for methodology.

4. Gold, platinum, diamond standards ... or none? In May, discussions referred to the notion of a supposed 'gold standard'. This term has been and is used frequently in the hot debates about rigour in impact evaluation. The notion has been seriously critiqued during Cairo, but also before and since (cf Deaton 2010; Khagram and Thomas 2010). It implies a reduced perspective on the evaluation focus and only one quality standard. The 'platinum standard' (Khagram and Thomas 2010) argues for holism and more contextually sensitive choices. Our own conference standard (see Figure 1, Section 1) was jokingly referred to as the 'diamond standard', explicitly integrating the four elements discussed in Section 2.

The critical issue in the discussion on standards is that values determine what we put in the standards – and making these explicit can help those working on evaluations to know what they are adopting when taking on a set of standards.

On this topic, however, Sheela Patel in her final comments warned us all about the notion of standards as examples of faddism: "On diamond standards I would like to say the following. I have a concern about best practices and diamond standards. If you use it like that it can become oppressive. Don't use words like 'the best, or terrific'. Those things will be become terrible in the end."

2.4 Complexity

Complexity has become the new buzzword in development discourses. Over the past 2 years, the term has been signalled in diverse contexts with high expectations. So what is the buzz around 'complexity'? Is it a powerful new perspective with practical implications or better viewed as a strategically useful metaphor to repackage marginalised ideas? Which elements within the extensive complexity science field might be of relevance?

The interest in 'complexity' seems to have centred on a growing recognition of the disjunction between the non-linearity and unpredictability of change processes and the protocols and procedures that govern development interventions that assume otherwise. The ensuing tensions are increasingly experienced as hindering poverty reduction and social change efforts.



The development sector is under much pressure to show impact, with its performance in terms of poverty alleviation under considerable critical scrutiny in recent times. In doing so, those active in the sector face various dilemmas. While development sector actors generally acknowledge the non-linearity of change pathways, linear program logic models, notably the logical framework matrix, still rule as the main instrument to track progress. Despite general agreement that change requires simultaneous actions by many actors at different levels, impact attribution is still considered possible and necessary.

And notwithstanding recognition that all is in flux and cannot be predicted, the aid system still demands precisely defined anticipated outcomes years ahead of expected realization.

So how can 'complexity' thinking help with these dilemmas? And how can we avoid it becoming an excuse to articulate expectations clearly and be accountable?

'Complexity science' is a loosely bound collection of ideas, principles and influences from diverse bodies of knowledge. It is a discovery of similar patterns, processes and relationships in a wide variety of phenomena. Ramalingam et al's (2006) study of this literature highlights that nine concepts are central in complexity science (see Box 7).

Do complexity and related concepts (see Box 7) help us think differently in relation to development interventions and the process of accompanying these through planning, monitoring and evaluation? How, if we are essentially dealing with systems thinking versus rational and linear planning logics, can we best learn about what we are doing in order to adapt? These questions bring us to processes, procedures, and methodologies that align with assumptions:

"...in decision-making at both policy-making and operational levels, we are increasingly coming to deal with situations where these assumptions [of order, rational choice and intentional capability] are not true, but the tools and techniques which are commonly available assume that they are." (Kurtz and Snowden 2003: 463)

**Box 7. Aspects of Complexity
(Ramalingam/Jones 2008)**

Features of systems

1. Interconnected and interdependent elements and dimensions
2. Feedback processes promote and inhibit change within systems
3. System characteristics and behaviours emerge from simple rules of interaction

Dynamics of change

4. Nonlinearity
5. Sensitivity to initial conditions
6. Phase space – the 'space of the possible'

Behaviour of agents

7. Attractors, chaos and the 'edge of chaos'
8. Adaptive agents
9. Self-organisation
10. Co-evolution

The dynamics in each context and the replicability of interventions can help evaluators define what can be expected and how to go about discerning change. The appropriateness of methodology then results from identifying the nature of what one is expected to assess. More standardisable interventions can be assessed differently from those that are context-unique and emerge in highly dynamic environments.

What norms of rigour would apply in different circumstances and for different types of interventions? Morell's new book on '*Evaluating the Unexpected*' (2010) is a plea for flexibility. Can an evaluation strategy that is not flexible in a complex situation be considered not rigorous? Many participants in May 2010 discussed that 'flexibility' is, indeed, a hallmark of rigorous application of evaluation method. While evaluators seek to clarify as many unknowns as possible, the more they encounter en route, the more adaptive the evaluation strategy needs to be.

Four issues surfaced in the May conference discussions.

1. *Varying understandings.* During the conference, 'complexity' led to much debate. Participants clearly had varying notions of 'complexity'. Complexity is still fairly new for many and the input on complexity was detailed and nuanced. Some used the term to mean 'difficult', others focused on the issue of emergence, and others on interdependence. Some found it academic,

questioning its practical value. Others found it hard to separate complicated and complex, thereby also questioning the ‘unknowability’ of the supposedly complex (see the third point). This variation in familiarity with central terminology limited in-depth discussions about what a complexity-oriented evaluative practice would look like (see Box 8).

As participants stressed, labelling and language are powerful and important. Participants suggested for example to set ‘rigour’ within a complexity evaluation framework, which would allow a contrasting of a conventional understandings of rigour with complexity rigour. Definitional understanding is essential, as is being aware how our ideologies shape definitions.

Box 8. Elephant in the room (thoughts by Catherine Kell, conference participant)

“... there was an elephant in the room – or perhaps a herd of them i.e. the *randomistas*! This resulted in a sort of asserting of “complexity” in the face of the elephants’ silent trumpeting of the “counterfactual”. I found Patricia Rogers’ articulation of the unhelpful ways in which complexity is used very helpful – but in my view, the concept of complexity was not made clear enough to really equip participants to argue strongly for why (rather than assert that) a diverse range of approaches is needed and is better, and what it is that these diverse approaches can offer that an RCT can’t offer. And because the experimental approach and its increase in influence were only really hinted at, the broader reasons for why this strong argument is needed were also not clear. So there was a kind of celebration of complexity, but without sufficient substance to it. I was left feeling that ‘taken-for-granted’ assumptions were being confirmed rather than understandings being deepened, challenged or sharpened. There are a number of economists working with only quantitative data who are articulating very robust critiques of RCTs and analyses of their limitations, a recent one coming from the World Bank itself.

2. *The link to evaluative practice.* Frameworks such as that of Cynefin can help to make sense of the sense-making inherent in evaluation (see Box 9). It makes clear the need for different methodological choices for different situations. It challenges the notion of good/best practice ambitions of many development organisations, as only of value for certain types of situations and issues (the ‘simple’ and ‘complicated’ domains). Furthermore, it highlights the need under conditions of complexity for a more adaptive mode, akin to the intentions of ‘developmental evaluation’. Real-time evaluative practice stays close to unfolding intervention so that it is part of it, or longitudinal study of impact evaluation.

Box 9. Seminar “Navigating Complexity” (CDI 2008)

“The notion of four (five) domains *explains and legitimises differences in learning and accountability* approaches. This gives space for an explicitly differentiated P, M&E framework, as long as we recognise that most projects have bits in all domains. Each domain (level) requires own learning/M&E process, and therefore legitimises diversity. It opens up space beyond the “one solution fits all” mentality that often abounds. Seeing the differentiated domains helps to manage expectations in terms of learning processes, and fairer accountability for different types of ‘outputs’. By opening the door for allowing ‘complexity’ to be recognised, it appears to open up space for *dealing with seemingly less tangible* issues, changes, and processes.”

But keeping it practical is essential. Sandip Pattanayak shares his insights: “The take-home for me was the difference between complicated and complex. This conference introduced me to complexity theory. However, the fact remains that improved understanding of what is complex is not going to make complex, simpler. Only appreciating what is complex will not help.”

3. *The level of the ‘unknowable’ with which we are dealing.* Is the unknowable only in the detail (true emergence) and will we always be able to ‘know’ some things ahead of time, or is the complex fully emergent? One participant discussed the tension between two aspects of the

complexity concept – path dependency and emergence, with implications for evaluative practice (see Box 10).

Box 10. How unknowable is the supposedly complex?

Participant: “I’m wrestling with complicated and complex. I know they are radically unknowable in advance, in the detail, but aren’t there structural things you can know in advance – the factors that create the environment in which things emerge one way or another? There are bottom up emergent details but there are also social trends that will lend to certain kinds of outcomes – for example David Burn’s tuberculosis research where he shows that there are certain consistent social arrangements that show up in the TB (Pulmonary tuberculosis) epidemic. If we don’t acknowledge these then our complexity approach will end up being an apology for neo liberalism, where ‘anything goes...’.”

Dr. Rogers: “TB is a complicated issue. Dr. Burns is interpreting – he has expertise, he is doing detailed analysis. This is thoughtful and we need more of this – causal analysis is often just simple, so this is a good improvement on this at least. If we just moved to incorporating the complicated domain at least, this would be a good improvement in evaluative practice.”

4. *Degrees of complexity and its recognition.* How much is truly ‘complex’? This question is important to clarify. And how do we recognise it? If it is correct that different contexts and different types of interventions require different methodological choices, then it is critical that we know how to discern the simple from the complex, and the complicated from the complex. This is a particularly difficult issue, with the example of cooking from a recipe considered ‘simple’ for some and complex for others.

Related to this were discussions on whether the methods and cases discussed in May actually dealt with complex contexts or interventions. Participants agreed that the cases varied considerably in their relation to complexity. In the Australian cases, questions were raised about whether the context was one of complexity or whether the evaluation team was dealing with complicated issues. In the case of Cambodia, was the intervention not just complicated? Table 2 comments on the degree of complexity of context and intervention, and the ability of the method used to recognise and value the complexity.

Table 2. Complexity in the Cases

(Note: For methodological aptness a scale is used: +++: very apt; ++: satisfying; +: contributes to understanding but not satisfactory; and 0: apparently low methodological aptness. Assessment made by conference organisers.)

Case	Complexity of Context?	Complexity of Intervention?	Methodological aptness to embrace complexities at play
1. Canada	Yes: different social realities	Emergent	++ Combine action research with appreciative inquiry
2. Indonesia	Yes: development of value chains that allow access to small producers	Yes, applying OM in dealing with messy value chain partnerships	++ Chain Intervention Framework combines outcome mapping and LFA
3. Somalia	Yes: fragile state	Yes, a continuous process of negotiation was implemented to ensure cooperation of various parties. The Somalian context is highly dynamic with interactions within, between, and among subsystems. So difficult to plan.	+ The methodology and process of the evaluation was quite indirect, starting first with practice rather than with theory. The deliberateness and consciousness about values, approaches and standards reflect the degree of professionalism.
4. SROI India	Partly: actors bring in their perspectives	Programmatic approach with a focus on sustainability and self-empowerment	+/0 SROI in its original concept applies linear ToC. M&E does include registration of unintended results and events
5. Aflatoun	Not clear	Openness for emergent requirements	+ Mixed methods to adapt as new needs and objectives arise
6. Cambodia	Yes, fragile state with governance issues	The assignment was pluralist, evolutionary and iterative.	++ During the pre-pilot phase, the project valued the process quality and engagement of stakeholders more than the initial findings. Some fundamental questions were asked like: "Why is this being conducted?"; "Who will benefit from it?"; "What changes are envisaged and how will the changes occur?"
7. Australia	Limited complexity	Outcomes not predefined, intangibles and emergent results	++ The "Collaborative Outcomes Reporting approach" has been successfully used with programs with higher degrees of complexity
8. BRAC Uganda and Tanzania	Yes: trying to achieve sustainable positive changes in the lives of adolescent girls requires involvement of various actors and systems (health, education, economy, gender norms)	Yes: starts from complex situation of an area with all its shocks and trends and interventions by diverse actors. New programme elements are added based on emergent understanding, like human rights and legal education	+ RCT-inspired design required methodological adaption in initial programme years and challenges remain to be addressed.
9. PADEV Ghana and Burkina Faso	Yes: impact understood as results, over decades, emanating from activities implemented by various actors and influenced by factors like yearly changes in weather patterns, world economic trends and political variability.		++ Careful documentation, use of an external reference group and other feedback mechanism especially from the involved population with an open mind for inter-subjectivity increases the validity and replicability without compromising the participatory and holistic approach

Table 3. Complexity from the Methods & More market

(Note: For methodological aptness a scale is used: +++: very apt; ++: satisfying; +: contributes to understanding but not satisfactory; and 0: apparently low methodological aptness. Assessment made by conference organisers.)

Method	Methodological design	Methodological aptness to embrace complexities at play
1. Developmental Evaluation	DE is an evaluation approach, assisting social innovators working in complex environments.	+++ This is the focus of DE
2. NGO ideas impact box	Mixed methods are applied to address assessment of changes	++ Toolbox and Tiny Tools are meant to combine continuous observation with occasional application. They allow for a long-term, systematic assessment of outcome and impact. 'Tiny Tools' can also be used for short-term evaluations to get community perspectives in a more validated way
3. Theory of Change and complexity	Way of looking at interventions with complexity central.	+++ One of the key messages of this book is the importance of taking into account the nature of the intervention when developing, representing and using program theory. The book highlights aspects of simple, complicated and complex interventions that might need to be addressed when developing, representing and using program theory.
4. Monitoring changes in capacity: 5 capability model.	The model provides a holistic assessment of capacity rather than a reductionist view. But it proves to be quite difficult in its application and needs further testing.	++ The 5 capability model clarifies what is being assessed, by providing nuanced vocabulary to address capacity at different levels (individual, group, organization), and societal. It introduces broader questions of legitimacy, relevance and resilience. The non-tangible issues in the framework resonate for practitioners. But it requires a high level of analysis to understand interrelationships among different elements of capacity.
5. Social Return On Investment	SROI in its original version is not difficult: 9 steps are identified which produce a SROI report.	+ SROI departs from linear ToC, but evaluation is open for different perspectives and experiences
6. Collaborative Outcome Reporting Technique	CORT uses a participatory process whereby information is generated and analysed through five phases that culminate in a report. Mixed methods are used for participatory involvement in all five steps.	++ CORT is especially useful when a programme has emergent or complex outcomes that are not fully defined at the onset of a program. For this reason a theory of change is refreshed at the start of the evaluation process. In addition qualitative inquiry is used to capture unexpected outcomes and deliberative processes are used to make sense of the findings.
7. Measuring complex systemic changes	Mixed methods to measure changes and impact	++ An impact measurement & learning system that is defined in terms of "empowerment" and measured consistently and iteratively over a longer period of time. The visualisation of the program's theory of change and mapping out its system of indicators helps partners to get their heads around the system and make sense of complex data.

3 Three Questions and Three Strategies for Change

This section returns to the remaining questions that guided the conference, providing initial ideas for answers. These questions are:

1. *What constitutes 'robust evaluation' for societal transformation, in terms of core values and quality standards?*
2. *Where do trade-offs occur in quality standards and core values? What are the consequences of these trade-offs for evaluation findings and for development?*
3. *What needs to shift in development in order to make possible high quality evaluative practices that influence societal transformation?*

3.1 What does 'evaluative practice that embraces complexity' mean in practice?

While complexity is perhaps an essential entry point that is forcing new questions of evaluative practice, it requires the interplay between the other three factors to answer the question. What are the implications for evaluative practice if one considers the implications of a complexity lens on the values, standards, and rigour that guide evaluation practice?

Patricia Rogers offers the image of a hand with five fingers for the practice we are seeking:

1. Clarifying and negotiating values;
2. Research design;
3. Data collection/retrieval;
4. Analysis, synthesis; and
5. Sense-making.

Box 11. Collingwood and 'rigour in practice'

Collingwood explains her understanding of 'rigorous evaluative practice that embraces complexity': "It is in the process of understanding together the choices made, that the change theory generating the programme is revealed. And so a rigorous evaluation practice seeks to evaluate initiatives on their own terms. It means trying to understand intentions and not only plans, choices and not only outputs. ... To my mind, striving to make a visual diagram that more fully represents an integration of the planning, unfolding and sense making in each programme intervention, would be rigour in practice."

In each of these, we need to *make choices explicit* (see Box 11). Evaluative practice that embraces complexity requires clarity about the elements jokingly referred to during the conference as 'the diamond standard' (see Figure 1): conscious and explicit reference to values, definitions of rigour through all evaluation tasks, and adherence to quality standards as informed by one's own values. These elements, together with clarity about what kind of - and degree of complexity one is dealing with, shape the choices.

Group discussions led to a set of *recurring assertions*, not all with consensus, about what constitutes evaluative practice that embraces complexity. These are described below, compiled from discussions during the two days, as issues that evaluators who want to base their practice on complexity of context and intervention might usefully consider.

Values. 'Values' are clearly essential to rigorous practice: values that underpin the choices made in evaluative practice, and taking into account different values in the evaluation process. Participants were firm about values as a non-negotiable and explicit condition of rigorous practice. Values can shift and keeping abreast of any changes is part of rigour. Being aware of the multiplicity of interconnected values, including one's own, requires time. We also need ways to surface these values that are often left tacit.

Jess Dart explains the values that underpin the CORT methodology and how this defines its quality: *“The focus on quality is largely associated with process quality: ethical conduct; culturally appropriate methods; ownership of the evaluation process; ensuring that the evaluation provides credible but useful recommendations to inform the next phase of the program. Interviews are usually taped and transcribed. Data is double analysed by participants at workshops and by the consultants using thematic coding.”*

Sense-making. Often not explicitly part of definitions of rigour, discussions stressed that rigour extends to analysis. This is particularly in conditions of complexity where multiple competing explanations of why change has happened exist. Part of this rigour means engaging those living with the changes in the sense-making and analysis, and not simply seeing them as data sources. This assertion does not come from a naïve notion about local knowledge but rather recognition that much understanding is tacit and that local eyes can help to verify explanations of change.

Agility and creativity. Societal change takes time, and in the process, much shifts in context and intervention. Evaluative practice needs to be agile to accommodate this. At EES (European Evaluation Society) conference in October 2010, this point was evident in presentations about two quasi-experimental impact studies by the French bilateral aid agency. The rigidity of method meant that, though contexts changed and the development intervention shifted, no changes could be made in the evaluation method, which resulted in very little useful information. This does not disqualify this method but stresses that in conditions of high volatility or where the interventions themselves are likely to change, other approaches may be more useful. Any changes in the evaluation processes themselves should be solidly documented. Why were changes made, what were/are the consequences? This makes the adaptive management of the evaluation process stand up to scrutiny and therefore more rigorous.

Discussions in May repeatedly stressed the need for good evaluators to use methods flexibly over time, thus requiring ongoing critical awareness. This requires flexibility on the part of evaluation commissioners to allow approaches that are appropriate for anticipated levels of complexity.

Focus. What should one look at in an evaluation process? What is the limit of what is considered? Discussions in May considered how much a rigorous evaluation would need to look at ‘the whole’. By this, participants meant not including all possible aspects, which is clearly impossible. But rather locating evaluations to the larger transformation being pursued, and not only to the narrow implementation of a plan of action. This comes back to the ‘golden eggs’ or the egg laying goose. With which perspective do evaluations look, assess and pass judgement? Aspects considered crucial to touch on the agency of people are power relations. Negotiating evaluation questions with stakeholders was mentioned as important to ensure utility and focus, both of which are essential for methodological choice.

In discussions, people also commented repeatedly on programme theory – ‘the devil is in programme design’. Social transformation, if conceived too rigidly, cannot evolve as contexts shift. Hence, the call for an explicit theory of change but one with clarity on assumptions and areas of uncertainty or multiple options, to accommodate emergence and non-linearity, with short cycles to reflect and adapt. This requires a focus on methods to explain what is changing, not just detect change, as well as methods to deliberate on the evidence – process and analysis.

Blending monitoring and evaluation. If complexity asks for evaluation to be more adaptive and ongoing, this suggests the value of shifting towards evaluative practice that emphasises more reflective practice and organisational learning consistently and goes beyond one-off contract-focused studies. Impact-oriented monitoring takes on importance. Participants encouraged more cross-fertilization between impact evaluation and organisational learning. As was discussed in the session on PADEV (Case 9):

“Complexity is related to the nature of the social systems we work in and try to influence, the problems we seek to alleviate and the arrangement through which we do so. From this case, but also from the keynotes of Patricia Rogers and Maarten Brouwer, it becomes clear that complexity-based evaluation has two sides. On the one hand intervening in complex systems requires more short-term and contextual evaluations for on-going learning, improvement and adaptation than mainstream evaluation allows for (evaluation as reflexive monitoring). On the other hand, impacts should be addressed more systemically (including the interaction between interventions and on-going social change) and long-term as apparent from the PADEV methodology. Both types of evaluation have widely different implications for the position of the evaluator, its relation to (on-going) planning, learning and intervention and the methods that should be used.”

Inclusion. Process is important; it is not a detail. It shapes evaluation findings and its utility. Evaluative practice that embraces complexity needs to focus on both result and process in cooperation with partners, instead of participation of partners. In particular, joint sense-making was mentioned in discussions as important for rigorous evaluative practices, in ways that allow multiple perspectives to inform the findings while being mindful of power relations among participants.



Learning. Dealing with the ‘emergent’ of social transformation requires ongoing deliberation on what seems to be working or not and what options exist. So rigorous evaluations are defined by their ability to help chart one’s way forward, meeting the learning needs and processes of the stakeholders involved. In evaluative practice, therefore, different learning needs need to be negotiated and methods for enabling learning are needed. And last but not least, building trust is germane for a learning experience to be possible, a value that needs specific attention and with methodological implications.

Standards. The existing AEA standards are valid and essential. They guide quality and make our work credible. However, the standards need to be balanced with situational interpretation of standards. We need to be asking: whose standards? We need to relate them more explicitly to the nature of contexts and interventions. We are not working on delivering material goods but on social transformation. Utility

becomes focused on whether an evaluation delivers learning rather than if it only delivers accurate findings (see above). In terms of those involved in evaluation process, a standard that emphasises plurality of perspective needs to be championed. To decide if new standards are needed to recognise complexity of social change, we can draw on fields that have always recognised complexity for inspiration, for example, leading to a standard that relates to the innate interconnectedness of change. Finally, a standard about honesty is needed that is clear about the methods and their assumptions and limitations, and, therefore, what can be claimed of the findings these lead to.

Methodology. Methodology has to do with choice in terms of expertise, time, scope of terms of reference, and context. For evaluation methodology to embrace complexity, we need to stay focused on the social transformation. Mixed methods become inevitable, given the nature of what one is looking at, requiring multiple perspectives, with relevancy and feasibility but also values being determinants of choice. However, some participants felt that the qualitative aspects of measuring need more weight and

importance. In many applications of mixed methods, more importance is given to quantitative methods. No one contested the need for both.

Importantly, methodology has to be open to see the surprising, able to read unforeseen realities, combined with rigorous methodology for sense-making.

3.2 Trade-offs and their Consequences

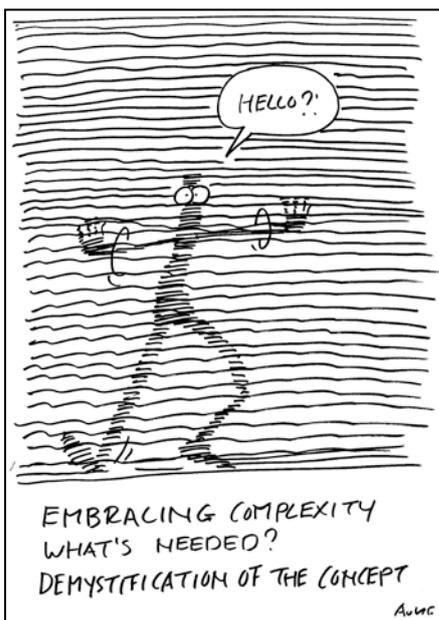
During the conference, not enough discussion was held explicitly on the topic of trade-offs. And yet every evaluation process entails multiple trade-offs. Where do *trade-offs* occur? A handful of examples emerged in the cases and methods.

More statistical rigour means that what one can assess is likely to become more narrow and social inclusion in analysis is less possible, greater rigour in social inclusion may mean trade-offs in who decides when what happens, more accurate portrayal of the causal linkages at play may mean less likelihood of ascertaining specific attributable impacts. A well known trade-off is that of evaluating a collective effort where disentangling specific attribution to final results are unlikely to be possible. Retrospective construction of a baseline may be more affordable but recollection bias creeps in. Doing a statistically credible study when budgets are limited may mean other contextual studies cannot be funded, leading to overall difficulty in making sense of discerned changes.

Simon Bailey of Aflatoun described how trade-offs are continuously made and shape the quality: “The translation of this [Aflatoun’s] approach into practice has had both successes and challenges. Challenges include maintaining quality across all projects, balancing different and sometimes competing organizational objectives, ensuring partner take up, and facilitating organizational and network learning. Successes have come from flexibility in developing and adapting approaches, finding and trusting interested individuals, as well as publicizing results.”

3.3 (Re)legitimise Choice for Complexity

During conversations in the May 2010 Conference, participants identified three areas for action: (1) documentation, (2) own practice and (3) sharing. In addition, a considerable research agenda (4) is also emerging that was briefly discussed by one sub-group and has since been pursued energetically. These four areas are described below. Some ideas are strongly aspirational and new – enduring difficulties were also listed. During the May discussions, people stressed the need to invest in *strategic* opportunities that can genuinely shift the debate and generate more legitimacy for rigorous evaluative practice that embraces complexity.



The case needs to be built for choice, it seems. Though contextually responsive methodology is widely accepted in professional evaluation associations and embedded in the agreed competency standards; a methods-driven hierarchy is increasingly dominating the discourse and dictating policy directions.

3.3.1 Action Area 1. Building the case for choice

Clarity about terminology is critical. Unclear or inconsistent terminology plagues evaluative practice. Terms like rigour, validity, quality, attribution, methodology, evidence-based practice, social change, transformational development, and complexity are used in a multitude of different ways. Even a term like ‘innovation’ is unclear – what is new for one, such as RCTs or MSC, is not for another. While unlikely to achieve a common understanding of ‘rigorous evaluative practice that embraces complexity’ we need to at least explore the diversity of understandings.

In particular, people stressed the need not to hide behind the term ‘complexity’ to refer to what was simply ‘difficult’ but to seriously understand what the term meant and its implications in complex situations. As labelling and language are powerful and important, a suggestion was made to differentiate between the definition and standards for ‘complexity’ rigour versus ‘conventional’ rigour.

Importantly, the discussions and terms need to be grounded. Fuzzy or theoretical concepts need to be demystified for people so they can understand them and work with them.

Rethinking evaluation. Discussions in May returned to the importance of viewing evaluation not as one-off studies but as accumulative evaluative practice, which requires rigorous programming. In the process deeper change the intervention is supposed to bring and the dynamic context in which it sits. We still face a strong split between evaluative ‘proving’ and evaluation-based improving. Participants are largely active from a learning perspective, and recognise that accountability is crucial. But the question remains of how we can have legitimization and learning within the same system, and whether certain methods favour one or the other purpose. Zaveri, a conference participant, commented in her post-conference blog:

“To then say that we can identify what that single intervention does or does not do is simplistic and arrogant, and ignores the many ripples of change it has produced. The conference reminded us of the need to address these complex ‘ripples’ and that we as evaluators must first acknowledge their existence and then make sense of it using our evaluation tools, methods and approaches. As one of the speakers, Sheela Patel from India mentioned, we ignore the deeper changes and are satisfied by evaluating the tip of the iceberg. What is worse is that we consider the tip of the iceberg evaluation to represent the whole iceberg. Such evaluations serve the narrow needs of budgets and timelines, selectively (sometimes erroneously) identify effects but worst of all, lose out on evaluating the richness of the human change that has occurred.”

Rigorously documented examples. Several participants commented on the need to document, rigorously, examples of good evaluative practice that embrace complexity. Few examples exist, as participant Catherine Kell stressed: “[Those working with (quasi) experimental methods] present a wide array of completed studies with measurable results

Box 12. The BetterEvaluation platform

BetterEvaluation is an international collaboration to improve evaluation practice and theory by sharing information about evaluation methods - in particular methods that are under-documented and invisible, such as those that address the complexity of social change, including non-experimental approaches to impact evaluation. The project consists of a searchable website of methods and network members who contribute content to the website and provide peer/expert advice.

More information is available at <http://betterevaluation.org/>. This project helps to showcase diverse methods of evaluation and examples of evaluations that have effectively used them, and link people working on similar challenges. The project is being undertaken by the Royal Melbourne Institute of Technology (RMIT University, Australia), the Institutional Learning and Change initiative of the Consultative Group on International Agricultural Research, the Overseas Development Institute, and Pact. *BetterEvaluation* is being coordinated by Patricia Rogers (keynote speaker in May), with Irene Guijt (conference co-organiser) as a project member.

that take on lives of their own. An alternative approach needs something similar, and I am not asking for measurable results here, but for completed accounts that simply say: "this is what we did, this is what it showed, and this is why understandings of complexity and emergence are important", for example. I think it is important to demonstrate the work rather than talk about values and standards or state what "should" or what "needs" to be done. Descriptions rather than prescriptions can better prove the point."

These cases need to systematically and thoroughly discuss the methodological choices made in order to be of use for others. Initiatives such as 'BetterEvaluation' (see Box 12) may help to move forward with this.

Articulating choices and tradeoffs. All case and method presenters made it clear that choices had to be made and tradeoffs accepted. In the interest in enhanced ownership, the action research on activism (Whitmore) shaped the calendar of work. Abebual discussed the choice for a comparative trial, which was so costly that BRAC decided against complementary qualitative studies. In Aflatoun, the education network, ideas can only be suggested from the central secretariat but not enforced.

Ibrahim and Rangan (2010) for example, point to how organizations need to start (re)conceptualising the scope of their evaluation practice. Currently all organisations stretch to all levels along an impact chain – or are forced to do so by bilateral aid agencies, irrespective of appropriateness. They say:

"Our normative argument is that it is not feasible, or even desirable, for all organisations to develop metrics at all levels on the logic chain. The more important challenge is one of alignment: building systems and structures for measurement that support the achievement of organizational mission, especially the goals that an organisation can reasonably control or influence. We contend that organisational efforts extending beyond this scope are a misallocation of scarce resources. (pg 4)."

Choices and their related tradeoffs are part of rigorous documentation of evaluative practice. Transparency of method greatly increases the potential for evaluations to be legitimate. Discussions during AEA with staff of 3ie showed that many evaluations that are not (quasi) experimental are not included in the database simply because the methodology is inadequately described and it is therefore unclear how findings were derived.

Meta evaluation to enhance credibility. Meta evaluations can greatly enhance the credibility of a certain set of ideas, development approaches, and of methods. However, to do meta evaluations of, for example, participatory management of natural resources, an adequate and comparable number of evaluations are needed.

3.3.2 Action Area 2. One's own practice as opportunity for change

Change starts with individuals, notably ourselves. And many opportunities were listed for where we can use our own practice to be the changes we are keen to see more of. A participant suggested the value of becoming critical historians of our own initiatives. By documenting clearly and sharing, we can contribute to enhancing the legitimacy of approaches to evaluation that are mindful of complexity.

Collaboration with clients – whether government or international donor. These debates are not about an either-or, us-them situation. Clearly there is much common ground and shared needs, and it requires collaboration to jointly improve. Together, openness for other ways of planning and being held accountable are needed to deal with the less linear that we encounter. Rigorous programmes are needed, which embrace complexity! As Patel urged:

"We act like a community of concerned people whose job it is to help measure change and look at ways of how change can be measured. That is very heartening. But if you are such a group you have a responsibility to help both those you evaluate but also those who commission the evaluations. You need to involve and convince them. The ultimate compliment that anybody can give you is that the method that you use is being internalised by those you used it with."

Risk taking and negotiating choice. The space given by clients to take risks is important but must be accompanied by the willingness, courage and open-mindedness of individuals to use that space. We can be bold and negotiate with funding agencies, taking a proactive stance in pursuing rigorous evaluative practice that embraces complexity. Negotiating scope at the beginning of assignments includes thinking through evaluation objectives of evaluation itself to go beyond the often limited focus on projects and push for a focus on the social transformation in question. We need to invest in tracking learning across much longer time-scales, and before and after the time/space boundaries set around projects. And if time and money are insufficient to do a rigorous job, then we need to say so and not take up assignments that can only lead to poor quality work.

Working on methodological gaps. Field building is needed. For example, we do not know well enough yet where and how to look for unexpected change. Nor do we have enough more real time approaches for short feedback loops, or ways to systematically surface different values during evaluation processes. To understand social transformation, evaluations can improve their analysis to come to counterfactuals or alternative explanations, while investing in ways that increase citizen participation in evaluation so it becomes a public good. How can the much requested (policy) prescriptions be accompanied with enough rich description to do justice to the context-specific relevance of the findings?

Focus on mastery of evaluation skills. No matter how wonderfully designed an evaluation process might be, the people who implement it will determine its ultimate quality. Evaluation requires *rigorous use* of methods, irrespective of which these are, and therefore balancing the bird's eye view with a focus on important detail. Evaluators need to be able to use and combine different methodological approaches, or at least know when what is appropriate. Flexibility is needed without falling into the trap of being opportunistic.



Participants noted a predominance of more participatory approaches among the cases and methods in the conference. Within this set of approaches, there is immense room for improvement on quality (using the evaluation standards) and benefiting from the rigorous thought process of moving from information – whether qualitative or quantitative- to an explanation of the phenomenon perceived.

Evaluative mindsets. And as if all this is not enough of an agenda, participants discussed the need for evaluators to work with courage, humour, connection, winning with trust and integrity. Using every opportunity when doing an evaluation to surface these issues and bring them to the fore. We need to scrutinize any agglomerations of new ideas, methods and techniques for faddishness – irrespective of which field it comes from. Including the faddishness that these discussions may represent! As Jess Dart commented: *"There is a lot of consensus in this room. We need to stay critical!"*

3.3.3 Action Area 3. Networking and linking

Sharing what works and does not. The evaluation world has spawned many e-lists, websites, lists of methods, manuals. We twitter, facebook, email and e-discuss. However, we need to create more opportunities to share work and subject it to comment. BetterEvaluation (see Box 12 above) will offer opportunities to open up cases for comment, generate peer-reviewed examples and discuss the trade-offs and evaluation-related choices at play.

Linking more widely. There is need to identify who we need to engage with, such as the more open-minded and the more sceptical. We need to refine the approaches to make them more understandable and accessible for people. As Sandip, sharing his case from Cambodia said: “*Stakeholders’ perspectives vary; so do the expectations from evaluation. These expectations compete with each other for attention. However, often, mediation of these expectations is poor due to methodological and political reasons. Improvement in practice requires that practitioners, commissioners and donors work together on this issue. What will provide impetus to change in practice is the way evaluations are commissioned. The coming together of commissioners, donors, and practitioners, in this conference, in my view, was an encouraging first step.*”

3.3.4 Action Area 4. (Action) Research

Theory and its influence on ways of working remain important areas where work is needed. Patel’s call to ‘learn better how to assess power struggles’ requires clarity on power and (action) research around this. But it also requires linking this into the funding organizations’ and their requirements. Patel gave one example of the disconnect:

“In the infrastructure project, SPARC was the only organisation who could say with credibility that people would move. As for the procurement specialists, they were taking something that makes sense in engineering terms and putting them on social things. They wanted the baseline to be done by one organisation, the design by another and the implementation by a third! But poor people need to have trust with the people engaging them. ‘I’m not going to take down my house, unless I’m sure of the relocation!’.”

One of the conference participants, Rosalind Eyben, discussed an emerging idea in May that has since taken root and been welcomed by dozens of individuals in international organisations – NGOs and bilateral. It was initially discussed as ‘The Big Pushback’ against a narrowing of what is valued and how value is measured. It is now known as ‘The Big Push Forward’, and is seeking to find constructive ways to advance conceptually and methodologically for assessing the impact of development aid in support of a fairer world, beyond the narrow bureaucratic protocols that assume guaranteed predictable outcomes.

During a first Big Pushback workshop held in the UK in September 2010, interest was expressed to undertake action research within organisations by agency staff on a range of topics (see Box 13). Since then other ideas have emerged, including, for example, tackling the role of middle management in imposing ‘over-bearing frameworks’ and engaging in the politics of methodology. Another bilateral aid agency has just commissioned more work on non-experimental impact evaluation methods.

Box 13. Areas of interest in action research ‘The Big Push Forward’

- *Challenging dominant discourses* with alternative discourses that stress the significance of history and context and emphasize accountability to those for whom international aid exists.
- *Communicating* in more innovative ways the complex nature of development by collaborating with and drawing on the expertise of development communication agencies in facilitating debates and expanding spaces for voices from the South, while building knowledge of how the public in the North understands development.
- Developing *different methods of reporting*, so that the requirement for aggregated numbers at Northern policy level does not influence the character of programming in complex development contexts.
- *Collaborating* with people inside donor agencies who are equally dissatisfied with the prevailing ‘audit culture’ and want to make space for supporting social transformation.
- *Re-claiming ‘value for money’* by communicating with donors and the public that some aspects of development work are valuable while irreducible to numbers; improve development organisations’ own internal practices in terms of value for money, e.g. in procurement; and work together to develop more self-critical standards.
- *Enhancing organisational learning* and reflective practice, using professional training to nurture out-of-the-box thinking and approaches
- *Scrutinizing the role of big business* in development aid and its impact on discourse and quality.

4 The Conference Process in a Nutshell

On request of the participants a summary is provided of the Conference process, to understand the design, the flow of the programme, facilitation techniques used and review what worked and what did not work, and how the programme elements related and provided focus. The programme is provided in Appendix 1.

To design the process, a mix of methods was identified that enabled participants to think through the five core questions of the conference:

- What values and quality standards are needed within evaluative practice in order to understand and influence such change processes?
- What are the strengths and weaknesses of innovative examples of evaluative practice that do justice to the complexity of change? (see Appendices 3 and 4)
- Where do trade-offs occur in quality standards and core values? What are the consequences of these trade-offs for evaluation findings and for development?
- What constitutes ‘robust evaluation’ for societal transformation, in terms of core values and quality standards?
- What needs to shift in development in order to make possible high quality evaluative practices that influence societal transformation?

These methods included keynote participant presentations (with careful steering of content), case clinics, ‘methods and more markets’, ritual dissent, individual and subgroup discussions, plenary discussions and reflections. This variety of methods used in an open and friendly atmosphere (with an entertaining chair!), whilst keeping a clear focus, helped to critically analyse and debate the core questions of the conference.

Day 1

The day started with organisers presenting background and reasoning for the Conference. The key concepts and a conceptual diagram were presented. Also the intended follow-up and positioning of the event in a bigger discussion was presented. The opening provided a road map of the event and acknowledged the different contributions to the event. It also allowed people to understand the background of those that were in the room and what is needed for an event like this. The opening was completed with an exercise called ‘the rainbow lexicon’: exchange between neighbours feelings and issues from the evaluative practice that tries to deal with complexity. This allowed people to connect and enter into core conference themes from their own experience.

The keynote address of Dr. Patricia Rogers brought into the debate the implications of ‘complicated’ and ‘complex’ for evaluative practice, locating this in current debates on what is considered rigorous. In the plenary session afterwards it became clear that a large group finds it difficult to relate in practical terms to complexity. The plenary session built a bridge from the presentation to the case clinics by stimulating critical questioning on ‘rigour’ and ‘complexity’.

The nine case clinics were included to assess the rigour of methodological options (dilemmas and good practice, standards and values, evaluation approaches for embracing complexity). Participants appreciated being exposed to a wide variety of cases. The time frame was too short to deepen the

discussion; most cases could not follow the case clinic programme guideline until the last part (a so-called 'fish bowl' exercise).

The keynote address from Sheela Patel was scheduled for the morning to illustrate a concrete case of evaluative practice in relation to societal transformation and provide a common reference. Due to travel difficulties she could arrive only in the afternoon, but this did not hamper the flow of the event. Many participants valued her focus on the people behind the abstractions being discussed and why we are discussing rigour in evaluative practice and the complexity of social change. In fact, it illustrated that emergent unplanned events are part of our live and we better strengthen our flexibility and openness to deal with this reality.

The first day ended plenary with the chair inviting to share insights from the cases, link the cases with the presentations and link back to the core question: what is rigorous evaluative practice that understands the complexity of social change.

Day 2

The second day started with a summary from the first day and explanation of the linkage to the programme of the second day. This worked well and it clarified the purpose of the second day: focusing, analysing, and looking forward.

The third inspiring keynote address by Maarten Brouwer highlighted critical trends in development cooperation that are shaping the space that exist for different approaches to evaluative practice and that legitimizes some over others, in certain circles. The presentation allowed the audience to understand what these trends imply for rigorous evaluative practice that embraces complexity.

Afterwards the morning session continued with the "Methods & More Market". The objective of this part of the programme was to share new practices, experiences and methodologies that show or seek rigorous evaluative practice that embraces complexity in understanding and influencing societal change. In addition the objective was to help participants on their reflection on implications for personal engagement in evaluative practice. Focussing questions were: Does this method/approach make possible rigorous evaluative practice that embraces complexity? Why/why not? What action (for personal evaluative practice) does the session trigger in me?



After lunch the Conference returned to the bigger questions to articulate solidly reflected responses to the methodological debate and the political debate:

- (i) What constitutes rigorous evaluative practice for embracing complexity?
- (ii) What are the most strategic opportunities to ensure space and capacity for rigorous evaluative practice that embraces complexity and influences societal transformation?

Firstly in plenary an overview was made about what the challenges are that people currently face in ensuring rigorous evaluative practice that also deals with complexity; to tease out methodological and

political dimensions. Afterwards participants addressed the two main questions in two sets of subgroups and applied a method known as 'ritual dissent' to deepen the argumentation and understanding. A strong engagement and participation was clearly present and participants contributed greatly in providing substance and backgrounds for both questions.

After the break in the afternoon the dots were linked through a panel responding to the results from the groups and interacting with the participants. This was followed by a guided meditation, facilitated by Roger Ricafort, to allow people to come back to themselves and their feelings with regard to the topics raised and discussed. The Conference was evaluated by assessing people's sense of value of the event and focussing participants around their own practice and next steps. The voting card method allowed both a transparent quantitative feedback as well as qualitative illustrations why people decided for their appreciation. Questions from the organisers and from participants were included in this evaluation (see also Appendix 6). During the closing, gratitude was expressed by CDI and Context to all present in the room: the organisers, presenters, participants and funding agents.

Appendix 1 – Programme

DAY 1 – Concepts, challenges and cases

Time	Topic
08.00	Registration
09.00	Opening by conference organisers
09.15	Welcome and road map of the event - Isaac Bekalo, event chair
09.45	Rainbow Lexicon
10.30	Break
11.00	<i>"Exploring the implications of 'complicated' and 'complexity' for evaluative practice"</i> Dr. Patricia Rogers, professor of public sector evaluation at RMIT, Melbourne and award-winning evaluation specialist
11.35	Plenary reflection and discussion
12.00	Lunch
13.00	Case Clinics: Round 1. Participants choose two out of nine cases. In the clinics, participants will analyse cases in relation to values and evaluation quality standards present in the case, how an understanding of complexity is enabled, how rigour has been sought, which dilemmas faced, where trade-offs have been made.
14.30	Break
15.00	Case Clinics: Round 2
16.40	'Breaking News!' plenary
17.00	<i>"Making evaluation work for SPARC – being rigorous and ensuring it serves societal change"</i> Sheela Patel, founding member of world-renowned NGO 'SPARC' in Mumbai, India and award winning social activist.
17.45	End of Day 1 with Drinks and Bites

DAY 2 – Trends, options and opportunities

Timing	Topic
08.30	Breakfast News! Plenary start
09.15	<i>"Development Cooperation at Crossroads: On what basis to choose new directions?"</i> Maarten Brouwer, special Ambassador of International Cooperation, DGIS (Directorate General for International Cooperation) in The Hague, The Netherlands
10.15	Break
10.45	Methods and More Market
12.30	Lunch
13.30	Returning to the big questions – group debate through ritual dissent The methodological debate. What constitutes rigorous evaluative practice for embracing complexity? The political debate. What needs to change and how in order to make possible high quality, value-driven evaluative practices that embraces complexity?
15.30	Break
16.00	Connecting the dots. Panel and plenary debate on lifting the game for evaluative practice in development
16.45	Assessing the past days
17.20	Closing

Appendix 2 – List of Participants

N°	Name	Organisation
1	Abbenante, Lioba	Deutscher Entwicklungsdienst
2	Adams, Laurie	Actionaid International
3	Adanutty Esenam, Mercy	Women And Development Project (Wadep)
4	Alpenidze , Irma	MDF Training and Consultancy
5	Aman, Aslam	H & H Consulting (Pvt.) Ltd.
6	Arkesteijn, Marlen	Capturing Development
7	Bailey, Simon	Aflatoun
8	Baser , Heather	ECDPM and CIDA
9	Batjes, Karen	Freelance Consultant
10	Bekalo, Isaac	IIRR Philippines
11	Belle van, Sara Bert	Prince Leopold Institute Of Tropical Medicine
12	Boxelaar , Lucia	World Vision Australia
13	Brouwer , Maarten	Ministry of Foreign Affairs
14	Brouwers, Ria	Institute of Social Studies
15	Brouwers, Jan	Context, international cooperation
16	Budge, Timothy	Plan
17	Causemann, Bernward	NGO-IDEAs
18	Ceban, Ludmila	Eadi
19	Chambers, Robert	Institute of Development Studies
20	Chambille, Karel	Hivos
21	Claessens, Lotte	War Child Holland
22	Collingwood, Catherine	Social Development Consultant
23	Conner, Ross	University Of California Irvine
24	Cremers, Petra	Wageningen University
25	Cuesta Fernández, Iván	Technival University Of Valencia
26	Cummings, Sarah	IKM Emergent / Context, international cooperation
27	Da Silva Wells, Carmen	IRC International Water and Sanitation Centre
28	Dart, Jess	Clear Horizon Consulting Pty Ltd
29	Davies, Richard	Independent M&E Consultant
30	Davis, Kristin	Global Forum for Rural Advisory Services (GFRAS)
31	de Lange, Piet	Ministerie van Buitenlandse Zaken
32	Demilew, Abebual Zerihun	Brac East Africa
33	Deprez, Stefan	VECO Indonesia
34	Desalos, Caroline	Independent
35	Dietz, Ton	University of Amsterdam
36	Dinges, Sabine	GTZ
37	Earl, Sarah	IDRC Canada
38	Elo, Elsebeth	Plan International (HQ)
39	Es, Yvonne	Oxfam Novib
40	Estifanos, Abiy Seifu	Westat Ethiopia Consulting Plc
41	Eyben, Rosalind	Institute of Development Studies
42	Faber, Koen	PSO Kenniscentrum

43	Fountain, Gerinke	NIZA
44	Fraser, Elvis	Bill & Melinda Gates Foundation
45	Gaasbeek, Timmo	ZOA Refugee Care
46	Gajo, Michael	GTZ
47	Gnagi, Adrian	Swiss Agency For Development And Cooperation
48	Godden, Kate	University Of Westminster
49	Grootoank, Inge	RNTC Radio Nederland Training Centre
50	Guijt, Irene	Learning by Design
51	Haagsma, Ben	IC Consult
52	Hamilton, Kate	Independent Consultant
53	Hartog, Kim	SOS Kinderdorpen
54	Hearn, Simon	Overseas Development Institute
55	Hédia, Hadjaj-Castro	Cota Asbl
56	Heider, Caroline	United Nations World Food Programme
57	Hemelrijckvan, Adinda	Oxfam America
58	Hilhorst, Thea	Royal Tropical Institute (Amsterdam)
59	Ho, Wenny	Action Aid International
60	Holland, Jeremy	Ids Sussex
61	Holzen von, Nadia	Eper-Heks
62	Hummelbrunner, Richard	Oear Regionalberatung
63	Huyse, Huib	HIVA/KU Leuven
64	Jans, Valerie	SOS Kinderdorpen
65	Jones, Harry	Overseas Development Institute
66	Kakooza, James	Award, Gender And Diversity, Icraft
67	Kell, Catherine	Twaweza
68	Kima, Laetitia	Alliance for a Green Revolution in Africa
69	Koekebakker, Welmoed	Consultant
70	Korppen, Daniela	Berghof Peace Support
71	Kumar, Shubha	University of California Los Angeles (UCLA)
72	Kusters, Cecile	Wageningen UR Centre for Development Innovation
73	Kuzmanoska, Irena	Sagittarius Ltd
74	Levin-Rozalis, Miri	Ben-Gurion University
75	López-Torrejón, Estela	Universidad Politecnica De Valencia
76	Luijter, Arjan	Consultant
77	Mack, Dorothee	Misereor
78	Magunda, Andrew	The Private Education Development Network
79	Mahandra, Shanti	Institute of Development Studies
80	Milabyo Kyamusugulwa, Patrick	Wageningen UR - Disaster Studies
81	Moore, Marah	I2i-Institute Inc.
82	Mulder, Arjen	OXFAM NOVIB
83	Mussambala Figueiredo, Timoteo	Plan International
84	Namukoko, Mary	Campaign for Female Education
85	Nassamula, Harriet	Pact Lesotho
86	Nichols, Paul	AUSAID
87	Noordergraaf, Bert	ICCO
88	Norgah, Samuel	Plan International

89	Olila Odhiambo, Tom	Strategic Connections
90	Oosterkamp van den, Mirte	Context, international cooperation
91	Ortiz Ochoa, Natalia	As Raizes
92	Oswald, Katy	Institute of Development Studies
93	Pal, Anirban	Erasmus University Rotterdam
94	Patel, Sheela	Sparc India
95	Pattanayak, Sandip	Catalyst Management Services
96	Pelgrom, Hans	Ministerie van Buitenlandse Zaken
97	Penninkhoff, Petra	KIT Royal Tropical Institute
98	Peris Blanes, Jordi	Universitat Politècnica De Valencia
99	Pieper, Ivet	Context, international cooperation
100	Plavcak, Barbara	Wotro Science For Global Development
101	Polvi, Johanna	Transparency International
102	Potters, Jorieke	Wageningen UR
103	Pradel Caceres, Willy	International Potato Center (CIP)
104	Prins, Ester	Context, international cooperation
105	Rauch, Janine	African Security Sector Network (Assn)
106	Ricafort Escobar, Rogelio	Oxfam Hong Kong
107	Rijneveld, Wouter	Stichting Woord En Daad
108	Roefs, Marlène	SNV Netherlands Development Organisation
109	Rogers, Patricia	RMIT Australia
110	Roman Millan, Inmaculada	Universidad Politecnica De Valencia
111	Romer, Friederike	Kncv Tuberculosis Foundation
112	Ruijschoot, Lieke	Context, international cooperation
113	Rutten, Rens	CORDAID
114	Santos, Francisco	Fundacion Etea Para El Desarrollo Y La Cooperacion
115	Scheers, Goele	European Centre for Conflict Prevention
116	Schmid, Marcus	Cbm
117	Schot, Paul	Dorcas Aid International
118	Servaas, Maurits	ICCO
119	Sharma, Madan Lal	CECOEDECON (India)
120	Soal, Sue	CDRA (South Africa)
121	Squires, Heather	Beacons For Public Engagement
122	Steenkiste van, Christ	Vredeseilanden
123	Sterenborg, Marieke	Context, international cooperation
124	Suleiman, Adam	States Accountability & Voice Initiative (Savi)
125	Taanman, Mattijs	Erasmus University Rotterdam
126	Taffa, Negussie	Botusa
127	Tanis, Marijse	Stichting Woord En Daad
128	Temmink, Cristien	PSO Kenniscentrum
129	Tijmensen, Yvonne	PSO Kenniscentrum
130	Til, Jan	Plan Nederland
131	Ton, Giel	LEI Wageningen UR
132	Chhoden Tshering, Phuntshok	Ministry of Agriculture Bhutan
133	Laan van der, Anita	SNV Netherlands Development Organisation

134	Es van, Marjan	Hivos
135	Lith van, Brechtje	War Child Holland
136	Velden van der, Fons	Context, international cooperation
137	Verdonk , Inge	Plan Nederland
138	Vincent, Robin	Panos London
139	Vugt van, Simone	Wageningen UR Centre for Development Innovation
140	Waid Lee, Jillian	Helen Keller International
141	Wally Mohamed, Nermine	The Social Contract Centre
142	Whitmore, Elizabeth	Carleton University
143	Wielinga, Eelke	Link Consult
144	Wigboldus, Seerp	Wageningen UR Centre for Development Innovation
145	Wilson-Grau, Ricardo	Ricardo Wilson-Grau Consulting
146	Wind, Tricia	International Development Research Centre
147	Woodhill, Jim	Wageningen UR Centre for Development Innovation
148	Wortel, Erica	Wortel Project & Interim Management
149	Wuite, Rosemarie	NIZA
150	Zaal, Fred	KIT Royal Tropical Institute
151	Zaveri, Sonal	Community Of Evaluators
152	Zeynalova, Bayaz	Aflatoun
153	Zhao, Kun	China Health Economics Institute

Appendix 3 – Summaries of Case Clinics

- Case 1: Using Action Research and Appreciative Inquiry to explore effectiveness In Social Justice and Environmental Advocacy Work, Canada
- Case 2: Planning, Learning and Accountability for Sustainable Agricultural Chain Development: The case of VECO Indonesia
- Case 3: Evaluation of Dutch Support to Capacity Development: Evidence-Based Case Studies, Case Presentation on SOCSIS, Somalia
- Case 4: Social Return on Investment: A case study on the Grain Bank to offset the risk of severe drought conditions in the Malpura Block of Rajasthan – India
- Case 5: Aflatoun Evaluative practice of a Children’s Education Network (Global)
- Case 6: Formative evaluation and outcome monitoring of democracy and governance, Cambodia
- Case 7: Evaluation of the Stronger, Smarter Realities program, Australia and Uganda
- Case 8: Application of Multiple Rigorous Methods in Impact Evaluation: Experience from BRAC ELA Programme in Uganda and Tanzania
- Case 9: Participatory Assessment of Development, Burkina Faso and Ghana

Case 1. Using Action Research and Appreciative Inquiry to explore effectiveness in Social Justice and Environmental Advocacy Work
Bessa Whitmore, Maureen G. Wilson and Avery Calhoun

How do you know that you are effective or successful in the social justice advocacy work that you do? The answer to this question is best expressed by our project participants:

"Canada did cancel the (Honduras, after Hurricane Mitch) debts. ... Yeah, it was a major success."

"People investing their energy is always a measure or a sign of success."

"Fifteen years ago, if you were to ask someone, 'What is a sweat shop?' [the answer would have been] 'I don't know.' Whereas now, if you ask people 'What is a sweat shop?' there is a fairly good idea about what a sweat shop is."

"I think there's a lot of successes that take place, like you work well together, you meet timelines, you see a turnaround, people are happy at the end of the day."

"I think the spin-offs that took place from this (event) are more important things that I would say constitute success. First of all, in each one of the settings, the number of networks that were built amongst the academic institutions, the students themselves, and the community groups that came to these lectures, were absolutely phenomenal."

"[Success is] ... taking the time to invest in the people ... when they're pulling in different segments of the community – you know pulling in the Aboriginal members, as well as the non-Aboriginal members – and helping to facilitate relationships amongst these members."

1. About the project

This case focuses on understanding what makes a difference, about "what works" from the perspectives of social justice and environmental activists themselves. Through workshops, interviews and a national symposium, we asked participants to tell us what success or effectiveness means to them, and what they think are some of the factors or conditions that contribute to their successes.

Our case is not about a formal evaluation but of evaluative practice in a broader sense, through a research project with two purposes. One purpose was to try to support, in a very practical way, the work of progressive activists by creating an additional space in which to step back and reflect on (in effect, evaluate) what they are doing and the impacts of their efforts – whether and how they were making a difference. In the often frantic world of activists, such reflection becomes a luxury. A second purpose was to add to our collective understanding about what success means in terms of activists' efforts to achieve broad social justice and environmental goals.

We worked with nine very diverse groups and organizations across Canada over a 4 year period (2005-2009):

1. Raging Grannies: A grassroots group of older women with no staff, no budget and no organizational structure
2. Pembina Foundation: A national environmental research/advocacy organization with a large professional staff
3. Youth Project: A gay/lesbian/bisexual/transgender youth group
4. The national chapter of an international development advocacy organization
5. Social Justice Committee: A Quebec-based social justice advocacy group focusing on international issues
6. Disability Action Hall: A high profile self-advocacy group of disabled activists
7. Storytellers' Foundation: An aboriginal group promoting education and local development in rural British Columbia

8. Anonymous: A national chapter of an international organization advocating for the rights of children
9. Alberta College of Social Workers: A provincial organization that regulates social workers and addresses social justice issues

2. Methodology

Assumptions about knowledge development: The project's methods were grounded in social constructionism, which emphasizes that knowledge derives from social interaction and reality is dynamic and indeterminate. Equally important, not only is knowledge contextually determined, its creation, in turn, influences. That is, from the constructionist perspective, how we interact is generative of reality. From the variety of methods that are in keeping with constructionism, we chose two– action research and appreciative inquiry – that are consistent with our philosophical position, social justice values and specific research questions.

Action research: Action research (AR) paved the way for our partners to be genuine collaborators in the research and in the evaluation of their own practice. The core aspects of AR – collaboration and dialogue among partners, and the focus on practical applications – offered a solid foundation for our work.

Appreciative Inquiry: Appreciative inquiry (AI) helped focus our participants' reflections on what was already working within their organizations and on preferred future outcomes. AI essentially shifts the focus from problem-oriented thinking to a process that examines and builds on positive experiences and successes in planning for, designing and implementing future actions. This process informs both our understanding of a situation (theory) and what we do about it (practice). Using tools of action research and appreciative inquiry, our process involved a series of workshops with each organization, a set of in-depth individual interviews with a variety of stakeholders (92 interviews) and a symposium that brought representatives from our partners together to share their experiences and refine our collective understanding what success means in social justice and environmental work.

3. Values and quality

A number of basic values and premises guided this research. One premise was the need for civil society interventions in a neo-liberal world. With the diminished ability (or willingness) of governments to protect their citizens from its impacts, civil society organizations worldwide have been moving into the breach, confronting the threats of corporate globalization to democracy, economic justice, the environment and protection of the commons. An important value stemmed from the imperative for social workers to address social justice issues and to evaluate their practice, as part of the professional code of ethics. Sharing power constituted an additional key principle or value. As allies with the participating organizations, we were very aware of the need to be flexible and adapt the methods to the needs of our activist partners. Above all, participation in this project needed to be useful for them, while also responding to our interest in knowledge creation. We used a number of strategies, drawn from various sources, to enhance the quality of the process: Guba and Lincoln's (1989) strategies to increase the trustworthiness of results, including prolonged engagement and persistent observation, member checking, triangulation and peer debriefing; their authenticity criteria (fairness, inclusiveness and action)(Lincoln, and Guba 2000); and Mertens' (2005) emphasis on standpoint, attention to voice, critical reflexivity, and reciprocity.

Our partners welcomed the opportunity to reflect in some depth on their work; indeed that is why they wanted to participate in the project. The process pushed them to articulate what they did, their criteria for success, and the evidence that their efforts were effective. The dialogic process integrated the sharing, feedback and critical analysis so essential for enhancing quality. Some found this helpful in strengthening funding applications, others for planning and evaluation purposes.

4. Complexity

Our study is grounded in the day-to-day realities of a highly varied set of organizations/ groups. Ramalingam et al (2008: 63-64) call this a pragmatic perspective, that is, we focused on the relevance of complexity to assess practical benefits for activists doing social justice and environmental work. Our deliberately flexible process was built in from the beginning, as we were well aware that each organization's needs and interests would be different.

All of the key concepts of complexity theory were reflected in individual stories and interviews and in our collective discussions. People could tell us, in vivid and detailed narratives or stories, about successful campaigns or events, but had difficulty putting these into simple cause-effect terms. They understood the interconnectedness of their work, illustrated in the ubiquitous presence of networks, coalitions and collaborations among organizations. They were well aware that unpredictability was part of the process, and that learning and adapting were ongoing. In spite of that, they felt strongly that their work did, in some way, contribute to the changes they wanted to make. It seemed to us that activists were often most excited, and felt most creative when they were operating "on the edge of chaos."

The *meaning of success* cannot be reduced to a singular, easy to define, sound-bite; its value is precisely in its complexity, its dynamism, its many meanings. Our challenge or, rather, opportunity, is to embrace this ambiguity, and harness the energy, commitment and enthusiasm of activists and their organizations (the attractors) that support progressive change.

The surest safeguard for rigour, according to Chambers (1997), is sharing with peers and inviting their analysis and critical review. These processes constituted the essence of our method. The negotiation process (initially and throughout) allowed for what Guijt (2008) calls 'messy partnerships' to be adjusted as circumstances and needs shifted. We built in flexibility, so that each organization could use the process in a way that best suited their interests. The symposium brought people together not only to share and network, but more importantly to critically examine and refine our results. Finally, eight organizations have contributed chapters about their experience to a book about the project, to be published next year. All these processes served to strengthen the robustness of the project and our results.

5. References

- Chambers, Robert (1997). *Whose reality counts? Putting the first last*. London: Intermediate Technology Publications.
- Guba, Egon G. and Lincoln, Yvonna S. (1989). *Fourth generation evaluation*. London: Sage.
- Guijt, Irene (2008). *Seeking surprise: Rethinking monitoring for collective learning in rural resource management*. Published PhD Thesis. University of Wageningen, Wageningen, The Netherlands.
- Lincoln, Yvonna S. and Guba, Egon G. (2000). Paradigmatic controversies, contradictions and emerging confluences. In Norman K. Denzin and Yvonna S. Lincoln (eds.) *Handbook of Qualitative Research* (Second Edition, pp. 163-188). Thousand Oaks, CA: Sage.
- Mertens, Donna M. (2005). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. (Second Edition). Thousand Oaks, CA: Sage.
- Ramalingam, Ben and Jones, Harry with Reba Toussaint, Reba and John Young (2008). *Exploring the science of complexity: Ideas and implications for development and humanitarian efforts*. Working Paper 285, Second Edition). London: Overseas Development Institute (ODI).

**Case 2. Planning, Learning and Accountability for Sustainable Agricultural Chain
Development: The case of VECO Indonesia**
Steff Deprez & Christ Van Steenkiste

Development and experiences of an Outcome Mapping based planning, monitoring and learning system for a sustainable agriculture chain development programme in Indonesia.

1. Introduction and background

The Belgian NGO Vredeseilanden aims to contribute to viable livelihoods for organised family farmers through improved income from sustainable agriculture with a particular focus on value chain development. Vredeseilanden operates in 7 regions including 13 countries in Central and South America, Africa and Asia. Vredeseilanden programme 2008-2013 emphasises the strategic importance of organisational and institutional learning for the management of its chain development programmes as well as in building its expertise as a supporter of family farmers to actively participate in markets. Subsequently, Vredeseilanden chose to apply a more learning-oriented planning and M&E system and decided to use core elements of *Outcome Mapping*⁸ for the programme design as well as the monitoring and learning process.

This paper describes the development and the experiences of the planning, monitoring and learning system of VECO Indonesia⁹, but also draws from the experiences in the other VECOs and the global Vredeseilanden programme. VECO Indonesia's SADC programme supports farmers producing healthy rice, cacao, coffee, cashew and groundnuts in Eastern Indonesia, Bali and Central Java. To achieve the objectives of its programme, VECO works through partnerships with local actors such as farmer/producer organisations and local NGOs to improve the position of farmers in the value chain. The VECO head office is based in Bali and manages five field antennas and operates with twenty-four staff. The programme is mainly funded by the Belgian government and co-funded by a variety of other donor organisations.

2. Intentional Design: Chain Intervention Framework à la Outcome mapping

The *intentional design* of VECO's programme describes the changes the programme aims to bring about and plans the strategies it will use. Vredeseilanden refers to the *Chain Intervention Framework (CIF)*. Every value chain supported by VECO requires a contextualised CIF. A visual presentation of the intentional design of VECO's SADC programme is presented in Figure 2 below.

⁸ Outcome Mapping (OM) is a planning, monitoring & evaluation approach developed by the *International Development Research Centre (IDRC)* in Canada (Earl, Carden and Smutylo 2001) and is an alternative and/or complementary model to the LFA. www.outcomemapping.ca

⁹ VECO Indonesia is the Vredeseilanden Country Office in Indonesia

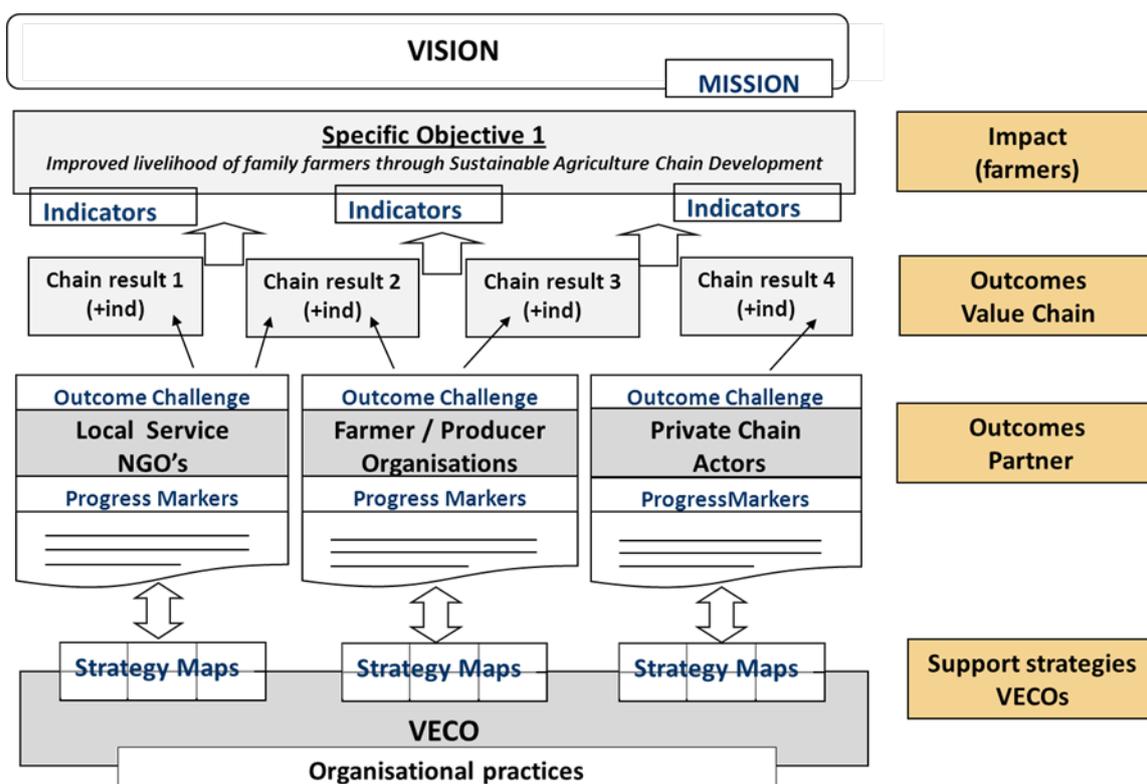


Figure 2: Chain Intervention Framework: Intentional Design of VECO's SACD programme 2008-2013

The *vision* describes the large-scale development changes that VECO hopes to encourage while the *mission* spells out how VECO will contribute to that vision and clarifies which part of the vision VECO's programme is going to focus on.

A dominant principle underlying OM is the idea that development is essentially about people relating to each other and their environment, hence its actor-centred focus. It introduces the concept of *boundary partner*, i.e. individuals, groups, or organisations with whom the programme interacts directly and with whom it anticipates opportunities for influence. The actions of the boundary partners - and the influence they have on the *beneficiaries* (e.g. family farmers) - contributes to the ultimate goals of the programme (Earl, Carden and Smutylo 2001).

VECO identified three key boundary partners for its SACD programme: farmer/producer organisations, local service NGOs and private chain actors. OM focuses on one specific type of results, i.e. *outcomes as changes in the behaviour, relationships, activities, and actions of people, groups, and organisations with whom VECO works directly* (Earl, Carden and Smutylo 2001) - the farmer/producer organisations, local service NGOs and private chain actors.

An *outcome challenge* describes the desired changes in the behaviour, relationships, activities, actions (professional practices) for each (type of) of boundary partner. It is the ideal behavioural change of each type of boundary partner for it to contribute to the ultimate goals (vision) of the programme.

The *progress markers* are a set of statements describing a gradual progression of changed behaviour in the boundary partner leading to the ideal outcome challenge. Progress markers articulate the complexity of the change process and represent the information that can be gathered in order to monitor partner achievements.

OM does not attribute outcomes to a single intervention or series of interventions, but looks at the logical link between programme interventions and behavioural change of its boundary partners. It is not based on a causal framework and acknowledges that impact changes are not caused by a single intervention or series of interventions by the programme; rather, it recognises that multiple, non-linear events lead to

change. Whereas programme planning is done in relation to the broader development context (vision), OM promotes the idea that assessment (M&E) should be focused on those changes *within the sphere of (direct) influence* of the programme, i.e. outcomes of the boundary partners (Earl, Carden and Smutylo 2001).

Although VECO underwrites this principle, both VECO and its donors also want to be informed about the changes in the *'sphere of interest'* (impact) in order to know the intended and unintended changes, understand and analyse the programme's contribution to these changes and to improve its interventions. In phase I (2008-2010), **specific objectives** (at farmer level) - and a respective set of **indicators** - describe projectable changes within the scope of the programme and are assumed to be achieved by the development efforts of VECO's boundary partners. During phase II (2011-2013), an additional 'layer' will be added, i.e. **results**, to be able to monitor the situational changes at value chain level.

Strategy maps are a mix of different types of strategies used by VECO to contribute to and support the achievement of the desired changes at the level of its boundary partners.

Organisational practices explain how VECO is going to operate and organise itself to fulfil its mission. It is based on the idea that supporting change in boundary partners requires that the programme team itself is able to change and adapt as well, i.e. not only by being efficient and effective (operational capacities) but also by being relevant (adaptive capacities).

Box 14. Twisting the models: Integrating Outcome Mapping & the Logframe

Although VECO's uses an OM-based programme framework for its ongoing planning and monitoring processes, it needs to report to its main donor by a logframe-based format. A pragmatic integration of OM into a logframe is established, although with some differences in phase I (2008-2010) and phase II (2011-2013) due to the changes in the Intentional Design:

1. **Objectives:** The vision of VECO was 'translated' in three more tangible and measurable objectives (SACD, Advocacy and Consumer Awareness) which seemed beneficial for both the Intentional Design (OM) and the logframe, hence it appears in both models.
2. **Intermediate results:** During phase I, the intermediate results in the logframe describe the changes of the direct partners and connected with the outcome challenges of the respective partners. The respective indicators are added to the set of progress markers as to guarantee inclusion in the data collection process and M&E process for future reporting. During Phase II, the intermediate results (+indicators) in the logframe will describe the changes in the value chain which correspond with the results in Intentional Design.
3. **Type of activities:** In Phase I and II, VECO's type of activities (in the logframe) resonate with the strategy maps of the Intentional Design. In Phase II, the main activities of the direct partners will be added in the 'type of activities of partners' and will be derived through monitoring of the progress markers. This implies that the outcomes of partners are as such not reflected in the logframe, but will only be used for ongoing planning and monitoring/learning for the specific value chain intervention.
4. **Usefulness:** The logframe is structured in such a way that it presents aggregated information across the different value chains (in one country and across different countries). In this way, the compilation of the logframe is also a useful exercise for VECO.

It might appear that VECO Indonesia is seemingly using two programme models for the same programme which might lead to a 'schizophrenic' planning and M&E practice. However, VECO Indonesia takes a pragmatic approach to this and views the *Chain Intervention Framework* as the core framework for ongoing programme steering and learning while the logframe format is seen as an (important) way of presenting and reporting progress and results.

3. Facilitating value chain development: embracing complexity

Value chain development is a complex process in which relationships, dynamics and mechanisms cannot be fully understood. Different actors (chain actors, chain influencers and chain supporters) are involved and have different roles, motivations and levels of power in the chain. And, it is affected by external factors such as government policies, institutional environment, environmental & cultural aspects, market dynamics, trends and consumer behaviours. Although the design and planning of value chain development programmes often assumes a rather linear relationship between interventions and desired results, it is highly contextual and effects from interventions can often only be observed and understood as the programme engages in action.

Acknowledging and dealing with this complex character of chain development requires a planning, monitoring and programme management approach which allows VECO and its partner to understand the process they are engaged in and make it usable for improved action. The particular logic of the Intentional Design (see above) has the potential to be instrumental in 'un-packing' the process as it provides a framework for continuous 'feedback' on intended and unintended results and outcomes in relation to programme interventions and strategies. However, how ongoing feedback is organised, how people make sense of reality and to what extent the programme can anticipate to emerging insights depends entirely on the quality and the nature of the ongoing monitoring and learning process, and the organisational/institutional conditions in which the programme is operating.

In 2008, VECO developed a learning-oriented M&E system that aimed to facilitate such a process and provides a framework for systematic data collection, sense-making and documentation which supports VECOs *planning* & management process, facilitates organisational and institutional *learning* and fulfils VECO's *accountability* requirements. Hence, the name *Planning, Learning and Accountability system* (PLAs).

4. Development of the Planning, Learning and Accountability system (PLAs)

VECO Indonesia took an action learning/research approach for this development process carried out in a period of more than one year. Throughout the process different programme actors participated at relevant times. The inquiry, reflection and data generation process was organised through workshops, focus group discussion, a virtual community of practice, observations, semi-structured interviews and document analysis.

Based on the theory and practice of Outcome Mapping (Earl, Carden and Smutylo 2001), Participatory M&E (Estrella and Gaventa 1997; Guijt 1998; IFAD 2002) and Utilisation-Focused Evaluation (Patton, 1997; Horton et al 2003), a seven step model was developed and used to facilitate the inquiry and development process of the PLAs (table X below). Each step of the PLAs design process involved an in-depth inquiry, discussion and negotiation guided by a set of questions. Although a 'one-step-at-the-time' logic might be assumed, the process was seldom linear and unfolded through different, often simultaneous, inquiry and reflection processes.

Table 4: Seven design steps of the Planning, Learning and Accountability system (PLAs); Case II Indonesia

1. Purpose and scope	Identify, clarify and share the main purpose and scope of the PLAs
2. Organisational spaces and rhythms	Identify the key moment/events and their frequency for planning, learning & accountability
3. Information needs	Define and prioritise the monitoring and learning questions and specific information needs
4. Plan for data collection and synthesis	Plan how the data will be collected, stored and synthesized
5. Plan for sense-making	Plan for critical reflection, analysis and conceptualization
6. Plan for documentation and communication	Plan how monitoring results will be documented and communicated
7. Plan for creating organisational conditions	Plan how the necessary organisational conditions and capacities will be established in support of the PLAs

Sufficient time was taken to design the first three steps of the process as they are crucial for the further development of the PLAs. In fact, they could be seen as the foundation of the M&E system. These steps need to be discussed and negotiated thoroughly with the actors involved. By doing so, they automatically provide direction for the timeframes, the depth and type of data collection, the sense-making process and the documentation. By implication, any changes in the first three steps during the implementation of the programme will have a direct effect on the process of the next steps.

This first step aims to create a common understanding on the **main purposes** and *the intended uses* (Patton 1997) of the M&E system (planning, learning & accountability). As many programmes have difficulty defining clear boundaries and units of analysis (Horton et al 2003), it is important to define the *scope* of the M&E system. It includes a debate on the *intended users* of the M&E results and on who should be involved in which parts of the M&E process. This step also included a reflection and a discussion on the aspirations with regard to the underlying values and principles of the PLAs (see further).

The second step is the identification of VECO's key moments and events for planning, learning and accountability. If M&E is to foster and facilitate organisational and institutional learning, it needs to be built into the regular organisational processes so as to become integral to the thinking and doing of the organisation (Dlamini 2006) and embedded in those **organisational spaces and rhythms** which are central to sharing, debate, learning and decision-making (Guijt & Ortiz 2007). *Organisational spaces* are the formal and informal meetings and events that bring organisations and programmes to life. *Rhythms* are the patterns in time, the regular activities or processes that characterises the tempo of organisational functioning (Reeler, 2001). In practice, defining the spaces and rhythms is to sit down with the people involved and ask them to list when they interact to share information and make sense of what is happening for their planning, learning and accountability processes. These moments are then listed and categorised followed by a mapping out of the rhythm, i.e., how often and when these spaces occur and what type of sense-making – sharing, debate, learning, decision-making – occurs (Guijt & Ortiz, 2007). The key organisational spaces of VECO's SADC programmes are home weeks (six times per year), bi-annual chain partner meetings and bi-annual VECO steering team meetings.

The third step is to clearly define and prioritise the required **information needs**. Information connected to the programme design – for example through the result indicators, progress markers and type of activities - are mostly the starting point in defining the information needs. However, as step two made clear, information has to be used and made usable for action. Therefore, defining and prioritising information needs to be linked with the outcomes of step one – the *intended uses* - and step two – the *organisational spaces* - in order to decide upon the importance and relevance of specific information

needs. Step three is centred around the following questions: *Which information is required? For which event? For who? And in what form?*

When the specific information needs are defined, a **plan for data collection** (step four) can be developed. It includes a decision on the *type of data* (primary/secondary, micro/macro data, quantitative/qualitative, etc.) as well as the *data collection methods*. Furthermore, it implies a plan for *data storage* and on how the data will be *synthesised* and presented in a way that is understandable and useful for the users.

For the key organisational spaces, the programme will need to plan how the data is used and analysed in order to strengthen concerted action and facilitate decision-making (step five). As social interactions are crucial sources for **sense-making** and critical for organisational and institutional learning, it is important to ensure that the organisational spaces are well-planned and facilitated. One of VECO's key organisational spaces is the bi-annual *chain meeting* during which farmers, boundary partners, VECO staff and other actors working in a specific value chain - i.e. a multi-stakeholder meeting - gather to share information, update each other on the progress and results in the value chain, discuss the role and contributions of partners and VECO, and to agree on joined chain interventions.

OM suggests to actively engage the programme team and boundary partners in the M&E process and promotes self-assessment as a key process in the M&E practice (Earl, Carden and Smutylo 2001). OM's learning character lies in the fact that it calls for reflection and analysis of the connections between changes at the level of the boundary partner (progress markers) and the support strategies of VECO's programme team. OM includes a variety of reflective questions to guide the sense-making process: 'what are the intended and unintended changes?', 'what are the hindering and contributing factors in relation to the desired changes?', and 'how effective and efficient were the support strategies in relation to the changes?'. In addition, VECO systematically collects data on the results and objective indicators. In this way, the M&E system facilitates a continuous process for thinking holistically and strategically about how it intends to achieve results and unites M&E of the process, outcomes and impact of the programme.

VECO experienced that there is no blueprint approach in planning for sense-making events and that each event should be designed according to the situation and context at the time of implementation. In fact, it is an evolving process of 'learning how to learn' with programme actors. The facilitation of each key event is guided by the desired outputs and the following principles:

- The collected data and information are presented and shared with those who produced the data or to those to whom the data pertains. By doing so, these events become an important feedback mechanism in the PLAs system;
- Participants engage in a critical analysis and debate on the data. By doing so, it is hoped that their understanding of the changes, progress, challenges and the context of the programme will improve and that insights for improved action will be co-created;
- The events are data generation events in itself during which additional data and information emerge from the analysis and the use of probing questions;
- Participants draw conclusions and formulate recommendations for future actions. Although consensus can be reached on some common points of action, multi-actor meetings do not necessarily need to include decision-making; and,
- The events should be organised in a way that motivates (e.g. by highlighting the achievements and progress made) and inspires people. This includes appreciative methods and approaches which avoid mechanistic analytical processes and are fun.

VECO acknowledged that the OM approach or the PLAs alone will not lead to learning-oriented M&E practice. At the end, real people in the real world have to 'translate' the principles and systems into action. Therefore, VECO concluded its PLAs design process with a critical analysis on the existing **organisational conditions** and the development of a plan for the institutionalisation of the PLAs. To do so, it developed an approach that is derived from the theory and practice of organisational learning. It is mainly based on the concepts presented in Britton's (2005) model on organisational learning in NGOs and Senge et al's (1994) work on learning organisations, i.e. that an organisation can develop a practical strategy for learning if it creates the right *motives, means and opportunities* to do so (see table X). The three elements

generate a synergy that will not occur when attention is paid to only one of the elements alone. For each of the three elements, four crucial organisational conditions were identified which need to be installed and maintained to ensure a successful implementation of the PLAs.

Table 5: Necessary organisational conditions for successful implementation of the PLAs

CREATING MOTIVES	<ol style="list-style-type: none"> 1. Formulate guiding ideas for a learning-oriented M&E system 2. Ensure support from management for (organisational) learning 3. Create a culture of learning 4. Install the necessary incentives to participate in or manage the M&E process
CREATING MEANS	<ol style="list-style-type: none"> 1. Strengthen human capacity (M&E, learning, reflective practice, reporting,...) 2. Provide specialist support (M&E, Learning, Knowledge Management, facilitation, ...) 3. Develop and use appropriate concepts, methods & tools 4. Ensure adequate financial resources for M&E and learning
CREATING OPPORTUNITIES	<ol style="list-style-type: none"> 5. Integrate the PLAs into the management & operational processes 6. Ensure clear and transparent structures, responsibilities & plans for M&E and learning 7. Develop a responsive information management infrastructure 8. Build relationships of trust among staff and partners

5. Monitoring and learning principles and practice

The following values and principles have inspired the development of the PLAs and VECO aims to translate - 'live' and act upon it – those into its monitoring and learning practice. From the start, VECO aimed to install a monitoring practice that is realistic to organize, that is (cost)effective, pragmatic and above all, ... that makes sense.

VECO wants to move away from the notion that M&E is '*done to the programme*' and aims to engage the programme team and partners in the design and implementation of the monitoring and learning process. The core of the PLAs is to create the necessary spaces where actors involved can share, reflect and decide together wherever that it is appropriate and useful. The monitoring practice relies on (facilitated) self-assessment as an important sense-making approach, rather than external assessments or 'complicated' information gathering processes and reporting systems.

The PLA system needs to be useful for the programme actors who produce and use the information (Utilization-Focused Evaluation). Data and information that is collected, discussed and synthesized by people need to be useful - and has to make sense - for those people. And if applied to the different levels of the programme (farmers, VECO partners, VECO programme field staff, VECO programme coordination staff and Vredeseilanden Head Office) it has a direct effect to the type of information and the methods for data generation and analysis. *Which information is required? In what form? For who? At what time?* are critical questions during the design (and the actual M&E practice) that assist in prioritizing the 'nice-to-know' from 'must-know' information. In addition, linking the required information to the key moments for programme planning and steering ensured that the info was used and that the M&E process is embedded in the programme management cycle (which aims to avoid the perception that M&E is an 'burden' and 'add-on' to the programme).

With the enhanced focus on spaces for interaction as a core event in the PLAs, VECO was able to reduce the reporting requirements and one-way information gathering flows from partner organizations to VECO Indonesia. During the last two years, report formats have been constantly updated with the aim to increase usefulness and relevance for further sense-making processes or for accountability purposes (which resulted in some case in simplified report formats).

In line with OM, the PLA system aims to provide the programme with a continuous system for thinking strategically about how it intends to achieve its programme goals and therefore focuses on both the process and the results. A central aspect in the monitoring process are the actors (VECO and its boundary

partners) and their respective contributions (strategies and outcomes) in relation to the desired – more tangible – results in the value chain. A regular in-depth reflection is organized with farmers, partners, VECO and other relevant actors during a bi-annual multi-actor chain meeting. Farmers, partner organizations and VECO come prepared with own initial reflections on their actions which is - along with additional chain data (primary and secondary) – shared, discussed and analysed in a facilitated participatory sense-making process. The info and insights generated during these events form the basis for decision-making and programme steering. At VECO level, a bi-annual programme steering meeting follows the chain meetings to share progress in the different working areas, discuss cross-cutting and strategic programmatic issues and facilitate knowledge sharing among VECO staff. The main benefits of the PLAs are aimed to come from the insights obtained during the monitoring and learning process rather than from the results presented in reports. The core idea is that the sense-making process generates new knowledge, supports learning, plans and motivates future activities, and builds M&E capacity and reflective thinking of the different people involved. In addition, it facilitates feedback and negotiation amongst the different actors involved and is seen as a crucial way to move towards improved accountability, i.e. a process through which VECO makes a commitment to respond to and balance the needs of stakeholders in its decision-making processes and activities, and delivers against this commitment (Blagescu et al., 2005).

VECO also invests in relevant systematic documentation of the information obtained, lessons learned and decisions taken during the M&E process. An internal Chain Interventions Report, a living document with facts & figures of the specific chains and an integrated finance-activity-strategy database are aimed to support better reflection and analysis as well as to allow M&E findings to be more easily shared and communicated.

6. Conclusions

The first reflections and insights of 1,5 year implementation show that the actual M&E practice is benefiting from the rather long and in-depth design process. There are clear aspirations and guiding ideas that have shaped (and still do) the applied monitoring and learning methods/processes. There is commitment from VECO staff at different levels and partner organisations actively participate in the newly established PLA events (chain meetings) and have welcomed the simplified reporting systems.

Adequate resources (staff and budgets) have been put in place to support and guide the PLAs and initial investments have been made to strengthen some of the organisational conditions.

Related to the main purposes of the PLAs - namely facilitating an improved *planning, learning and accountability* process – some initial benefits can be observed. VECO programme staff and partners have a better understanding of the value chains and the roles, contributions and expectations of the different key actors in the programme have become clearer during both the Intentional Design process and the ongoing monitoring process. The continuous chain intervention and outcome monitoring has already resulted in a series of strategic programme adjustments during the last 1,5 year: new chain results, more focused chain interventions, new (type of) boundary partners, phasing-out existing partnerships, a changed organisational structure and an investment in new expertise (new staff, ...).

Although there is a general feeling that the PLAs is well designed, fits the realities of the programme and can potentially lead to a meaningful practice, the monitoring and learning process is still fragile and could erode or undermined if not carefully taken care off.

- It is clear that a continuous investment in preparation, methods, facilitation, motivation and keeping momentum for the face-to-face meetings - as the core PLA events - is required.
- Upward accountability is still dominant and influences the monitoring and learning practice towards an *intra-organisational monitoring* perspective - focusing on VECO's own monitoring needs, learning process and information flows – and can, if not consciously paid attention to, undermine the monitoring process which aims to facilitate a change process based on the viewpoints of and in collaboration with the local chain actors, i.e. *institutional monitoring and learning*.
- A focus on the process and the contributions of actors involved still requires good insights and relevant data on the more tangible chain outcomes and impact changes. A further exploration on *'intelligent'* information seeking methods and processes for this level will be required: detailed

quantifiable data if relevant and useful and/or macro information or emerging patterns when large number of variables are at play and where the measurement of pre-determined indicators is not possible or not contributing to an improved understanding of the reality.

- Through the monitoring of progress markers and strategy maps, OM tends to generate an overload of data. For example, the time spent on collecting, recording and reporting this data limited at first the attention for other monitoring processes such as sense-making.
- A related challenge for programme staff is the compilation of accurate, useful and presentable synthesised information, from the data generated during chain meetings or through other data collection methods. If data is not presented in a synthesised and 'absorbable' way, it has a direct influence on the quality of the interpretation.
- It is clear that a learning-oriented and OM-based M&E system requires a high level of M&E capacity of the programme staff and partners. For example, assessment and analysis of behaviour changes of boundary partners is less straight forward than a discussion on the activities carried out by the partners. Different approaches and methods for assessing progress markers are explored. It is clear though that to maintain the reflective approach in the monitoring process, continuous support and facilitation by a specialist M&E coordinator will be necessary.
- Having an OM-based programme framework is no guarantee for 'dealing with complexity'. It definitely has elements in its design that support a *process-oriented* monitoring process, allow for adjustments and to anticipate to important unintended changes in the course of the programme. But it is perfectly possible to use the OM-based programme framework as a linear and causal programme model based on a set of pre-determined information needs - whereby indicators and progress markers are used as a checklist and yardstick for success. Using OM to its full potential for guiding/managing complex processes depends entirely on the quality and nature of the monitoring practice.

7. References

- Blagescu, M, De Las Casas, L, and Lloyd, R (2005), *Pathways to accountability: A short guide to the GAP framework*. London: One World Trust.
- Britton, B (2005). Organisational learning in NGOs: Creating the motive, means and opportunity. *Praxis Paper, No.3*. Oxford: The International NGO Training and Research Centre.
- Dlamini, N, (2007), Transparency of process: Monitoring and evaluation in learning organizations. *In CDRA Annual Report 2005/2006*, 20-31. Retrieved December 15, 2007, from www.cdra.org
- Earl, S, Carden F and Smutylo, T (2001), *Outcome mapping: Building learning and reflection into development programs*. Ottawa: International Development Research (IDRC)
- Estrella, M and Gaventa, J (1998), Who counts reality? Participatory monitoring and evaluation: A literature review. In *IDS Working Paper No 70*, Brighton: IDS
- Guijt, I and Ortiz, N, (2007), *Outcome Mapping 2nd generation*. Powerpoint presentation Outcome Mapping workshop Wageningen (December 2007, Wageningen International)
- Guijt, I (1998), Participatory Monitoring and Impact Assessment of Sustainable Agriculture Initiatives: An Introduction to the Key Elements. *SARL Discussion Paper No. 1*. London: IIED Sustainable Agriculture and Rural Livelihoods Programme (available at <http://www.iied.org/pubs>)
- Horton, D, Alexaki, A, Bennett-Lartey, S, Brice, K, Campilan, D, Carden, F et al (2003), *Evaluating capacity development: Experiences from research and development organizations around the world*. Ottawa: ISNAR, IDRC and CTA.
- International Fund for Agricultural Development (IFAD) (2002), *Managing for Impact in Rural Development. A Guide for Project M&E*. International Fund for Agricultural Development (IFAD), Office for Evaluation and Study (available at <http://www.ifad.org/evaluation/guide/index.htm>)

- Patton, Q M (1997), *Utilization Focused Evaluation: The New Century Text (3rd Ed.)*, London: Sage Publications.
- Reeler, D (2001), *Unlearning – facing up to the real challenge of learning*. CDRA: Cape Town
- Senge, P, Ross, R, Smith, B and Kleiner, A (1994), *The fifth discipline fieldbook: Strategies and tools for building a learning organisation*. London: Nicolas Brealey.

Case 3A. Evaluation of Dutch Support to Capacity Development: Evidence-Based Case Studies.
Case Presentation on SOCSIS, Somalia
Tom O. Olila and Piet de Lange

1. About the Case

a. Basic context

This case study concerns the Strengthening of Somali Civil Society Involving Systems (SOCSIS), a capacity strengthening programme implemented with the support and facilitation of Oxfam Novib in conjunction with regional consultants. The focus of SOCSIS has been on institutional capacity building for local Civil Society Organizations (CSOs) in Somalia, focusing mainly on human rights/rule of law, gender and peace building. The SOCSIS capacity development process targeted trainee consultants, local CSOs, traditional elders and local authority leaders. The SOCSIS study focused on the period 2000 – 2008 and was carried out by Strategic Connections (Tom Olila) in collaboration with Context, international cooperation (Fons van der Velden).

The SOCSIS evaluation has been one of the case studies under a series of seven evaluations titled '*Evaluation of Dutch support of capacity development; evidence based case studies*' that were initiated by the Policy and Operations Department of the Dutch Ministry of Foreign Affairs (IOB). The purpose of the overall capacity development (CD) evaluation is to respond to the need for knowledge and new insights that contribute to further policy development on capacity development of IOB and Dutch Non-Governmental Development Organisations, jointly herein referred to as the Dutch Development Partners, (DDPs) and their partners in developing countries. The evaluation thus looked into how and under what circumstances capacity has developed with an aim to generate knowledge and insights for further policy development and how support for CD can be provided more effectively.

b. Type of evaluative practice

The study can best be described as being a developmental evaluation – it was largely explorative and descriptive in nature and has a strong focus on learning with a view to support innovation, adaptation and to stay in touch with what's unfolding. The evaluation design in this regard aimed amongst others to capture system dynamics, interdependencies, and emergent interconnections. The evaluator was part of a 'larger' study team, acting as a learning coach bringing evaluative thinking to the table.

c. Methodology

One key aspect of this process was to ensure that the methodology of the study (including indicators) were as far as possible 'bottomed up'/contextualized. The evaluation stressed the importance of the Southern views and experiences. The study was carried out as a collaborative learning process with the active participation of Southern partner organisations, Oxfam Novib, and the study team, based on principles of action learning and grounded theory. The nature of the evaluation was 'explorative and descriptive' with an emphasis on 'what has happened' and 'what has emerged' ie both planned and emergent development.

A briefing workshop aimed at ensuring a joint understanding of the study's conceptual framework; joint development of sub-questions and indicators; fine tuning the evaluation methodologies; preparation for the data collection, documentation, and triangulation was held in Nairobi between September 28 and October 3, 2009. This was followed with a start up workshop with Oxfam Novib aimed at collecting primary data at the level of SOCSIS and to prepare for further field based primary data collection.

The actual 'field visits' for data collection took place between 4th and 14th January 2010. For practical logistical and security reasons¹⁰, it was been decided that extensive 'field visits' would not be possible, most respondents were thus invited for a data collection workshop in Hargeisa. To facilitate data gathering, the study team used a multifaceted approach during and outside of the evaluation workshops. Some of the key methodologies used for this evaluation included secondary data review; learning histories (timeline analysis); story telling; targeted/key informant interviews; focused group discussions; questionnaires as well as gender and power analysis. The outcome of the collaborative research was then compared with information obtained from other sources for purposes of triangulation.

The data collected from the various sources (including the questionnaires) has been analyzed and the report developed by the consultant. Finally a Joint learning workshop was organised for the larger Partos evaluation with representatives of the DDPs, Southern partners and members of the study teams to prioritise/cluster issues and decipher learning.

2. Values and Quality

a. Values (and their negotiation)

Participation - the evaluation was carried out as a collaborative learning process (involving donors, Southern partners, and evaluation team). This involvement was at entailed active involvement these actors in the preparation of the study, primary data collection, and distillation of the major research findings, triangulation and documentation of major lessons learned as well as dissemination of research findings. All key methodologies, questions and indicators have been bottomed up with the active involvement of key stakeholders.

Context sensitivity - The evaluation appreciated that capacity development takes place within a certain context and is influenced by it. Organisations are thus seen as non bounded entities rather as open, complex systems. The evaluation appreciates capacity development as an internally driven (endogenous) process and tried to understand the interrelatedness of context, capacity development, outputs and outcomes with the role and support given by external actors (including donors).

Methodological appropriateness - The evaluation has been initiated and, to a large degree designed, by DDPs with minimal involvement Southern partners and national consultants at the onset. Space was thus created during the study process, for contributions from Southern partners/national consultants with ample attention paid to their understanding, acceptance and internalisation of the conceptual framework, design, methodologies and drivers for the study as well as to check the appropriateness of the same.

Negotiation - The organization of the briefing workshop in Nairobi was appreciated, but participants initially experienced the study as fixed and foreign in many ways. The overall Terms of Reference and concepts such as the open systems approach, and certain aspects of the research methodology (power cube, storytelling etc) were seen as having been developed from 'elsewhere'. However, there was a reasonable level of acceptance to work with these; national consultants also had the room to use other additional methodologies that they were familiar with, as long as these complemented a number of 'obligatory' methods that had to be used by every other case study within the larger evaluation.

b. Quality

A number of quality aspects/standards were implicitly aimed e.g. ensuring relevance and usefulness; participation of key stakeholders; impartiality and fairness; report quality; proper planning; respect for human dignity and diversity; financial prudence and accountability; context specificity and clear/accurate documentation and accuracy. A number of safeguards/ strategies were adopted to ensure these quality/standards amongst them the following:

¹⁰ Oxfam Novib Protocol at the time did not allow for visits to South/Central Somalia & Puntland.

i) Utility (=the evaluation will serve the information needs of intended users)

All key stakeholders were identified beforehand and strategies for their involvement set – the principle users of the evaluation outcomes viz. the Dutch development partners have been part and parcel of the process throughout the evaluation.

The selection of study team was carefully done with the participation of key stakeholders (Partos, Oxfam Novib and Context international corporation) while considering their experience, expertise amongst other set criteria.

Information collected was tailored to address specified evaluation questions – sufficient time was dedicated towards appreciation of these study questions. Data collection methodologies were 'tailored' to ensure that information towards the same was gathered.

Sufficient guidance was provided on the outlook of expected study reports to ensure that these clearly described the program, context, purpose, procedures and findings of the evaluation. Sufficient peer review and feedback was provided on the report both before and after the joint learning workshop by key stakeholders to ensure quality of the same.

ii) Feasibility (= evaluation will be realistic, prudent, diplomatic, and economical/prudent).

Being a truly multi-stakeholder process, a continuous process of negotiation was put in place to ensure the cooperation of the various parties. A lot of tact has been applied to handle the power dynamics at play within the aid chain and amongst stakeholders involved. Key stakeholders whose interests had to be balanced included the Dutch development partners (IOB, Partos, Oxfam Novib Netherlands & Nairobi); the evaluation team (Dutch team leader, national consultant, evaluation counterpart); the beneficiaries/respondents (Somali NGOs & consultants) as well as authorities in Somalia.

A lot of detailed information was developed to guide the evaluation as part of the inception report as well as the briefing and start up workshops reports - these clarified amongst others the evaluation timetable, purpose, methods, resource requirements needs, as well as the scope (geographical, time period, unit of analysis).

iii) Propriety (=evaluation is conducted ethically, & considers welfare of those involved)

A key belief of this CD study was that there is rarely a capacity vacuum. The evaluation thus endeavoured to ensure a fair examination and recording strengths and weaknesses of the program so that strengths can be built upon and problem areas addressed.

Security concerns were seriously taken into consideration during the study - stringent security protocol had to be pursued to ensure as far as possible the safety of the evaluation team and the respondents.

Due to the complex nature of the process, obligations of the various parties involved in the study was carefully reflected upon and agreed to in writing (what is to be done, how, by whom, when) – a few sticky points had to be continuously be renegotiated.

The resource allocation towards the process was negotiated with and contributed by key parties to the process to ensure prudence, cost effectiveness and accountability.

iv) Accuracy (= evaluation will reveal and convey technically adequate information)

The study design was such that a thorough appreciation of the context was done in advance and its likely influences on capacity development detailed out.

The purpose & procedures of the study were clearly discussed and described in enough detail as part of the inception as well as the briefing and start up workshop reports.

Through the peer review process by the national consultants involved in the larger Partos evaluation, as well as the feedback from the study team and Oxfam Novib, the study report was subjected to a thorough review to ensure quality of the same (legitimacy of facts, fairness, response to evaluation questions, clarity, simplicity, completeness and general adherence to agreed reporting guidelines, etc).

3. Complexity and Robustness

a. Complexity

There has been the involvement of many different (and at times competing) stakeholders with diverse vested stakes in SOCSIS including donors, local NGOs, local consultants, 'authorities' etc – all these involvements, interactions and interests had to be looked into

Socially complicated situations such as the insecurity (and its causes, consequences) in Somalia posed a challenge of coordinating and integrating many players involved in the SOCSIS program and process

Further, the uncertainty in the context of Somalia makes cause-effect relations difficult to track - outcomes are thus largely context dependent and emergent in nature. Relations are largely nonlinear in which actions have unpredictable reactions and implications.

The situation within the Somali context is also highly dynamic in which interactions within, between, and among subsystems and parts within systems are volatile and always changing hence difficult to plan (including the evaluation process itself)

SOCSIS in itself is not an independent distinct organization rather a bringing together of various persons, processes and methodologies. This makes the boundaries of the unit of analysis to be rather unrigimented which further complicates issues of study scope.

As a result of this complexity, a wide array of methodologies has been applied to this study with a view to 'fitting' the various study respondents (see section above).

b. Robustness

The methodology and process of the evaluation was quite indirect, starting first with practice rather than with theory. A lot of preparatory work in form of the inception reports; theoretical foundations/frameworks as well as the briefing, start up and joint learning workshops demonstrate the rigour involved.

Additionally, the deliberateness and consciousness about certain values, approaches and standards reflect the degree of professionalism that the process entailed.

Further, the complexity of the context of the evaluation (Somalia), which is more within the periphery (of the ordinary), makes the evaluation unique as it ventures into the rarely explored territory.

The process also is further complicated by the vastness of the stakeholders whose interests had to be taken into consideration.

4. Improvements

Some of the given methodologies such as the power cube/analysis, were difficult to use with various stakeholders as these were new to them - where such are used in future, sufficient time ought to be dedicated to ensuring appreciation of the same – as far as possible, familiar methodologies that would generate the same information would be applied (and or used complementarily).

All key persons involved in an evaluation, especially the Southern based evaluators should be involved as far as possible in the earlier stages of the development of the evaluation approach and methodology.

The conceptual framework and evaluation tools, approaches and methodologies should as far as possible be simplified and deconstructed in order to make these less voluminous, theoretical and technical in language.

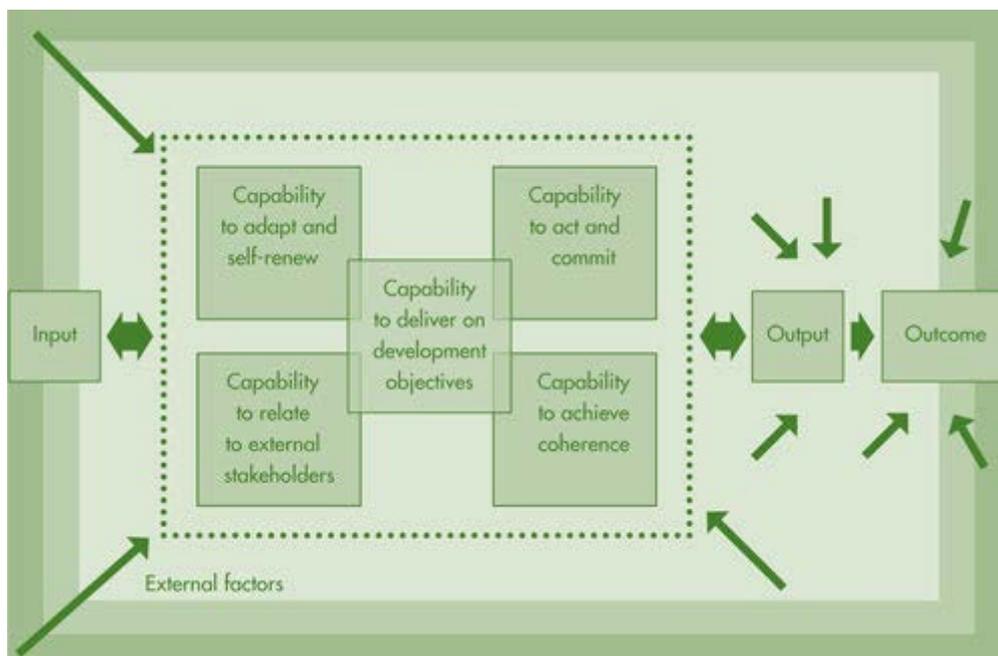
For an evaluation with as wide a scope as was the Partos study, the evaluation framework should allow sufficient room to capture the diversities and complexities of various case studies (e.g. flexibility with regard to methodologies, processes and time allocation for data collection etc).

Case 3B. Evaluating capacity development support

Piet de Lange

IOB carries out independent evaluations of policies and operations in all fields of development cooperation. Recently, IOB launched an evaluation of Dutch support for capacity development that will result in a synthesis report based on a series of evaluations of the support for capacity development provided by seven organisations in 17 countries, most of them in sub-Saharan Africa. The seven organisations are the Ministry of Health (Ghana) and six Dutch NGOs: Agriterra, the Netherlands Commission for Environmental Impact Assessment (NCEIA), the Netherlands Institute for Multiparty Democracy (NIMD), Partos, PSO and SNV (Netherlands Development Organisation). Although these organisations work in different fields, they are all directly involved in promoting and supporting capacity development. The evaluation is intended to respond to the need for knowledge and insights that will contribute to the future policies of the ministry, Dutch NGOs and their partners in developing countries. The evaluators will look at how and under what circumstances capacity has developed, and attempt to identify the factors that have influenced the effectiveness of the support provided by the Netherlands government and NGOs.

1. Open systems approach



Recognising that capacity is elusive and often transient, the evaluation will not use a predefined concept of capacity, and will regard organisations and networks as open systems with permeable boundaries. This approach, summarised in the diagram, will allow the evaluators to focus on how capacity has developed from within, rather than to look only at what outsiders have done to support and promote it.

The adoption of the open systems approach has significant methodological implications. In particular, the framework and the indicators used in each evaluation must be contextualised and related to the perspectives of both the Dutch and Southern partners with regard to capacity development. Thus the indicators and operational criteria will be determined in cooperation with local stakeholders. Southern partners will be fully involved in the evaluation process from the outset, whether as members of reference groups, as resource persons, or in conducting the fieldwork for each of the seven evaluations. In summary, the evaluation will underline the relevance of Southern partners' views of and experiences with capacity development.

In the analytical framework shown in the diagram, the broad concept of capacity is divided into five core capabilities that every organisation and system possesses. None of these capabilities can by itself create capacity. They are strongly interrelated, and provide the basis for assessing a situation at a particular moment, after which the capacity of the system can be monitored and tracked over time in order to assess how it has developed.

The IOB will conduct the evaluation in collaboration with a network of partners:

- external advisors, including staff of the European Centre for Development Policy Management (ECDPM), Utrecht University, Tilburg University and Erasmus University Rotterdam, and Southern advisors;
- facilitating organisations (methodology development, communication);
- and consultants based in the North and the South.

For each evaluation, a reference group and an evaluation team have been established, consisting of Northern and Southern members with a background in capacity development theory and practice.

The final synthesis report of the evaluation, which will be available in December 2010, will present the key findings and the lessons learned. Together with the more detailed reports on each of the seven organisations, it is hoped that the evaluation will make an important contribution to the international debate on capacity development.

2. Further reading

- Engel, P., Keijzer, N. and Land, T. (2007) *A Balanced Approach to Monitoring and Evaluating Capacity and Performance: A Proposal for a Framework*. ECDPM Discussion Paper 58E. www.ecdpm.org/dp58e
- De Lange, P. and Feddes, R. (2008) *General Terms of Reference, Evaluation of Dutch Support to Capacity Development?IOB*.
- IOB: www.minbuza.nl/job-en

Case 4. Social Return on Investment- A case study on the Grain Bank to offset the risk of severe drought conditions in the Malpura Block of Rajasthan –India
Mr Manish Prasad and Mr Madan Lal Sharma

1. Context

Harsh living conditions, poor quality land and water resources, few biological resources- these are conditions with which the people of Rajasthan have lived for centuries. But successive years of drought are undermining the coping mechanisms of the people. Under these pressures of deteriorating climatic conditions CECOEDECON (an Indian NGO) has continued to work with the partner communities for developing their capacities to develop innovative strategies to effectively confront the impending challenges that they face with.

Here is an example of a village Ratanpura in the Malpura Block of Tonk District in Rajasthan, where the community members were able to establish a 'Grain Bank' to offset the risk to their food grain requirements during the drought period of 2008 - 2009.

The village Ratanpura is a multi caste village having a population of more than 500 people especially belonging to the schedule caste and the other backward castes. The main occupation of the village is dry-land agriculture and animal husbandry. Around 95 percentages of the households have land but most of them have to go for daily labour during the lean season to meet the family requirements. Both men and women have to work hard to earn their living.

As draught is a regular phenomenon, a project to address the issue of disaster risk reduction was implemented in the area. The establishment of a Grain Bank was one major activity of this project. The main stakeholders of this project include the Small Farm families, the village development committee members, (The VDC members include men, women and youth members) the CECOEDECON Staff, the farmers from other villages, traders, etc.

2. Type of evaluative practice and methodology

This case study is measuring the impact of the Grain Bank at the community level using Social Return On Investment (SROI).

SROI is a methodology to understand, measure and report on the social, environmental and economic results of an intervention or social firm. Conventional financial return methods like cost-benefit analysis do not capture the social and environmental changes of such organisations (positive as well as negative), whereas SROI recognises their full economic, social and environmental 'value'. Conducting an SROI analysis is a way of making the invisible value of things that are essential to quality of life both visible and measurable; in short, of valuing what matters. At the Grain Bank, both the economic and social returns are assessed through the SROI steps explained below.

All the nine steps involved in the SROI process were applied. These steps are –

- I. Defining the boundaries (objective and scoping)
- II. Identification and selection of key stakeholders
- III. Developing a theory of change
- IV. What goes in (identify inputs for each outcome)
- V. What comes out (identifying results)
- VI. Valuation
- VII. Calculation of SROI ratio
- VIII. Verification
- IX. Narrative

The time taken for this was four and a half hours. It was organized in a village setting. Appreciative enquiry and problems analysis were other methods combined in the process. The data collection process was

mainly through the direct interaction with the groups of community members, which included men, women and youth. Observation of the grain bank operation and triangulation of the facts with individual case stories and records were done.

Everyone in the evaluation team made their own notes during the interaction and they were shared and discussed in a separate meeting. The compilation of those points was assigned to a particular individual who was responsible for the final reporting.

3. Values

The values that prominently reflected in the process were 'mutual understanding and trust' among the members who could openly share their efforts; transparency in revealing and divulging all the information, records and 'shared ownership' of the results etc. The other values manifested in the process were of 'effectiveness' and 'empowerment'.

The process itself was so empowering that when the people were narrating their sincere efforts openly, the other members were very appreciative and encouraging. There were differences of opinion on the issue of the prioritization of the most important impact/change.

4. Quality

The SROI methodology has been a *tailor made tool* for the existing monitoring and evaluation system of the organization which has been doing a lot in terms of impact monitoring with the active participation of the community members.

The SROI tool has been used in many case studies and we have been using it very *flexibly* without compromising the quality. Transparent, participatory, objectivity, usable and sustainable are the other qualities envisaged.

The evaluation standards that we have tried to uphold through this tool include utility, feasibility, propriety and accuracy.

5. Complexity

Since its inception, Cecoedecon has tried and tested various approaches to development emphasizing multiple strategies. The organization believes that there is a need to contribute to innovation in the development sector and that alternative, innovation approaches and models need to respect or build on what is already there in the development practice especially at the level 'where the action is', such as amongst the members of the Grain Bank in the village of Ratanpura. Since 2008 the organisation and its focal groups have been experimenting with SROI, as a methodology to measure impact. SROI allows for respecting Multistakeholder Processes, institutionalising the learning and to go beyond the different levels of staff. Another important aspect is that SROI respects unexpected outcomes and recognizes the contribution of the communities. For Cecoedecon it is important to value respect, build upon what is already there and moreover, create empowerment within the communities by using tools such as SROI.

6. Robustness and improvements

SROI is a new methodology that has been piloted in the social sector and as such it has become a trial and error process. No given solutions to the problems are documented and available. The perceptions of change and their value to individuals and groups vary very much at different levels. The standardization process is cumbersome. Therefore the methodology has to take into consideration a majority perception to estimate the value of changes and results.

SROI is a participatory tool where in which the direct beneficiaries and their perspectives are given prominence. The validity and authenticity could be verified easily and the result is declared immediately as a part of the process. The process leaves out great confidence and a sense of ownership as well. Presenting the financial value created by social and environmental benefits can be particularly important in

justifying investment in activities that do not generate a direct financial return. SROI can demonstrate the value to society of social or environmental returns of any intervention that from a strictly economical viewpoint may look unviable.

The preparations for this exercise especially in terms of selecting the participants for the direct interaction need to be carefully made. Even though the number of participants could be more but the people who were directly involved in the intervention could be more actively engaged. Since it has been a time consuming process, it would be good to structure the exercise in phases as per the convenience of the participants.

7. More information

More information on Cecoedecon can be found at: <http://www.cecoedecon.org/>

More information on SROI can be found at: <http://sroiseminar2009.wordpress.com> and www.developmenttraining.org (Contextuals, 4, 8 and 9)
http://contextinternationalcooperation.files.wordpress.com/2010/11/sroi-practical-guide_context-international-cooperation.pdf

Case 5. Aflatoun applying mixed methods in a network that supports children through education

Simon Bailey and Andrew Magunda

1. Organizational Overview

Aflatoun's mission is *'inspiring children to socially and economically empower themselves to be agents of change in their own lives and for a more equitable world.'* It works in partnership with organizations to deliver a programme of Child Social and Financial Education in schools and non-formal centres. The Secretariat, based in Amsterdam, coordinates programme dissemination, advocacy, research and network activities. Currently, Aflatoun is working in 41 countries with 50 different partner organizations. We divide its work into three distinct pillars:

- **Programme:** To deliver high quality social and financial education to children in 75 countries
- **Concept:** To become thought leaders in the area of social and financial education
- **Network:** To develop partnerships and raise resources for the Aflatoun movement

Each pillar has its own strategic objectives, indicators and activities that guide work within the organization and that have to be assessed.

2. Methods Used

Being a new organization with multiple organizational objectives has led to a number of challenges related to determining the impact of our work. This includes:

- How do you assess an organization that has a number of different goals in a coherent way?
- How do you work in partnership with organizations on learning and evaluation of a programme?
- How do you evaluate the effectiveness of a Secretariat in a network to ensure accountability?

Balancing these different goals and dealing with these challenges has required the creation of a coherent single framework that we call the Aflatoun Quality Assurance and Impact Assessment System. This has been done through conceptualizing our work as consisting of quality assurance, output tracking, outcome and impact assessment and through the assignment of appropriate assessment tools to each area. This includes partner selection, output and outcome surveys, qualitative research, network/secretariat assessment, SROI, and random assignment.

While the different tools and aims at times seem disparate, having an overarching approach keeps learning, monitoring, and evaluation activities close to the key organizational objectives. This planning keeps this work integrated and has begun to provide information at appropriate times for organizational and network decisions.

3. Quality and Values

The translation of this approach into practice has had both successes and challenges. Challenges include maintaining quality across all projects, balancing different and sometimes competing organizational objectives, ensuring partner take up, and facilitating organizational and network learning. Successes have come from flexibility in developing and adapting approaches, finding and trusting interested individuals, as well as publicizing results.

Partnership is the key value of the Aflatoun approach, therefore a representative from one of the local organizations that delivers the programme will present about his involvement with Aflatoun. Issues that will be discussed include his role in how evaluation tools are developed, how they are shared with partners, the types of activities that partners participate in, and some preliminary results.

4. Complexity, Robustness, and Improvements

This case will try to show how mixed methods can satisfy different organizational ends. While the individual different objectives may be either simple or complicated, managing all these aims together makes the approach complex. Complexity implies different trade-offs and balancing competing aims but this has its benefits as well.

While not all findings can be equally robust, flexibility allows for new approaches to be easily developed and new findings generated. Using multiple approaches has led to a number of key findings that has helped orient the organizations about key issues.

Like all approaches, Aflatoun's approach is a work in progress. We have made many mistakes but are young and flexible enough to acknowledge them and appropriately change course. This is made possible for two reasons. First, partners are consulted and are able to participate in the development/redevelopment of approaches and tools. Second, our approach is based on internal –as opposed to donor- objectives. For this reason, we are able to take risks and also change or alter course when things need to improve.

Case 6. Formative Evaluation and Outcome Monitoring System for Governance in Cambodia *Sandip Pattanayak, CMS (India)*

1. Context

Cambodia is one of the poorest countries in South-East Asia. More than a third of the population of Cambodia lives below the \$1/day poverty line (129 in the Human Development Index). Emerging from years of civil conflict, Cambodia has made remarkable progress since 1991. In its effort to deepen its development reforms, the Royal Government of Cambodia has integrated the Millennium Development Goals (MDGs) into its overall strategies and plans, which are reflected in the Rectangular Strategy (2004) and National Strategic Development Plans (2006-2010). High levels of poverty, limited government accountability and weak civil society are key challenges for development.

Decentralisation and Deconcentration (D&D) of investments, programmes and services to political structures and administrative institutions at the sub-national level is considered central to poverty reduction and development. The D&D efforts aim at consolidating and deepening the process of democratisation at the grass roots and promote local development and poverty reduction. One of the principles is to strengthen local Commune/Sangkat Councils (CS Councils) and expand their powers, responsibilities and resources. CS Councils are the local self-governing bodies elected by people and govern a cluster of villages. The National Committee for Sub-National Democratic Development (NCDD) under Project to Support Decentralisation and Deconcentration, which is one of its major governance reforms initiatives, conceived the idea to establish a Formative Evaluation and Outcome Monitoring System focussing on the performance of the CS Councils. The objectives of the support are to:

1. Design and demonstrate a locally manageable, user-friendly and learning oriented formative evaluation system specific to the NCDD/PSDD context and to its projects and investments;
2. Demonstrate and document ways and means of feeding back the formative evaluation findings and products into the D&D policy debate and into the NCDD/PSDD programming and decision making systems;
3. Develop local capacities to manage and sustain the above formative evaluation systems and processes via know-how transfer.

CMS is working in partnership with the Economic Institute of Cambodia in the M&E support to NCDD.

2. Type of evaluative practice

The system being established is a Formative Evaluation System, which aims at progressive improvement in the performances of the CS Council. It includes outcome monitoring, which informs and feeds into the performance assessment process. The performance assessment is an organisational self-assessment process leading to reflection, learning and action for improvement. This central process is complimented by feedback from the Citizens (Client of CS Council), Civil Society (Partners in Development) and Technical Support Staff (Facilitators of CS Council). The self-assessment by the CS Council completes the loop of the 360-degree approach.

3. Methodology

The methods employed to do the study were mostly qualitative in nature, except Household Survey, which used questionnaire for data collection. However, the entire study was based on a constructive engagement process with the CS Councils from the beginning of the study. Hence, the methods should be seen in conjunction with the processes. The qualitative methods employed were:

1. **At the CS Level** – Rapport building and engagement with CS Council, Analysis of secondary data¹¹, Process Mapping¹², Self Assessment of Performance, Strength and Weakness Analysis, Improvement Planning;
2. **At the Citizens' Level** – HH Survey and Focus Group Discussion¹³;
3. **At the Partner Level** – Feed-back from NGOs, Feed-back from Government Line Departments;
4. **At the Facilitators' Level** – Functional Analysis.

The sequence of the methods and time taken¹⁴ to complete each process/method was as follows: (i) Engagement with CS Council – 1 Week, (ii) Secondary data analysis – 2 Weeks; (iii) Focus Group Discussion – 2 Weeks, (iv) HH Survey – 6 Weeks, (v) Feed-back from NGOs and Feed-back from Line Departments – 2Weeks; (vi) Functional Analysis – 1 Week; (vii) Commune Performance Assessment (including process mapping, self assessment of performance, strength and weakness analysis and improvement planning) – 2 Weeks; and (viii) Feed-back to Provincial Authorities. For each of the three provinces, a province level report has been prepared, which analyses the performances of the four CS in that province.

4. Values (and their negotiation)

The process is being guided by the following values and principles:

- a. Engagement with the CS council to dispel fear and enabling the councillors to identify strengths and weaknesses and take improvement actions;
- b. The system design was kept flexible to allow for experimentation and innovation during the course of the study. The design, methods and processes evolved over a period incorporating learning from the field processes;
- c. Plurality of methods and data – The 360 degree approach was adopted to generate feedback from various stakeholder groups who influence and facilitate the performance of the CS Council.
- d. More emphasis was given on the quality of engagement processes rather than the completion of the study;
- e. Identify what was done well and what was not done well; identify learning and improvement actions and initiate actions;
- f. Identify strengths and weaknesses and work on it.
- g. Do not criticise. Encourage to embrace mistakes.

5. Evaluation Quality

The quality-related challenges foreseen were:

- Delivery of the assignment due to changes to the team of the local partner
- Slow absorption in terms of understanding study concepts, methods and overall study requirements within the local
- Stakeholder management on the concept;
- Data quality in terms of collection and analysis;
- User-friendly System Protocol and TOT development
- Capacity building of EIC team;
- Coordination and communication with the client

¹¹ Secondary information available with the CS Council on development projects implemented including projects implemented by NGOs and Government Line Departments; revenue and expenses; capacity development initiatives taken up for the CS Council and availability of key records and regular maintenance of records as a measure of transparency and accountability.

¹² Processes followed by the CS Council in development planning and implementation as against the processes described in official guidelines and manuals.

¹³ FGDs were conducted with community members in each village (One FGD per village).

¹⁴ The time given is average time to complete the process for a set of four CS.

The following processes were put in place to address the challenges:

1. CMS took the lead organisation role and was responsible for overall design, planning, execution, management and monitoring of the assignment. A full-time team of two consultants was placed on-site for a period of one year to anchor the process and take lead.
2. Capacity building was done through daily engagement with the national consultants, on-field demonstration and handholding and through in-house orientation. Concepts and process were kept simple. It started with methods and their application and then moved to processes and concepts.
3. Senior resources of the national partners were mentored to ensure that the skills stay within the organisation and institutionalisation takes place.
4. Fort-nightly briefing was done to the client on the progress of work and challenges. Post-pilot briefing was provided on the findings of the pilot. This was done to ensure continuity of communication and coordination.
5. The evaluation was pilot in nature, as it intended to pre-test the design, tools, methods and approaches. This was done through piloting at three different provinces and assessing utility, feasibility and accuracy of the proposed system. Improving the performance of the CS Council through a participatory process of reflection – learning and action was the intention of establishing the system. From the very beginning of the assignment, intensive effort was made to engage the elected councillors in the reflection-learning-action process. This was done to reduce fear of externally facilitated assessment and promote ownership of the processes and improvement actions. Care was taken to see that the findings help the C/S in reaching at an actionable conclusion. However, often, the requirements of process quality clashed with the demands of accuracy. In the evaluation principles it was decided that greater emphasis will be on quality and empowerment rather than accuracy and completion.

6. Complexity. Why is the evaluation focus, i.e. the object of study, considered ‘complex’? To what extent did this affect the choice of methodology, and how?

Cambodia is going through social, economical and political transformation. Decentralisation and deconcentration is ongoing and rapid changes are being made to create and strengthen political and administrative institutions at the sub-national level. Commune Councils existed as traditional institutions in Cambodia. These structures were constitutionally recognised and were allowed to contest politically in 2002-03. Since then a series of reform measures have been taken to strengthen these institutions of local self-governance. The CS is continuously evolving with progressive evolution of the decentralisation and deconcentration scene in the country. Cambodian society is strongly hierarchical in nature. Questioning positions of power or authority is generally not well accepted in the society. Moreover, the experience of Khmer Rouge period, deeply entrenched this behaviour. In such context, provoking the councillors to reflect, self assess, identify areas of improvement and take improvement actions were challenging. Hence, a more objective assessment method was adopted. The whole exercise was viewed with scepticism and fear. Councillors were apprehensive to rate something low or below average and then provide evidence for their low scoring. Lack of sufficiently trained human resources, limited financial resources, corruption affects governance at national, provincial and local level. Often, while discussing about strengths and weaknesses, the councillors were agitated and were of the opinion that organisational strengthening should be done at the national and provincial level first, then at the local level. All these elements required a lot of confidence building as a part of the process requirement.

7. Robustness

The assignment was pluralist, evolutionary and iterative. Study pilots were intended to experiment the design, tools and approaches and evolve over a period of time. It started with a conventional evaluation design, which covered key aspect of organisational performance assessment and developmental outcomes. However, during the pre-pilot phase, it was understood that the process quality and engagement of CS Councillors and other stakeholders was important than the findings per se. Some fundamental questions were asked:

- Why is this being conducted?
- Who will benefit from it?
- What changes are envisaged and how will the changes occur?

8. Lessons learned and potential improvements

In a multi-stakeholder evaluation assignment, create a coordinating mechanism like an Evaluation Core Group. The Evaluation Core Group would act as a decision-making body and would steer the evaluation assignment. It would have representation from all principal stakeholders and would meet at regular intervals for review, planning and trouble-shooting.

- Set the principles of evaluation and evolve a consensus around it. Be flexible to evolve and add to the principles when something new is learnt.
- Bring clarity and agreement between parties to the evaluation on what is to be delivered, what can reasonably be delivered and what is not to be delivered; if scope of work changes, renegotiate.

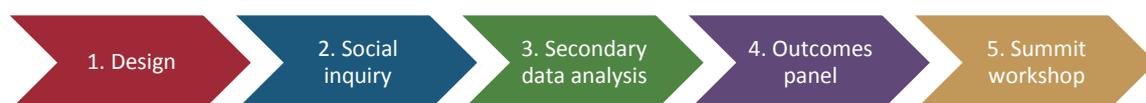
Case 7. Evaluation of the Stronger, Smarter Realities program, Australia. *Jess Dart and the Stronger Smarter Institute*

In Australia there is a big disparity between educational outcomes for Indigenous children compared to non-Indigenous children, and in the last 8 years educational outcomes have been either stable or declining.¹⁵ While Indigenous children have been staying longer at school, too many Indigenous students leave school without a formal Year 12 qualification. Nationally, the proportion of Indigenous students who achieved a Year 12 Certificate (as a proportion of students who were enrolled in Year 11 in the previous year), has decreased from 51% in 2001 to 46% in 2006. During that period the proportion of non-Indigenous students increased from 80% to 86% and the gap between Indigenous and non-Indigenous outcomes has widened.

This case is about a participatory evaluation of the first phase of the Australian “Stronger Smarter Realities Program” (SSR) that ran from 2006 to the end of 2008. This project was about creating systematic and transferable change by arming Australian educators with the belief, skills and capacity to make profound changes to the learning outcomes of Indigenous children. Over 3 years, the project aimed to engage principals, teachers and Indigenous community leaders from 240 schools with high Indigenous student populations, and support them to transform their schools in such a way to deliver dramatically improved educational outcomes for Indigenous students. The program is based on the premise that this can be achieved by providing a supportive educational environment, by providing excellent teachers and by having high expectations.

The program is delivered by the Stronger Smarter Institute (formerly Indigenous Education Leadership Institute (IELI)), Queensland University of Technology, Brisbane, Australia. It was funded jointly by two donors: the Sidney Myer Fund and the Telstra Foundation. The project has two streams: the Principal Leadership Program and Teacher Leadership and Community Leadership Program. This evaluation was primarily concerned with the Principal Leadership Program. The evaluation was completed in 2009 at the end of the first phase of the project by external consultants using a participatory approach. It was guided by two key questions i) to what extent has the SSR project contributed to excellence in Indigenous education in participating schools? And ii) to what extent did the SSR project influence the overall Indigenous education agenda?

The evaluation was both summative and formative in nature and largely focused on outcomes. It was summative in that it was conducted at the end of the first phase of the program and was required by the program funders. It was formative in that it was intended to influence the next phase and scaling up of the program. The evaluation used the “Collaborative Outcomes Reporting Technique” developed by Jess Dart. This is a mixed method approach that involved key stakeholders in a number of process steps.



Firstly, a design workshop was held where the theory of change was clarified and evaluation questions developed. This was conducted with program team members and key stakeholders in a participatory manner. Social inquiry included over 50 semi-structured interviews incorporating the Most Significant Change technique and 3 case studies from Indigenous communities. The data trawl involved collection and analysis of secondary documents and quantitative data on student outcomes from 10 schools.

The quantitative data, case studies and qualitative summaries were used as evidence to feed into an ‘outcomes panel’ with Indigenous educators who examined the data and created statements about: the extent to which the outcomes had been achieved; the plausible contribution of the program to these

¹⁵ National Report on Indigenous Education, 2006 DEEWR

outcomes and the quality of the data. The panel were selected as they were highly respected, had no vested interest in the program and had an excellent knowledge of Indigenous education policy and practice. The process culminated in an evaluation summit workshop that saw key stakeholders and staff deliberating over qualitative and quantitative data and creating recommendations. The consultants' role was collection and synthesized data and facilitation of the sense-making process with recommendations created by workshop participants.

While the quantitative data was limited in scope, the evaluation was noteworthy as it managed to capture some of the less tangible outcomes concerning 'breakthrough learning' and raised expectations for Indigenous children. The program itself has been very successful and is being scaled-up and delivered on a national scale. This evaluation has been highly influential as evidenced by all the recommendations been successfully implemented, and one Philanthropic funder stating that the evaluation was well-balanced and gave them confidence to provide further funding for the program.

Values and Quality

This evaluation was based on the premise that the values of the Indigenous panel members, program staff and key stakeholders were of highest priority. Funders' views were not considered. The evaluators attempted to "bracket off" their opinions and instead presented a series of data summaries to panel and summit participants for them to analyse and interpret. The evaluators felt they were not the right people to make judgements concerning the effectiveness of a program.

Values were surfaced and debated throughout the evaluation process. The theory of change created a 'normative model' for how program staff viewed the program and this was used as the organising construct for the evaluation. Program staff debated and helped create the evaluation questions. Quotations and data were presented in a non-homogenised manner to panel and summit participants. Vignettes were analysed using the most significant change approach and the reasons for their selection were documented. Participants debated the value and significance of data sources and came to agreement on the key findings of the evaluation. Dot voting was used to determine how different participants rated each issue and outcome. Participant analysis and judgements were used to frame the evaluation report in terms of how findings were grouped, and the recommendations came from the workshops.

The focus on quality was largely associated with *process quality*: ethical conduct; culturally appropriate methods; ownership of the evaluation process; ensuring that the evaluation provided credible but useful recommendations to inform the next phase of the program. All interviews were taped and transcribed. Data was analysed in participatory ways and by the consultants using thematic coding. A weakness of the evaluation was the quantitative data; the data sets were patchy and the sample was limited. It proved extremely difficult to acquire this data from schools. One of the recommendations of the program was to create a more robust quantitative monitoring system, for the second phase of the program.

While the program was not overly complex, the program outcomes and process were somewhat emergent, intangible and hard to measure. The "Collaborative Outcomes Reporting approach" has been successfully used with programs with higher degrees of complexity – often with no clear outcomes defined at the onset of the program. It is widely used for evaluating biodiversity conservation programs.

Case 8. Application of Multiple Rigorous Methods in Impact Evaluation: Experience from BRAC ELA Programme in Uganda and Tanzania

Abebul Zerihun

1. The case

Based on the lessons and experiences of BRAC's programme for adolescent girls in Bangladesh and with the support from Nike Foundation and MasterCard Foundation, a pilot was initiated in Uganda and Tanzania in 2008. This programme, named Empowerment and Livelihood for Adolescents (ELA for short), is implemented at the community level. All girls aged 13-22 who reside in one of the project's villages are eligible to participate. The club provides a platform for the programme's six components: provision of a safe space for socialization, life-skills training, livelihood training, financial literacy, credit services, and community sensitization. Each club has one Adolescent Leader/mentor, who is trained by BRAC to operate the club and carry out various activities. Livelihoods training are conducted by trained professionals.

This programme is seen as a pilot in which to experiment with innovative approaches, measure the results, and provide quantitative and qualitative evidence for possible scaling-up of similar interventions in the future. Because the evidence on what works in facilitating the transition of adolescent girls and young women to productive work is limited, particularly in Africa settings, impact evaluations are deemed as an essential part of this initiative. In May 2008, an impact evaluation was launched by BRAC Africa Research Unit in partnership with researchers from World Bank, London School of Economics, and University London College. The impact evaluations of ELA projects relied on rigorous quantitative and qualitative methods to measure the effects of the livelihoods and life-skills training, and microcredit services on the socio-economic well-being of adolescent girls and young women and their households.

The core tool of the impact evaluation is a randomized control trial (RCT), supported by rigorous qualitative tools, and systematically captured programme administrative data. In early 2008, the programme identified 10 branch offices in each country (5 in rural areas and the other 5 in urban) for implementation of ELA. In each of these branches, 15 to 23 clusters (or villages) were marked as potential sites for setting up ELA centres. Then a census was conducted to list adolescent girls between the ages of 13 to 22 in each of the locations. For each branch, 10 villages have randomly been selected as treatment and 5 as control. Therefore, we have 100 treatment clusters and 50 control clusters in each country (300 study clusters in both countries). Treatment groups are further randomized by 'financial + non-financial intervention' and 'only non-financial intervention'. 40 female adolescents in each village were interviewed, sample=6000girls/country. The activities related to setting up of ELA clubs started as soon as the baseline survey is done (June 2008). The first treatment group will not receive microcredit until the 2nd round of survey is conducted (June 2010).

In addition to the quantitative data collected, a series of qualitative studies were conducted from the initial stage of the programme to document the qualitative aspects of the evaluation programme. In January 2009, after few months of programme operation a documentation process was done to capture qualitative data on initial setting up and opening of clubs and how and why adolescent girls join such a club. Around the same time additional qualitative study was conducted to document selection process of mentors and performance of mentors in service delivery. As a result of mentors' study, a separate mentors' survey module was developed and administered to all ELA mentors in May 2009. Mentor's data was later merged with ELA panel data.

Key administrative data are linked with the core evaluation design (RCT). Adolescent girls participating in club activities are provided with *unique idno* and baseline survey participants who are participating in the club are assigned with the same *unique idno* they were given at baseline. As a result all administrative data (days attended, whether participated in life-skills training, whether participated in livelihoods training etc) is integrated with impact evaluation dataset.

During data analysis two types of comparisons will be conducted. First, programme participants will be compared to a matched sample of girls with similar characteristics from control villages. Note that

because girls self-select into the programme, programme participants cannot directly be compared to a random sample of girls from the control villages. The matched comparison will allow identification of the causal effect of the ELA programme, conditional on participation. Second, girls in treatment villages will be compared to girls in control villages, which will identify the causal effect of residing in a village in which the ELA programme participates even if the girl herself does not participate in the programme. Such girls may be affected through demonstration effects of the programme, or through the diffusion of information from other village residents.

2. Values and Quality

The two most debated values among various ELA stakeholders are: 1) *further classifying intervention clusters in to two groups, microfinance and social interventions*. Practitioners and donors strongly argued that this would complicate and negatively affect programme implementation as there is a strong assumption that microcredit service is the most attractive programme component (this fact was later supported by the result from a qualitative study conducted at the initial stage of program implementation). In the other hand, the evaluation team thought that the pilot programme would be a great opportunity to provide evidence to the debate on approaches for adolescent girls around microfinance versus social intervention. After weighing its pros and cons all stakeholders reached on an agreement to further randomize treatment villages into 'only club activities' and 'club + microfinance'. The group that are not receiving microcredit service will receive after the second round survey. 2) *Community participation versus random assignment of communities*. In ELA, active community participation is core for programme success and sustainability. Each treatment community are expected to donate free houses to host club activities, parents/guardians are required to pay around \$1 for each adolescent girl they send to BRAC ELA club, and community members are required to mobilize an additional amount of \$200 for each club. Resources mobilized are used to sustain the club in the long term. With random assignment of treatment villages (RCT), in some cases program staff were forced to work in communities were its extremely difficult to gain the minimum required support. Such difficulties had brought greater challenge in maintaining the integrity of evaluation procedures. For example program team were forced to abandon implementation in 3 treatment villages (out of 100) because they could not manage donated houses. To address such challenges and to maintain overall quality of this complex evaluation project, BRAC has set up an in-house but independent research unit in both Uganda and Tanzania. Researchers from the Unit are responsible to conduct periodic monitoring to ensure that the evaluation is being implemented uniformly as per the design. Frequent feedback was provided to senior management regarding treatment status of villages and potential risks of contamination. Similarly, programme management team keeps the research unit in the information loop regarding program implementation and changes in implementation.

3. Complexity and Robustness

A key focus of the impact evaluation of ELA is to measure the overall impact of the intervention on empowerment outcomes among adolescent girls and their families, but another; probably complex part of the study is to unpack *how* the parts of the intervention work differently. The impact evaluation includes several outcome indicators that measure the wider impact of the programme. For example in Uganda, the high prevalence of new HIV infections among young women and the ability of livelihoods programmes to mitigate those risk through reduction of behaviours such as transactional sex is the important outcomes to influence through adolescent programmes. The other general outcomes includes their health and social knowledge, level of social cohesion and reduced gender discrimination, whether programme participation can enhance their level of aspiration and cognitive ability to make more informed decisions, and inter-generational transmission of attitudes. Such complex interventions that have a multi-disciplinary base and that aim to capture the wider impacts of the programme have impacted the complexity of the evaluation design. For example to separately estimate the effect of training and microfinance the study used two groups of treatment communities.

The other complex aspect of this evaluation is its adaptation to intervention changes. A few months ago, the program identified human rights and legal education as a key intervention component effectively tied to economic empowerment to adolescent girls and young women. Currently as programme prepares to develop human rights and legal education, the research team is preparing to collect data on the

challenges of and success of ELA program that incorporate human rights and legal education using ethnographic research.

The use of randomized comparison group (RCT), the triangulation of most data collected by in-depth interviews through focus group discussions and direct observation, and continuous process documentation-paying appropriate attention to the empirical complexity of process of change induced by an intervention, are key aspects of the impact evaluation process made to ensure rigor and quality. The application of RCT design however does not mean that there is no room for improvement. For example, we are aware that, it is evidently difficult to have high quality income, expenditure and financial flow data with one-time structured surveys. Recently, data collection tools such as “Financial Diaries” have successfully shown as a best option. We believe that the integration of such data collection tools will best capture the wider impact of ELA programme. Considerable amounts of effort were made to incorporate Financial Diaries as a tool to continuously track cash flows on small samples of ELA RCT households. However, this could not be materialized due to financial constraints.

Case 9. Participatory Assessment of Development

Wouter Rijneveld and Fred Zaal, PADEV

1. The Case

The PADEV project develops an Impact Assessment Method that is participatory and holistic. Impact evaluation is usually linked to specific interventions. Observed impact is often predetermined by those who initiated the intervention. In international development with its inherently unequal power relations this raises the political question of who defines impact. Secondly, it raises the methodological dilemma of choosing between objective, externally defined operational variables for development often far away from beneficiaries or using participatory methods that easily suffer from bias, because of the 'dependency' relation. Thirdly, the focus on a single actor chain easily leads to a tunnel vision where the object of research is insufficiently placed in a complex relation to other local or international interventions and contextual shocks and trends. This diminishes the strategic usefulness of conclusions.

Here we start from the premises that any intervention with a measurable, that is felt impact, will be assessed and valued by the population for which it was intended. They are therefore the true evaluators of the impact of such an intervention. Also, they will be able to assess the impact as compared to other similar or different interventions, and against bigger trends that occur in the area. A particular intervention that a partner organisation is interested in will thus always appear not as a stand alone evaluation of just that intervention, but as an intervention in context.

The project takes place in the Northern Part of Ghana and the Southern part of Burkina Faso. Stakeholders in this project are the donors (ICCO, Woord en Daad, Prisma, and the University of Amsterdam, UvA), the implementing organisations (UvA, Royal Tropical Institute (KIT), University of Development Studies (UDS) in Ghana, Expertise pour le Developpement du Sahel (EDS) in Burkina Faso), and local partner organisations of ICCO, Prisma members and Woord en Daad, but also the local population. In three series of three workshops, an assessment methodology was developed. Each workshop is a three-day exercise where the various components of the toolbox are implemented. These components consist of exercises and small surveys or sets of questions that systematically collect information on the major trends in the area, the changes in particular capitals or capabilities, the personal history of all participants and their families, and an assessment of local definitions of wealth and poverty. Following this, a full list of interventions in the various domains is made, and of these interventions information is collected on the agent that introduced the intervention, the capital it had impact on, and the time it was implemented. After that, the five best and worst projects are selected through negotiated consensus and are assessed on their relation to the major trends in the livelihood domains, their impact on wealth groups, and changes in perception over time. All exercises are done with groups that are split between gender, age and whether the person is an official (linked to government or projects) or civilian.

This gives a very detailed picture of why certain interventions are appreciated, by which social group in society and for what reasons. Analyses can be made with regard to differences between types of projects, timing of projects, actor type, gender differences in perception et cetera. This can be used to compare the basic judgements expressed by the population with those on the particular intervention the partner organisation is interested in (ex post evaluation of impact). It can also be used to develop policies and interventions that assist particular groups in society (ex ante policy development).

2. Values and Quality

There are two sets of values involved. One consists of the values and principles within the team of researchers, basically on what constitutes reliable impact measurement tools, discussed below. The more important set derives from the people who were involved: we take those values as the core of the method. Those values, which differ for men and women, young and old, ethnic and religious groups, and socio-economic status, should come to expression precisely through the methodology developed. Those values are diverse, positive impact is obviously a relative concept.

In such a situation, quality in implementation is also not absolute: we aim for replicability through meticulous reporting, developing the method in as clear and transparent a way, and for complete openness for peer review processes. An External Reference Group is installed to evaluate the method and its implementation, adjusting regularly. We have published our initial results and distributed the tool and the results. Feedback was sought from the participating population on the conclusions. Also, we consider a long time frame, as many results and effects of interventions develop over time and often after the life time of such an intervention. An historical analysis of impact is explicitly part of the method. The methodology aims for inter-subjectivity, not fake objectivity or externally defined objective truth. Whenever there is any form of interpretation of the researchers, or categories given by the researchers, this is carefully documented. This increases validity and replicability without compromising the participatory and holistic approach.

3. Complexity and Robustness

We take the challenge of complexity head-on: this method is radically different as it starts from the complex situation of an area with all its shocks and trends and interventions by a great diversity of actors. The perspective is not from a single intervention chain, so the linear causality bias is absent. Complex society can never be described fully: where detail is won in one aspect, details in other aspects are lost. The core of the method is the position of individuals, representing social groups, in local society. They are the actors defining the relevant aspects in their social and natural environment and the complex and changing relations between these aspects. The risk is that the particular intervention that an agency that commissions the evaluation is interested in, is lost from sight in the myriad inputs people provide.

The method is robust in the sense that it has been developed iteratively over nine workshops in systematically differing areas (density of intervention, historically differing areas). The methodological differences in the iterative steps are reduced progressively, and a convergence of results is the result: comparable quality of data is now achieved under quite differing local conditions, for example whether there was or wasn't a long history of development activity. A manual was developed by the team on the basis of the final toolbox, and this manual seeks to improve the application of the method across teams and countries.

Improvements are being formulated for the fourth and last round of workshops (which in itself will allow an evaluation of the evaluation development process between the first and fourth rounds and the robustness of the findings).

One is that the link with the intervention or the intervening agency needs to improve. We aim to link the agency to the results of the method by introducing the method to agency staff and have them follow the same assessment procedure! This should allow a better understanding of the results at the agency level, and a comparison of the understanding of staff and population of trends, interventions and their impacts. This will be a strong learning experience, and could directly improve the policy of an agency.

Another improvement will be to compare subjectively described trends and occurrences with externally developed datasets. Both a test of the quality of the data provided by the population, and a way to integrate external 'objective' data with local subjective data, this will make the method as a whole even more robust, without losing the subjective basis of the valuation process.

Part of the process is the reduction of the large number of trends mentioned by the various groups of participants into a smaller set that we ask participants to choose from, and link to interventions (causing or mitigating). This pre-selection is a subjective judgement in essence, and there should be a check on the quality of the judgement through a scrutiny of it by the population involved in the exercise.

Lastly, the selection of participants should be carefully done, and this process is yet to be standardised. The solution so far is to repeat, with MSc students projects, the same exercise for a different level of society. This allows an assessment of the robustness of the findings. Examples are the selection of one village among the whole set of villages in one workshop, to see whether similar results are obtained. A second test was to change the type of participant and start from a group of school children and from

there select their households and repeat the exercise. And thirdly, the most excluded and least visible group of all in all instances, the very poor, were purposely selected and the exercise repeated. The results of this comparison are awaited.

Appendix 4 – Summaries from the ‘Methods & More’ Market

- Method 1. Developmental Evaluation
- Method 2. NGO ideas impact box
- Method 3. Theory of Change and complexity
- Method 4: Monitoring capacity and capacity development in Papua New Guinea
- Method 5. Social Return On Investment (SROI)
- Method 6. Collaborative Outcome Reporting Technique (CORT)
- Method 7. Measuring complex systemic changes

Method 1. Developmental Evaluation

Ricardo Wilson-Grau

1. The background/reasoning behind this approach/methodology

Evaluation is to support the development of an innovation. This evaluation mode began to emerge from the practice of Michael Quinn Patton¹⁶ about twenty years ago as he **realised social innovators would benefit from evaluative** thinking that was not offered by traditional evaluation. Social innovators do not think and act in terms of theories of change and logic models. As Patton says, they “don’t follow a linear pathway of change; there are ups and downs, roller coaster rides along cascades of dynamic interactions, unexpected and unanticipated divergences, tipping points and critical mass momentum shifts, and things often get worse before they get better as systems change creates resistance to and pushback against the new.”¹⁷ In this complexity and turbulence they can benefit from support for evidence-base decision-making. Formative and summative evaluation approaches, which focus on improving or judging an intervention model, are not suitable when the intervention model is being created. Social innovators require hard data and hard questions about what they are doing and where they are going as they develop their intervention strategies. That is the role of Developmental Evaluation.

Initially through Patton his own evaluation practice and then more recently with the experiences of other evaluators, often mentored and coached by him, the incipient field of Developmental Evaluation has emerged. To date it has been the subject of a few articles, blogs and in 2006 the book *Getting To Maybe*, co-authored by Patton. His new book *Development Evaluation* was published in 2010 and addresses the topic in-depth. This ‘open market session’ at the May 2010 Conference is based primarily on that manuscript, for which I commented and made small contributions, and secondarily on my own practice in this new field.

2. Description of the methodology/approach

Developmental Evaluation supports the creation of innovative approaches and solutions to major social problems and needs when conditions are changing and the environment is complex and dynamic. (Note that it is distinct from the *evaluation of development*.) A developmental evaluator generates and examines data on actions taken by innovators and their consequences to inform ongoing decision-making. In tracking what happens as an innovator innovates, and communicating their implications, the developmental evaluator calls upon diverse evaluation methods and tools. The key to success, however, is a trusting, respectful working relationship between innovator and developmental evaluator with both tolerating ambiguity and uncertainty and dedicating time together to interpret data from multiple and often conflicting sources, and raise and answer questions about the evolving innovation.

In addition to engaging with Patton as he wrote his manuscript, my practical experience with Developmental Evaluation has revolved around my support for organisations creating planning, monitoring and evaluation systems. This has involved adapting the principles of Outcome Mapping across cultural, social, political and economic divides. These organisations include the Global Water Partnership and its 70-plus national water partnerships in 13 regions around the world; the NGO Free The Slaves in India, Nepal and Western Africa; the Global Platform for the Prevention of Armed Conflict in 15 regions, and an IDRC-sponsored action-research Eco-Health project in Lima Perú. In addition, I am advising the Dutch Doen Foundation’s development of a new strategy for arts and culture grant making in Senegal, Mali, Kenya and Uganda.

¹⁶ Patton is one of the founding fathers of the evaluation profession in the 1970s and to which he has dedicated his life. He is former president of the American Evaluation Association and author of *Utilization-Focused Evaluation*, now in its fourth edition, and *Qualitative Research and Evaluative Methods*, now in its third edition. His 2006 book *Getting To Maybe*, with Frances Westley and Brenda Zimmerman, introduces Developmental Evaluation.

¹⁷ Michael Quinn Patton, *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, Guilford Press, 2010.

3. The advantages and disadvantages in respect to understanding and influencing societal change

Social innovators operate in a disorderly and uncertain world characterised by complex, nonlinear dynamics. In these circumstances Developmental Evaluation can support the process and results of developing an intervention model. It offers systematises information and reflected upon it in real time to inform the ongoing development of the innovation. Thus, the promise of Developmental Evaluation is that it meets this need and supports the development of a viable solution – a model – to meet new, unusual and messy social problems.

Developmental Evaluation is inappropriate, however, when there is already a model to be improved or to be judged for its merit, validity or worth, which is the terrain of formative and summative evaluation. Developmental Evaluation can, however, determine when and if an innovation is ready for formative evaluation as a pilot intervention or for a summative evaluation as a model to be assessed for its potential to be brought up to scale or replicated.

4. References

- Gamble, J.A. (2008). *A Developmental Evaluation Primer*. Montréal: The J.W. McConnell Family Foundation
- Patton, M. Q. (1994). "Developmental evaluation". *Evaluation Practice* 15 (3), 311-20.
- Wehipeihana, N. & McKegg, K. (2009). "Developmental evaluation in an indigenous context: Reflections on the journey to date." American Evaluation Association Conference, Orlando, Florida, November 14.
- Westley, F., B. Zimmerman & M. Q. Patton. (2006). *Getting To Maybe: How the World is Changed*. Toronto: Random House Canada.

Method 2. NGO IDEAS Impact Box

Bernward Causemann

1. Set up

14 German NGOs cooperate with more than 30 Southern NGOs from South Asia, East Africa and the Philippines to develop new ways of assessing impact (in technical terms: outcome, impact and unintended side-effects). The programme runs until 2011 with three regional offices. It also gives technical and conceptual support to VENRO, the umbrella body of German development NGOs, to develop a German civil society response to the aid effectiveness agenda.

The new tools combine three aspects the specific issue of “**the difference we make**”, i.e. the question of how to observe, measure and describe change, combined with the analysis of what has contributed to this change, and to what extent our own action has consequences; an outcome and impact assessment which can be **carried out by the actors themselves**: by the primary stakeholders, their CBO representatives, and the NGO staff who support them; tools and procedures that contribute to awareness creation and empowerment.

The tools are available on www.ngo-ideas.net/publications. They are constantly being revised. Examples and specifications will be published over time.

2. The Impact Toolbox

The Impact Toolbox builds on tools that have been developed in the savings and credit sector in South India in 2006/7. Currently, the toolbox is made more generic to be adaptable in other regions and sectors. It combines four tools.

a. **SAGE** (Goal establishment and assessment by groups for individuals and households).

Members of community groups develop their goals (objectives, aims, aspirations) – usually the same goals for each individual – for development. In intervals (e.g. every year) they assess where they stand. The tool is flexible: rating can be yes/no or on a range of 1-10, or against standard indicators. Goals can be developed by each group, or the same goals are jointly developed by a federation. Crucial is the discussion of the assessment: Members discuss how each individual should be rated. That leads to validation, challenges members and promotes awareness of opportunities for change.

Challenges: a) Often people initially find it difficult to develop goals. b) Field staff often facilitates towards goals that are on an output level, not on an outcome or impact level. They often think in terms of

Expected Benefits from NGO-IDEAS

Communities and groups assess the change that happens to them and around them. Through structured reflection, they

- realise what change happened,
- come to understand what led to this change and how they themselves contributed to it,
- become aware of how they can influence developments,
- come to conclusions what they want to change in the short and medium term.

Southern NGOs interact with communities and groups in assessing change. In addition, they use their own means of assessing change and its causes.

- NGOs get a better understanding of the change that happens.
- They can document change better, and they get valid information when expected change is not happening.
- They understand better the dynamics and priorities in a community.
- Staff are better oriented towards results, rather than just activities.
- NGOs therefore can improve their work.

Northern NGOs get more information on the change they enable, can give more directed support to approaches that bring comprehensive change and can show a specific civil society approach to the aid effectiveness agenda.

indicators, not goals. c) Developing goals for new sectors. Currently, exemplary goals for inclusion of Persons With Disabilities are being developed in East Africa, and goals of street children in the Philippines.

b. **PAG** (Goal establishment and assessment by community groups)

In PAG, community groups develop goals for the groups, their working together and benefits they want to have from the group. Application is as in SAGE. Both SAGE and PAG should be followed by a reflection on what could be done to bring about improvements.

Challenges a) Members can find it difficult in the beginning to think about the benefits they want to have from the group beyond the benefits that are planned for by the supporting NGO. b) Sometimes individual goals that would fall under SAGE are put under PAG.

c. **PIAR** (Participatory Impact Analysis and Reflection)

The NGO-tool in the toolbox is also applicable to reflections of federations, field offices and donors. It describes a systematic approach to analysing the data from SAGE, PAG and PWR, and introduces the need to cross-check the results of this with other existing or generated data (cf. the 6 elements in PIAR). PIAR is currently being revised.

Challenges: a) NGOs often find it difficult to analyse data and to assess what additional information would be relevant. b) SAGE and PAG generate a lot of data that is difficult to handle and analyse. We are testing for this purpose a software that can be used on field office level (www.grafstat.de).

d. **PWR** (Participatory Well-being Ranking)

In order to be able to differentiate impact according to poverty status, PWR is part of the toolbox. It is presented in two forms: As Ranking and as Categorisation. Alternatively, PQL (Participatory Quality of Life Assessment) can be used.

3. The Tiny Tools

Tiny Tools makes the PRA philosophy available for impact assessment, empowering communities to assess change, attribution and impact. They all build on community validation, are mostly visual. They are almost all suited for analysing “the difference we make” with only one application; a repeated application, though, can give a deeper insight to outcomes and impacts. Most work with visualisation. With some Tiny Tools, existing tools have been adapted for impact assessment, others have been developed new. Most Significant Changes light has been adapted in such a way. It is part of the NGO-IDEAs philosophy that the application of the tools should always be concluded with a discussion on what next steps to take.

Challenges: a) Aggregation is more difficult than with SAGE and PAG. b) With some tools, the community needs to understand that this is about change in general, not just the effects of a project.

4. Relevance for the conference

Toolbox and Tiny Tools are meant for a combination of continuous observation with occasional application. Together, they are for a long-term, systematic assessment of outcome and impact. The Tiny Tools can also be used for short-term evaluations to get community perspectives in a more complex and validated way than through interviews (e.g. DEZA 2008 on their water programme: <http://www.seco-cooperation.admin.ch/shop/00010/02022/index.html?lang=en>).

5. Complexity

The NGO-IDEAs tools capture complexity in two ways: a) They build on community perceptions. Communities, when empowered to speak what they experience, look at developments holistically. b) PIAR and some of the Tiny Tools embrace complexity through the way they are designed.

6. Values and principles

NGO-IDEAs puts a high value on community empowerment and on directing development interventions in a way that leads to results that are attractive to the communities. Amongst the three pillars of M&E: to prove, to improve and to empower, NGO-IDEAs puts particular emphasis on stimulating Learning and Action amongst the stakeholders, and particularly the primary stakeholders.

7. Quality standards

NGO-IDEAs tools can meet high quality standards: They can be well documented, can lead to high validity and reliability of data and are mostly replicable. The combination of tools allows a high level of triangulation. Whether these standards are met, depends on application: skills in facilitation, data collection and analysis are required that cannot always be taken for granted in the NGO-sector and among communities that these tools are made for. Empowerment has priority.

Method 3. Theory of change and complexity

Patricia Rogers

This material is drawn from a forthcoming book on program theory and logic models Sue Funnell and Patricia Rogers (2010) Purposeful program theory. Jossey-Bass.

One of the key messages of this book is the importance of taking into account the nature of the intervention when developing, representing and using program theory. Are many organizations involved in making a strategy work? Does a program work the same way for everyone? Will a project only succeed if other services are also available for participants? These different issues all relate to whether it is reasonable to think about an intervention as simple, or whether there are important complicated aspects such as multiple components or complex aspects such as adaptive responsiveness. Taking into account complicated and complex aspects of interventions helps to produce better program theory and better use of program theory.

Aspects of simple, complicated and complex interventions that might need to be addressed when developing, representing and using program theory are listed below.

What it looks like			
	Simple	Complicated	Complex
1. Focus	Single set of objectives	Different objectives valued by different stakeholders	Emergent objectives
		Multiple, competing imperatives	
		Objectives at multiple levels of a system	
2. Governance	Single organization	Specific organizations with formalized requirements	Emergent organizations in flexible ways
3. Consistency	Standardized	Adapted	Adaptive
How it works			
	Simple	Complicated	Complex
4. Necessariness	Only way to achieve the intended impacts	One of several ways to achieve the intended impacts	
5. Sufficiency	Sufficient to produce the intended impacts. Works the same for everyone	Only works in conjunction with other interventions (previously, concurrently, or subsequently)	
		Only works for specific people	
		Only works in favourable implementation environments	
6. Change trajectory	Simple relationship– readily understood	Complicated relationship– needs expertise to understand and predict	Complex relationship (including tipping points)– cannot be predicted but only understood in retrospect

Method 4. Monitoring capacity and capacity development in Papua New Guinea

Heather Baser

1. Background

Many organizations in both developing and developed countries are struggling with how to effectively monitor and evaluate (M&E) capacity and capacity development. One of the challenges is the variety of understandings of the concept of capacity and hence of what the goals of capacity development should be.

Seven years ago, the European Centre for Development Policy Management in Maastricht in the Netherlands set out to better understand what capacity is, how capacity develops, and what outsiders can do to stimulate it. The final report of what became a major study entitled *Capacity, Change and Performance*¹⁸ provides a way of thinking about capacity based on complexity theory. I subsequently used this framework to help three organizations or networks that are part of the law and justice sector in Papua New Guinea assess their progress in developing their capacity. The process allowed those involved to think about questions such as: What makes up capacity? Where does it come from? What does it look like? The final analysis was based on their views of what they had been doing. I played the role of facilitator and scribe.

2. Framework of analysis and methodology

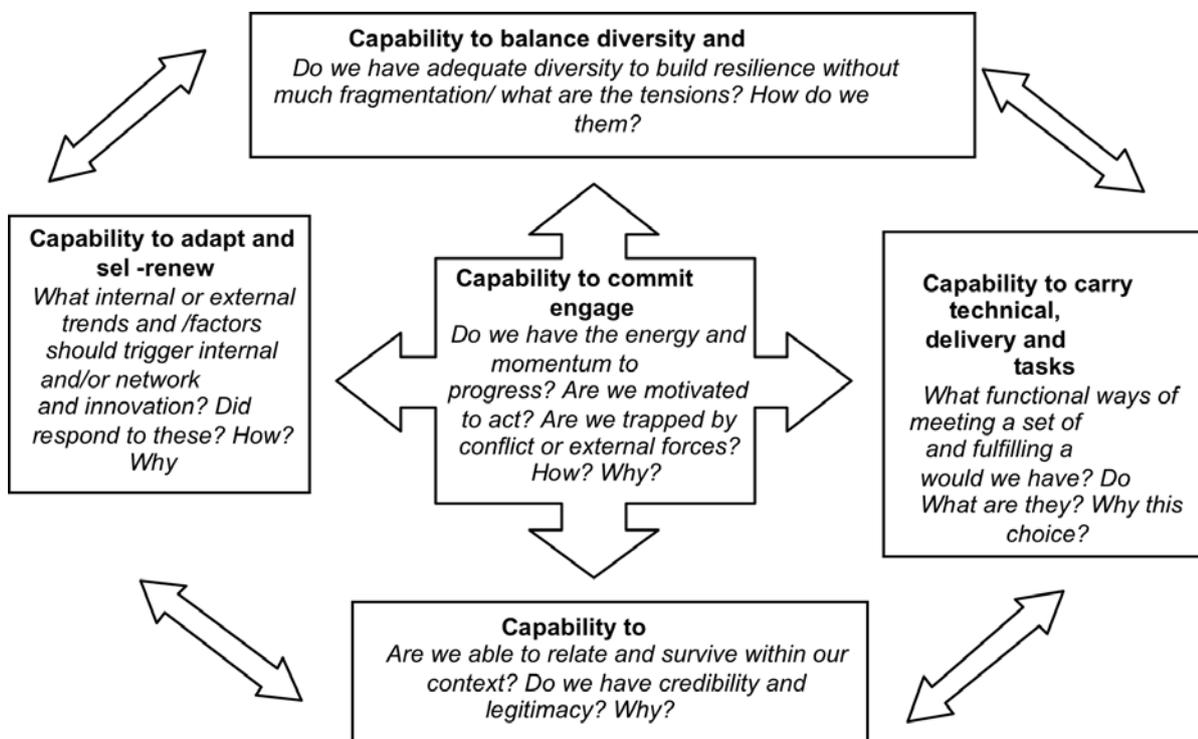
The Capacity, Change and Performance framework sees capacity as emerging from the relationships among the following elements:

- *Context or the political, social and institutional factors within which the organization or system struggles to make its way - Context includes the tangible and intangible features of culture, history and governance.*
- *Individual competencies or the direct contributions of individuals to capacity.* Competencies are the ability to do something and can include a range of skills and abilities from technical or logistical to mindsets, motivations and hopes. The ECDPM research found that the energy of an individual or a small group at the core of a structure contributes to the progress of any capacity development process. Perhaps the most obvious contributions at the individual level were those of leaders at all levels.
- *Collective capabilities or the skills of a group, an organization or a system to do things and to sustain itself - Organizations and systems can be seen as combinations of capabilities and areas where capabilities are weak (their disabilities). The framework sees five core capabilities which interrelate as shown below.*
- *Capacity development is the process of enhancing, improving and unleashing capacity. Capacity development is about how competencies and capabilities interrelate to encourage virtuous cycles that support broader capacity, for example and ideally, improved individual leadership reinforces the group's capability to build supportive networks and attract resources. This, in turn, increases overall capacity. This increased capacity opens up new opportunities for the organization and the individuals within in it.*
- *System capacity* is what emerges from the process of capacity development – the overall ability of a system to perform and make a contribution. It is the outcome of the interrelationships among competencies, capabilities and the context, with the outcome being different from any one of the inputs. This ability includes combining and integrating the competencies and capabilities into a functioning system. Some aspects of such a 'capacitated' system would be legitimacy, relevance, resilience and sustainability.

¹⁸ Baser and Morgan, 2008.

- Performance or the ways in which organizations or systems¹⁹ apply their capabilities and the results of that application in terms of the ability to deliver and function - Performance is about execution and implementation: the results of the application and use of capacity. It is capacity in motion.

The five core capabilities



3. Advantages and disadvantages

I see the main advantages of the methodology as follows:

- It provides clarity for M&E on what is being assessed, by providing nuanced vocabulary to address capacity at different levels – individual, group or organization, and societal.
- It introduces broader questions of legitimacy, relevance and resilience that many M&E systems do not address.
- The non-tangible issues in the framework such as relationship, relevance and motivation have resonance with many managers in PNG (although some middle level staff have more difficulties with them).
- It provides a holistic assessment of capacity rather than a reductionist view.

On the other hand, the challenges of the framework include:

- It is relatively complex and takes time for country partners to understand. It would take even more time for them to use it without support. Distinguishing between collective capabilities and system capacity was particularly challenging for country partners.
- It requires a high level of analysis to understand the interrelationships among the different elements of capacity.

¹⁹ A system is an entity that maintains its existence and functions as a whole through the interrelationships of its parts.

- There is a shortage of proven methodologies for judging aspects of capacity such as relationships, adaptation and legitimacy. They do not lend themselves easily to quantitative measures.

My experience would suggest that this framework would be most useful when applied:

- from the beginning of an activity to allow country partners time to become comfortable with the concepts,
- in medium to large programs (because there is a reasonable investment of time needed to understand the concepts), and
- where country partners have a genuine interest in looking at capacity issues for their own sake, rather than for symbolic purposes.

The framework does need more development, preferably through application.

4. Links to public documents

For general information on the basic analytical framework and the rationale behind it, see *Capacity, Change and Performance*, April 2008 (particularly pages 22-3 and 103-4) and *Capacity, Change and Performance: Insights and Implications for Development Cooperation*, Policy Management Brief no. 21, December 2008. Both are available on www.ecdpm.org/capacitystudy

A summary of the work in PNG will be available shortly through AusAID that financed it as part of the Papua New Guinea-Australia Law and Justice Sector Partnership.

Method 5. Social Return On Investment (SROI)

Jan Brouwers and Ester Prins

1. Background

In 2007, Stichting Het Groene Woudt/The Green Forest Foundation (SHGW), International Child Support (ICS) and Context, international cooperation started to work together on the methodology, Social Return on Investment (SROI). This methodology, already applied in the social sector, is new to the development sector. Cooperation between these organisations includes an exploration of the relevance, positioning and applicability of SROI and is based on a literature review, interviews and field research, developed according to the principles of participatory learning and collaborative action research. These principles place the knowledge, insights and experiences of participants at the centre of the entire process. A number of Southern organisations have taken part in this collaborative venture: amongst others, ICS offices in Asia and Africa; and four India organisations, namely Cecoedecon, Gram Vikas, the Agriculture and Organic Farmers Group (AOFG) and Shri Jagdamba Samiti/Partnership for Clean Indoor Air (SJS).

Based on a write-shop organised in November 2009, in which various development practitioners experimenting with SROI participated, a practical and a visual guide to SROI in the development sector is currently being developed and will be launched in the autumn of 2010. Furthermore, Context, international cooperation is working together with the Social Evaluator, an SROI software supports for organisations interested to applying SROI.

2. What is SROI?

Activities generate economic, social, environmental and, potentially, many other results. An SROI analysis tries to bring these results together and to express them as one value. Such an analysis tries to understand, measure and report on these social, environmental, economic and possible other results, created by organisations or networks. It identifies the sources of value, tries to find adequate indicators of this value, and develops qualitative and quantitative expressions of these indicators. In the end, SROI monetises (some of) these indicators by assigning financial equivalents to social, environmental and other returns. Examples of such returns include: more jobs, increased life expectancy, less absence from work due to sickness, reduced waste, a cleaner environment, lower crime rates, higher tax returns and lower public health expenditures.

The process of an SROI analysis leads to the so-called SROI ratio. This is the ratio between the value of the benefits and the value of the investment. For example, a ratio of 3:1 indicates that for every Euro invested by an organisation, three Euro worth of value (economic, social, environmental and/or other) is delivered to society. When calculating an SROI ratio, it has to be realised that such a figure is, in itself, of limited use. Data about similar organisations are needed to assess whether a SROI ratio of 3:1 is good, to be expected or below average.

The framework of SROI has broadened since it was originally developed. Methodologically, the original focus on monetisation is considered not to be sufficient and/or feasible because not everything can be sensibly monetised. Ideally, the SROI methodology should combine quantitative and qualitative measurements: providing a quantitative investment ratio, in combination with a narrative, based on qualitative methods, such as storytelling exercises. In recent versions of SROI, more emphasis has been put on ascribing value to unintended consequences and negative impacts next to intended and positive ones. A typical SROI analysis thus consists of (i) a specific process by which the SROI ratio is calculated, (ii) contextual information in quantitative, qualitative and narrative terms to enable accurate interpretation of that ratio, and (iii) additional information on values which could not be monetised during the analysis. SROI bases its understanding of value on the perceptions of stakeholders. To take a simple example, the monetary value of a pair of fashionable, brand-name shoes will be very different for a child who wants to show off to friends than for his or her parents who have to pay for the shoes. In the development sector, stakeholders have to be regarded as the owners of the intervention. They are the actors in their own development process and they are the ones who should define the intended changes and results to be

achieved. As such, SROI can be employed to make a deliberate choice to analyse results from the perspective of the beneficiaries.

Thus, SROI can provide a good basis for improvement of an organisation's performance because the point of view of the organisation's programme can itself be placed in the centre. In this way, the results of development cooperation can be measured 'developmentally', which implies measuring in a participatory way, self-respect and contributing to the development objectives of local organisations. SROI also has the potential of providing learning opportunities.

3. The SROI steps

Step 1: Defining the boundaries (objectives and scoping)

As a first step, one needs to clarify what is going to be measured and why a measurement process is being started. The scoping phase helps to define the boundaries of the analysis and this is critical to making the SROI analysis practicable. In the scoping phase, explicit boundaries of what is to be included or excluded from measurement are defined.

Step 2: Identification and selection of key stakeholders

Once one is clear about the scope of SROI, the next step is to identify and involve stakeholders.

Step 3: Theory of change

This is one of the most important steps within the SROI framework. It tells the story of how stakeholders involved in the project believe their lives have changed or will be changed.

Step 4: What goes in?

In order to make a project possible, inputs need to be provided. In this step you will identify what has been contributed by all stakeholders involved in order to make the project possible.

Step 5: What comes out? (Identifying results)

This step visualises the results of the project. These can be on both outcome or impact level, and both positive and negative. In this step, stakeholders formulate indicators in order to express a simple and reliable manner to measure achievement or to reflect the changes connected to an intervention.

Step 6: Valuation

This step entails assigning a value (often expressed in monetary terms) to the social and/or environmental outcomes or impact of an intervention.

Step 7: Calculation of the SROI ratio

By calculating the SROI ratio, a comparison can be made of the investments (inputs) on the one hand and resultant social and environmental returns (outcomes and impact of an intervention) on the other. Calculate the rate of return: all (adjusted) valued results/invested inputs. When the SROI ratio is higher than 1, there is a net positive result.

Step 8: Verification

Validate the data obtained.

Step 9: Narrative

An SROI ratio, as such, provides interesting information but it should be embedded within the larger context in order to fully understand its meaning. The narrative provides information on the process and allows for describing elements that could not be measured or could not be valued.

4. More information

For more information on SROI, please visit the following websites:

<http://sroiseminar2009.wordpress.com/>

www.developmenttraining.org (Contextuals No. 4, 8 and 9)

<http://socialevaluator.eu>

Box 15: SROI ratio calculation of the Tanzania Grain Storage Scheme

The Grain Storage Scheme (GSS) in Mwabusalu started in 2004. The region used to have food-shortages, due to an uneven distribution of grain-availability over the year. The introduction of the GSS was meant to store grain to use it in economically difficult times. In 2007 the project was handed over to the community. A monitoring committee is now in charge.

Calculation of the ratio

The community of Mwabusalu, facilitated by its GSS monitoring committee, identified the following changes in their lives (impact) as a result of the project intervention:

- Improvement in the quality (and quantity) of the stored grains;
- An increased food availability, especially in the 'hunger period';
- Increased pupils' attendance at school by reducing shortage of food.

The community calculated as break-even point for the project if 1,250 bags per year would be stored against a price of Tanzanian Shilling (TSh) of 6,000 per bag. This means that the total input can be equalised with $1250 \text{ (bags)} \times \text{TSh } 6,000 = \text{TSh } 7,500,000$ per year.

They decided to calculate the Economic Return On Investment ratio first and to add, in a second process, the social ratio: Research revealed a season high and a season low price of maize in 2006 of TSh 39,000 and TSh 22,800 respectively. In case farmers could store their maize until the 'hunger period' with the seasonal high price, the total undertaking could result in $1,250 \times$ the difference between the two prices, which is 16,200. In total this will amount to $1,250 \times 16,200 = \text{TSh } 20,250,000$. The economic ROI, which is calculated by dividing the output by the input, is 2.7.

As indicator for the improvement in the quality of the stored grains the group came with the germination capacity of the maize. Before the project started the germination of maize was low, because of a weevil. It eats the most nutritious part of the maize, which is also responsible for seed germination. To kill the weevils the maize stored is fumigated regularly. Stored maize thus has a higher germination-rate and a significantly higher yield. When used for sowing, only 2 kgs per acre are now needed, compared to 12 kgs before, when stored at home. With a price of TSh 500 per kg of seed this means a reduction per acre of TSh 5,000. Calculating with the average of 7 acres per family and 67 families in the scheme this comes to a total of $7 \times 67 \times 5,000 = \text{TSh } 2,345,000$.

Increased food availability is defined as additional yield per acre, which, as a result, will provide the year around availability of food. As shown above, the quality of the fumigated grain is much higher than the one kept at home. This, according to the farmers, leads to a 80% production rise, from, on average, 5.5 bags per acre if not fumigated, to 10 in the latter case. Using the 2006 high season prices, this leads to $7 \text{ (acres)} \times 4.5 \text{ (extra bags)} \times 67 \text{ (participants)} \times \text{TSh } 39,000 = \text{TSh } 34,190,100$ additional gain.

Increased school attendance due to a reduction of food shortage is measured by cost reduction for a family in search for medical treatment. The average treatment costs were estimated at TSh 1,000 per doctors visit for consult and medication, added with costs a parent is making by not being able to work of TSh 2,000 per occasion. The group found the number of visits to be cut by half, decreasing from 12 to 6 per year. This would mean a gain of $6 \times 3,000$ per child per year. With an average of two ill children per family this leads to an annual cost reduction of $2 \times 18,000 = \text{TSh } 36,000$. For the participating community members this leads to an annual saving of $67 \times 36,000 = \text{TSh } 2,412,000$.

The SROI ratio, which can be calculated by adding these three figures together and dividing this by the input, comes to 5.2. Combining this ratio with the economic ratio of 2.7, this comes to a total ratio of 7.9. This ratio will grow if, as expected, more members will participate in the scheme.

Method 6. Collaborative Outcome Reporting Technique (CORT)

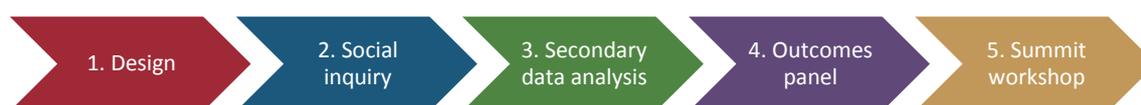
Jess Dart

The Collaborative Outcome Reporting Technique (CORT) was developed by Dr Jess Dart and is characterized by two elements: a participatory process whereby the information is generated and a five-part structure in which the report product is presented.

Report structure: the report aims to explore and report the extent to which a program has *contributed* to outcomes. Under the CORT, reports are short and generally structured against the following sections:

1. A narrative section explaining the **program context** and rationale.
2. A **'results chart'** summarising the achievements of a program against a theory of change model.
3. A narrative section describing the **implications** of the results e.g. the achievements (expected and unexpected), the issues and the recommendations.
4. A section that provides a number of 'vignettes' that provide **instances of significant change**, usually first person narratives.
5. An index providing more detail on the sources of **evidence**.

Participatory process: CORT uses a mixed method approach that involves participation of key stakeholders in a number of process steps:



Participation can occur at all stages of this process for example:

1. **In the planning workshop.** In this workshop the theory of change is clarified, existing data is identified and evaluation questions developed. Consultants play the role of facilitation and documentation.
2. **In the data trawl.** Program staff may be enlisted to help with the collation of data, although in our experience consultants usually lead this process as the evaluation managers.
3. **The social inquiry process.** Volunteers can be given a short training session in interviewing and with the help of an interview guide can conduct the interviews. This is a very effective way to involve staff in the data where there is sufficient enthusiasm around the process. Otherwise consultants or the evaluation managers conduct all or a proportion of the interviews.
4. **Outcomes panel.** People with relevant scientific, technical, or sectoral knowledge are brought together and presented with a range of evidence compiled by the consultants. They are then asked to assess the contribution of the intervention towards goals given the available knowledge. We call this an outcomes panel and it is usually facilitated. It is sometimes also referred to as an expert panel. It can be substituted for a citizen's jury.
5. **Summit workshop.** At a large workshop instances of significant change are selected (incorporating aspects of Most Significant Change Technique) and key findings and recommendations are synthesised. The summit should involve broad participation of key stakeholders such as program staff and community members.

Advantages: Organisations often place a high value on the reports because they strike a good balance between depth of information and brevity and are easy for staff and stakeholders to understand. They help build a credible case that a contribution has been made. The participatory process by which reports are developed offers many opportunities for staff and stakeholder capacity building. Compared to standard outcomes evaluations approaches they are relatively straightforward. They are a great way to kick off a new monitoring and evaluation system, because they involve synthesising and reflecting on all existing data and data gaps (a great platform to think about what data is really needed!). It has been used in a wide range of sectors from overseas development, community health, and Indigenous education. But the

majority of work has occurred in the Natural Resource Management Sector, with the Australian Government funding 20 pilot studies in 2007-9.

Limitations: PSR's have been criticised for being too appreciative, or for being incapable of telling a bad story. While this is certainly a risk, the technique does attempt to address this in a number of ways. Firstly all informants are asked to describe the strengths and the weaknesses of the program. These weaknesses or issues are documented in the report. Secondly, the outcomes panel is encouraged to report on negative as well as positive trends in terms of the outcomes. So the "negatives" are not avoided in this process. However, the choice of topic for an outcomes report is often purposeful rather than randomly selected. Topics for reports are often selected on the basis that they are likely to show some significant outcomes. In addition CORT only address one type of key evaluation question. That is the question concerning the extent to which an investment contributes to outcomes. It is an extremely important question, but it is not the *only* type of key evaluation question that is important. This needs to be understood and acknowledged. For this reason, CORT should not be seen as the only reporting tool. The idea is that it should complement other reporting processes or be extended to encompass more.

Values: CORT is based on the premise that the values of stakeholders, program staff and key stakeholders are of highest importance in an evaluation. The evaluators attempt to "bracket off" their opinions and instead present a series of data summaries to panel and summit participants for them to analyse and interpret. Values are surfaced and debated throughout the process. Participants debate the value and significance of data sources and come to agreement on the key findings of the evaluation.

Quality: The focus on quality is largely associated with *process quality*: ethical conduct; culturally appropriate methods; ownership of the evaluation process; ensuring that the evaluation provides credible but useful recommendations to inform the next phase of the program. Interviews are usually taped and transcribed. Data is double analysed by participants at workshops and by the consultants using thematic coding.

Complexity: CORT is especially useful when a program has emergent or complex outcomes that are not fully defined at the onset of a program. For this reason a theory of change is refreshed at the start of the evaluation process. In addition qualitative inquiry is used to capture unexpected outcomes and deliberative process are used to make sense of the findings.

Resources: Clear Horizon runs a two-day training program on this technique. We have also drafted a comprehensive User Guide that lays out all steps of the process. See www.clearhorizon.com.au. Examples report can also be found here. Jess hopes to write the book soon!

Method 7. Measuring complex systemic changes

Adinda van van Hemelrijck

In the past two decades a debate has been going on about the effectiveness of aid and development, how to measure its impacts and make evidence-based arguments about what works and what doesn't. The debate has culminated in the old war of methods, between logical positivism and interpretative relativism, the "scientific" way of collecting "hard evidence" versus the qualitative and more participatory approach producing "soft(er) evidence". While recognizing the depth and importance of the methodological dispute, I find it more productive to try to move beyond the dispute and make the best use of all worldviews in an integrated, flexible and responsive manner. At Oxfam America, we have used this proposition to develop a rights-oriented approach to planning, evaluating and learning, based on the understanding that fighting poverty and injustice requires fundamental systemic changes at multiple levels, and consequently a methodological fusion that can capture complexity and present it in a way that can meet and influence stakeholders' different world views.

This introduction paper gives a brief overview of the basic premises of Oxfam America's approach to impact measurement and learning from a right perspective, a short description of the case on productive water rights in Ethiopia that shows this approach, and the main challenges we face not just in this particular case but in all programs. A selection of background literature is added that has influenced the thinking behind this approach.

Oxfam's approach

a. Fighting the root causes, not just the symptoms

Local realities are embedded in wider systems that influence and shape them while also the local systems influence its surrounding environment. The root causes of poverty and injustice are multi-dimensional, varying across different contexts but entrenched in wider and more complex interdependencies. Poverty and injustice can be described essentially as rights issues that are complicated by the multi-level nature of rights violations in socio-political relationships, institutions and "glocal" markets. Hence, it cannot be fixed by short-term interventions, neither by the "scale-up" of such quick fixes. Its symptoms can be fought temporarily (as famine is by food aid, lack of water by digging wells, lack of cash by savings & credit, etc.). Its root causes, though, require fundamental systemic changes of the individual, collective, societal and institutional competencies and behaviours that are reinforcing and reproducing exclusion, discrimination and deprivation at various levels. Breaking somewhat with conventional definitions, Oxfam America measures "impact" therefore as

a significant and sustainable change in power relations that enables excluded and marginalized people to realize their rights to access and manage the resources, services and knowledge they need for strengthening their livelihoods, improving their well-being, and influencing and holding accountable the institutions that affect their lives.²⁰

Development is shaped and done by people –not for people. In order for people to be able to influence and change individual, collective and institutional behaviours, they need to understand how the underlying system works. Development can therefore be understood as freedom or empowerment: the ability of people to influence the wider system and take control of their lives. This implies that development efforts – and thus its planning, evaluation and learning processes – should focus on building both people's capabilities to understand and work the system (agency) and the enablers that help them doing so (the institutions and complex webs of relationships).²¹

b. Measuring complex systemic change over time

²⁰ From LEAD (2008).

²¹ From Van Hemelrijck (2009).

Obviously no organization can do this on its own. Impact, as defined above, can only be realized through collaborative efforts over long periods of time around specific rights issues in a specific context. So Oxfam America develops, together with its partners, 10-15 years programs consisting of both project and non-project activities²² that are strategically aligned and, based on a defensible theory of change, geared towards achieving a common impact goal. Clearly, partners and stakeholders in these programs cannot be motivated to contribute consistently over a longer term if they cannot observe and understand how a program's impact fundamentally changes the system.

Hence the importance of a robust impact measurement and learning approach that

- a) can reveal complex (non-linear) causal relationships between changes at different levels and at different moments in times;
- b) is simple and cost-effective enough to last for many years;
- c) can be debated and understood by partners and key stakeholders, particularly the poor people themselves; and
- d) can help build the case for plausible contributions to fighting the root causes rather than try to attribute such changes to any single actor, or any single project, or any single intervention.

A program's theory of change visualizes the complex systems changes we try to realize and measure, and reveals the set of assumptions about the most effective ways to get to impact. By pitching indicators on its most crucial outcomes and hubs we can measure the complex interactions and change patterns. The theory of change and indicators do not have to be perfect and worked out in great detail, but "good enough" to enable partners and stakeholders to understand the system and learn about the patterns of change. More sophistication is obtained through the design of the methods and tools that are required for ongoing monitoring of project and non-project contributions, iterative research on important causalities, and longitudinal evaluation of impacts and change patterns. Combining ongoing outcome monitoring and iterative research should help probing and sculpturing a program's change theory over time, by: (a) filling critical gaps, (b) bridging the time lags, (c) probing the assumptions, and (d) keeping track of intervening or "unexpected" variables in the theory of change. Good "benchmarking" of the change theory in manageable phases of three to four years, should enable us understand distant relationships, and plan different interventions accordingly.

The right choice of methods, then, depends on what questions about what particular parts of the system are investigated at what point in time, at what scale or level of complexity, to convince or influence whom, for what purposes. Individual methods become rigorous in as much as they comprehensively and consistently can generate valid and reliable data that speaks to the indicators and questions in the program's theory of change. This requires setting boundaries, and at the same time recognizing the politics and fuzziness of boundaries.

c. Dealing with the politics and fuzziness of boundaries

Program outcomes cannot be studied in a totally "objective" manner by a non-engaged external observer, because they cannot be isolated from the wider socio-economic and political environment and its more localized interdependent variables. Therefore, researchers cannot stay out of the system they are observing –neither the localities where they conduct the field studies, nor the wider system of which their institutions and contractors are part. Once they start to collect data through observations, interviews, surveys, diaries, and focus groups, and process and qualify data to draw conclusions, they are actually creating and attributing value and meaning, thus interacting with the embedded power structure.

The purpose of involving external researchers for evaluating outcomes and impacts of Oxfam America's programs, therefore, is not so much to obtain "objective" evidence of the effectiveness of its interventions within set boundaries. Rather, it is to provide a fresh perspective on the observed state of play, which can challenge the boundaries of what is accepted as "truth" or common sense, identify false assumptions and

²² E.g. global-to-national advocacy, movement & constituency building, community mobilization, local-to-global market inclusion, private sector engagement, primary research, etc.

deficits in the knowledge system, and reveal wider patterns of behaviour that create and perpetuate the root problems. Hence the importance of building in feedback loops into evaluations, and between internal and external monitoring and evaluation. For this, collected evidence needs to be carefully triangulated with other data sources, and validated by its multiple stakeholders, around the program's theory of change. Data sources can include verified stories from people, monitoring reports from partners, statistical records, baseline & evaluation reports, and other research studies. The different data streams from monitoring, research and evaluation converge in an annual impact reflection, in which partners and stakeholders collectively try to discover the patterns of change in the system.

For creating consistency in the research methodology, and coherence in the evidence collected from different sources over a program's lifetime, we seek to establish a long-term collaboration with a locally or regionally based research institution. Our assumption is that this also will contribute, on its turn, to building a country's capacity for dealing with complex systemic change through a process of institutionalization of the knowledge acquired in the program.

d. An emblematic case

The case presented here is the impact measurement and learning system of a 10-year, rights-based program around smallholders' productive water rights in Ethiopia, that is aiming at enabling smallholders to access and manage water resources (through strengthening their organization) in order to enhance their productivity and food security in a fair, equitable and sustainable manner. The program is designed around a theory of change that builds on the core-proposition of "co-investment" by smallholder communities, NGOs/CSOs and local/regional/federal government, supported by the necessary institutional, legal and regulatory changes at all levels. The program has a strong gender component, with specific impact indicators measuring equitable access and control over strategic resources within the household, the communities and their organizations. Women are expected to gain decision-making power and take greater leadership in community institutions as well as in local NGOs/CSOs and government offices at local, regional and federal levels.

A program impact baseline research is in the process of being finalized. The research focused on a core set or "system" of impact and outcome indicators related to the core proposition in the change theory. Secondary literature review and comparative case studies within case studies were conducted (mainly process tracing of "in" and "out" household panels, combined with key informant interviews and "in" and "out" focus groups, village transects and participatory diagramming, secondary, and statistical data analysis at federal and regional state levels, and policy and organizational chart analyses). Research findings will be validated by key stakeholders in a workshop in Addis Ababa on June 7th, 2010. After three to four years, a formal step-back will be taken and an impact evaluation conducted on the same indicators, with the same groups of people and the same household panels, while the baseline will be expanded through primary research on additional impact indicators and outcomes relevant to the next program phase.

Ongoing iterative research will be carried out every one to two years (sequenced over time) for assessing crucial causal mechanisms and specific questions that require special expertise (e.g. water rights codification & distribution), sophisticated methods (which could be an experimental design), and additional funding (for instance, for piloting a productive sustainability index). The impact research and evaluation agenda is developed and implemented by our research partner, the IWMI (International Water Management Institute). The primary research is conducted by local researchers speaking the local languages and knowing the local contexts, who are supported by a small advisory team of high level and progressive IWMI researchers (including a gender and water rights expert, an economist, an agronomist, and an impact assessment specialist).

The overall objective of the impact research and evaluation agenda is to provide partners and key stakeholders with

- (a) accurate, valid, and useful data on the most crucial impact indicators, which help them to assess to what extent significant changes are taking place and discover patterns of change; and

- (b) narrative analyses of the causal processes that are contributing to these changes to assess whether the program's theory of change is actually working (or not) and probe its major assumptions.

Impact research and evaluation is complemented by an annual impact monitoring, reflection and reporting cycle that is managed by a group of strategic partners. On an annual basis, they will convene key stakeholders to make sense collectively of the different data streams, and advise the strategic partners to make the necessary course adjustments. The collective sense-making processes are essential to success: they form the basis of a downward accountability and reflexive learning practice we want to develop. The annual collective impact reflections will be combined with (and build on) empowering methods for ongoing impact monitoring (such as most significant change, action research on traditional conflict mediation, participatory value chain analysis, and farmer constituency feedback committees).

e. Challenges

The long-term impact measurement and learning system presented here is emblematic for nine other programs under development at Oxfam America. We are building the planes very much as we are flying them and as a result these systems haven't been completed, tested and revised yet. Most have gone through a long participatory design process and are now at the stage of having established a programmatic impact baseline covering the impacts and most crucial outcomes in the change theory. "Soft" evidence through outcome and impact monitoring will be obtained this year; "harder" evidence through impact evaluation, two to three years from now. An agency-wide peer review²³ of the processes and products that will deliver the soft evidence on these programs will be carried out next year. An external evaluation of all of Oxfam America's ten impact measurement and learning systems is planned for 2013-2014, when the first streams of hard evidence has come out of the program impact evaluations.

Among the challenges we face in particular in this Ethiopian case, are whether there will be enough commitment from key players to achieve significant systematic changes that are hard to measure, and whether sufficient funding can be secured for elements of the program with indirect and muddled returns on investment. Despite the widespread acknowledgment of the need for longer term approaches in development, and the need to focus on root causes, few donors, foundations, or social investors are willing to invest in such a complex methodology and measurement system. Finally, it is also a challenge to develop the competencies that are needed to manage these multi-level, multi-dimensional and multi-actor measurement and learning processes, understand the methodology and think outside the traditional development box. For most development workers and managers on the ground, in their daily routine, they prefer approaches that rather simplify the managerial requirements and challenges. Being committed to achieving predetermined targets, they are tempted to adhere to an approach that tries to prove in a relative short time frame the unambiguous success of "golden bullet" solutions for then replicating them at larger scale. Managers in general don't like insecurity, uncertainty and fuzzy boundaries. Being confronted with complexity and uncertainty is risky and scary. In a context like in Ethiopia where people tend to be more risk-averse and stick with tradition, this can be particularly challenging. Although Oxfam's approach is confining some of these uncertainties through using a theory of change, dealing with it in a systematic manner and building the interventions from a theory of change obviously remains somewhat counter-intuitive.

Methodologically we are confronted with two major questions, which are:

- How to establish a rigorous enough relationship between indicators and related phenomena that are actually quite distant from one another in time or space? How to qualitatively keep control of intervening variables over many years, most of which we can't anticipate?
- What are then appropriate methods for measuring both qualitatively and quantitatively how groups/clusters of indicators move together (or not), how "leverage points" do (or do not) create wider ripple effects of influence?

²³ We have developed a participatory methodology called APPLE that is used for agency-wide horizontal review and planning exercises at meta or strategic levels.

f. Conclusion

Oxfam America seeks to obtain robust evidence on complex changes in conditions, behaviours, relationships and institutions required to empower poor people for asserting their rights and transforming their lives, through a systems approach to measurement and learning. With the adequate leadership and sufficient commitment from key stakeholders, such complex changes can be realized and measured if the right combination of methods is deployed. The optimal match helps partners and stakeholders to probe why and when certain changes occur, understand the system behind it, and sculpture the program's theory of change. Cognizant of the fact that these are "living" systems created through human interactions and shaped by power relations, it is extremely difficult, though, to find an "objective" way of proving causal connections in particular when changes happen over long distances in space and time and multiple (known and unknown) variables are into play. Evidence that is robust enough to serve its purpose of revealing and influencing complex systems change, I believe, can be obtained therefore only through applying:

- a) an appropriate mix of methods in impact research and evaluation for the purpose of triangulation;
- b) appropriate collective analysis & sense-making methods for the purpose of validation; and
- c) appropriate methods for obtaining data on the individual grant and non-grant activities of the program, which implies verification.

Groundbreaking in Oxfam America's impact measurement & learning systems is

1. The use of a system of indicators to measure "impact" which is defined in terms of "empowerment" and measured consistently and iteratively over a longer period of time (10 years);
2. The use of a change modelling approach by visualizing a program's theory of change and mapping out its system of indicators, which helps partners and stakeholders get their heads around the system and make sense of complex data;
3. A serious attempt to sidestep the dichotomy of "objectivism" versus "subjectivism" through a systems approach that brings different data streams together -from ongoing internal tracking by partners and external impact research and evaluation by an external research partner- in a curriculum of impact reflection and meta-learning with key stakeholders; and
4. The attempt to contribute to the institutionalization of the knowledge acquired over time within the country itself, through working with regionally or locally based research institutions (instead of consultants).

Appendix 5 – References and Further Reading

- Barlett, A. (2008), *No more adoption rates! Looking for empowerment in agricultural development programmes*. In: Development In Practice (Vol 18, Nrs 4-5: 524-538). Oxfam GB. Routledge Publishing.
- Byrne, A. (2009), *Working Towards Evidence-Based Process: Evaluation That Matters*. MAZI Articles. (<http://www.communicationforsocialchange.org/mazi-articles.php?id=356>). South Orange NJ: Communication For Social Change Consortium.
- Bourdieu, P. (1995; 10th Ed), *Outline of a Theory of Practice*. Cambridge University Press.
- Chambers, R. (July 2007), *From PRA to PLA and Pluralism: Practice and Theory*. IDS
- Chambers, R.; Karlan, D.; Ravallion, M. & P. Rogrs (July 2009), *Designing Impact Evaluations: Different Perspectives*. New Delhi: International Initiative for Impact Evaluation (3ie).
- Creswell, J. W. & V.L. Plano Clark (2007), *Designing and Conducting Mixed Methods Research*. London, New Delhi: SAGE Publications.
- Cummings, S. (2010), *Evaluation revisited I: rigorous vs. vigorous*. <http://thegiraffe.wordpress.com/2010/06/17/evaluation-revisited-i-rigorous-vs-vigorous/>. Post conference blog, June 17 2010.
- Davies, R. & J. Dart (Apr 2005), *The Most Significant Change (MSC) Technique. A Guide to Its Use*. Version 1.00. Funded by Oxfam International *et al*.
- Deaton, A. (2010), Instruments, Randomization and Learning about Development. *Journal of Economic Literature* 48 (June 2010): 424 – 455.
- Earl, S. & F. Carden (2003), *Learning from Complexity: The International Development Research Centre's Experience with Outcome Mapping*. In: Development and the Learning Organization. Oxfam GB.
- Easterly, W. (Dec 2010), Sometimes it IS about the money. <http://aidwatchers.com/2010/12/sometimes-it-is-about-the-money/>. Posted December 20, 2010
- Eyben, R. (May 2008), *Power, Mutual Accountability and Responsibility in the Practice of International Aid: A Relational Approach*. IDS Working Paper 305.
- Gamble, J. (2008), *A Developmental Evaluation Primer*. Canada: The J.W. McConnell Family Foundation (on the Sustaining Social Innovation Initiative).
- Geertz, C. (1993), *The Interpretation of Culture*. London: Fontana Press.
- George, A.L. & A. Bennett (2005), *Case Studies and Theory Development in the Social Sciences*. Cambridge (USA): MIT Press.
- Guijt, I. (Nov 2008), *Critical Readings on Assessing and Learning for Social Change: A Review*. Sussex (UK): IDS.
- Guijt, I.(2008), *Seeking Surprise. Rethinking Monitoring for Collective Learning in Rural Resource Management*. Wageningen University Press, The Netherlands.
- Ebrahim, A. and V.K. Rangan. (2010), The Limits of Nonprofit Impact: A Contingency Framework for Measuring Social Performance. Harvard Business School. Working Paper 10-099
- Kabeer, N. (Dec 2000), *Resources, Agency, Achievement: Reflections on the Measurement of Women's Empowerment*. In: *Power, Resources and Culture in a Gender Perspective: Towards a Dialogue Between Gender Research and Development Practice*. Conference Report. Uppsala University: Collegium for Development Studies, in cooperation with SIDA.
- Keystone Accountability. *Developing a Theory of Change. A Framework for Accountability and Learning for Social Change*. A Keystone Guide.
- Khagram, S. and C.W. Thomas (2010), Towards a Platinum Standard for Evidence-Based Assessment by 2020". *Public Administration Review*. Special Issue on the Future of Public Administration in 2020. Volume 70, Issue Supplement s1, pages S100–S106, December 2010
- Krznicar, R. (2007), *How Change Happens. Interdisciplinary Perspectives for Human Development*. Oxfam GB Research Report.

- Kurtz, C. and D. Snowden (2003), *The new dynamics of strategy: Sense-making in a complex and complicated world*. IBM SYSTEMS JOURNAL, VOL 42, NO 3, 2003
- Lacayo, Virginia (2007), *What Complexity Science Teaches Us About Social Change*. MAZI Articles (<http://www.communicationforsocialchange.org/mazi-articles.php?id=333>). South Orange NJ: Communication For Social Change Consortium.
- LEAD (Dec 2008), *Rights-Oriented Programming for Effectiveness. Designing and Evaluating Long-Term Programs*. Boston: Oxfam America. (Available upon request)
- Lehrer, J (2007), *Annals of Science. The Truth Wears Off. Is there something wrong with the scientific method?* December 13 2010.
http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer#ixzz1AdUHGvpG
- Mayoux, L. & R. Chambers (2005), *Reversing the Paradigm: Quantification, Participatory Methods and Pro-Poor Impact Assessment*. In: *Journal of International Development* (17: 271-298).
- Mayoux, L. (2007), *Evaluation and Impact Research for Rights-Based Development. Issues and Challenges*. Paper presented to Oxfam America Impact Evaluation Workshop, September 17-21, 2007, Lima, Peru.
- Mertens, D. M. (2009), *Transformative Research and Evaluation*. NY: The Guilford Press.
- Morell, J.A. (2010) *Evaluation in the Face of Uncertainty: Anticipating Surprise and Responding to the Inevitable* Guilford Press N.Y.
- Narayan, D. (Ed; 2005), *Measuring Empowerment: Cross-Disciplinary Perspectives*. Washington DC: The World Bank.
- Natsios, A. (2010), *The Clash of the Counter-bureaucracy and Development. Center for Global Development Essay*. July 2010.
- Nyamu-Musembi, C. & A. Cornwall (Nov 2004), *What is the "rights-based approach" all about? Perspectives from international development agencies*. IDS Working Paper 234. Brighton: Institute of Development Studies.
- Nitipaisakul, W. (March 2007), *Theories of Change: A 'Tipping Point' for development Impact?'*. In: *Impact* (Issue 3, March 2007), Barton (Australia): Hassal & Associates International.
- Offenheiser, R.C. & S. Holcombe (2001), *Challenges and opportunities of implementing a rights-based approach to development: an Oxfam America perspective*. Paper presented at the Conference on Northern Relief and Development NGOs: New Directions in Poverty Alleviation and Global Leadership Transitions (July 2–4). Oxford: Balliol College.
- Parks, W. (2005). *Who Measures Change? An Introduction to Participatory Monitoring and Evaluation of Communication for Social Change*. South Orange NJ: Communication For Social Change Consortium.
- Patel, S. (2007), *Reflections on Innovation, Assessment and Social Change Processes: A SPARC Case Study, India*. Discussion Paper. IDS, United Kingdom.
- Patton, M.Q. (2002, 3rd ed), *Qualitative Research & Evaluation Methods*. London, New Delhi: SAGE Publications.
- Patton, M.Q. (2008, 4th ed), *Utilization-Focused Evaluation*. London, LA, Singapore, New Delhi: SAGE Publications.
- Perrin, B. (2010). *What is Rigour in "Rigorous" Impact Evaluation?* Presentation on Rigour to the European Evaluation Society, Prague Oct 8, 2010.
- Ramalingam, B. (2011). *The globalisation of vulnerability*. January 11, 2011.
<http://aidontheedge.info/2011/01/11/the-globalisation-of-vulnerability/>
- Ramalingam, B., Jones, H., Reba, T., & Young, J. 2008. *Exploring the Science of Complexity: Ideas and Implications for Development and Humanitarian Efforts*. London: Overseas Development Institute.
<http://www.odi.org.uk/resources/download/583.pdf>
- Reeler, D. , *A Theory of Social Change and Implications for Practice, Planning, Monitoring and Evaluation*. Community development Resource Centre. Capetown: <http://www.cdca.org.za/>.
- Reyes, C. & E. Due (Nov 2009), *Fighting Poverty with Facts. Community-Based Monitoring Systems*. IDRC

- Rossi, P.H.; Lipsey, M.W. & H. E. Freeman (7th Ed; 2004), *Evaluation. A Systematic Approach*. Thousand Oaks, London, New Delhi: SAGE Publications.
- Schwandt, T. A. (1981), *Defining Rigor and Relevance in Vocational Education Evaluation*. Paper presented at the Annual Meeting of the American Educational Research Association (Los Angeles, CA, April 13-17, 1981).
- Sen, A. (2001), *Development as Freedom*. Oxford, New York: Oxford University Press.
- Senge, P. (1990), *The Fifth Discipline. The Arts & Practice of the Learning Organization*. NY: Doubleday/Currency.
- United Nations (2003), *The Human Rights Based Approach to Development Cooperation: Towards a Common Understanding Among the UN Agencies*. UNDP.
- Tacchi, J.; Slater, D. & G. Hearn (2003), *Ethnographic Action Research: A User's Handbook*. New Delhi: UNESCO. (<http://mobilecommunitydesign.com/2005/10/ethnographic-action-research>)
- United Nations (2004), *Human Rights and Poverty Reduction: A Conceptual Framework*. Geneva & New York.
- Uphoff, N. (2003), *Some analytical issues in measuring empowerment for the poor with concern for community and local governance*. Paper presented at the Workshop on "Measuring Empowerment: Cross-Disciplinary Perspectives" held at the World Bank in Washington DC, February 4-5, 2003.
- Van Hemelrijck, A. (2009), *LEAD Measurement Note on Empowerment in Rights-based Programming*. Implications for the work of Oxfam America.
- Veneracion, C.C. (Ed; 1989), *A Decade of Process Documentation research. Reflections and Synthesis*. Quezon City: Institute of Philippine Culture, Ateneo de Manila University.
- Williams, B. & I. Imam (2007), *Systems Concepts in Evaluation. An Expert Anthology*. American Evaluation Society, Point Resys: EdgePress of Inverness Publishers. Watson, D. (April 2006), *Monitoring and Evaluation of Capacity and Capacity Development. A Case Study Presented for the Project "Capacity, Change and Performance."* Discussion Paper No 58B. Maastricht: European Centre for Development Policy Management
- Wesley, F.; Zimmerman, B. & M. Q. Patton (2006), *Getting to Maybe. How the World is Changed*. Random House Canada.
- Woolcock, M. (March 2009), *Toward a Plurality of Methods in Project Evaluation: A Contextualized Approach to Understanding Impact Trajectories and Efficacy*. In: Journal of development Effectiveness (Vol 1, nr 1). Routledge Taylor & Francis Group.

Appendix 6 – Evaluation of the Conference

The Conference was evaluated in two ways: (1) evaluation at the end of the Conference (voting); and (2) evaluation after the Conference (Survey Monkey).

1. Evaluation at the end of the Conference

A voting method was applied. All participants (153) received a green, yellow and red card. They voted according to the following appreciation:

- Green = Good, objective reached, satisfied
- Yellow = OK, could be better but not bad
- Red = Weak, not satisfied

First the four objectives of the Conference were assessed. Afterwards participants were invited to suggest other questions to be evaluated.

Question 1: Greater clarity about what is rigorous evaluative practice that embraces complexity.
Appreciation: About 60% green, 40 % yellow, 2 red cards

Question 2: New insights about methodological options to understand and influence societal change.
Appreciation: About 50 % green, 40% yellow, 10 % red

Question 3: Inspired to take the agenda for change forward in my own role and practice.
Appreciation: About 80% green, 20 % yellow, 2 red cards

Question 4: New contacts that will help me/my organization improve evaluative practice.
Appreciation: About 60% green, 40 % yellow, 2 red cards

Participants then suggested their own criteria for assessing the Conference.

Robert Chambers: How do we appreciate the facilitation of the Conference organisers?
Appreciation: 100 % Green, two yellow cards, no red cards

Welmoed Koekebakker: How do we value the process we have gone through these two days?
Appreciation: 100% green cards

Patricia Rogers: Do we think our capability to explain complexity to others has increased as a result of this conference?
Appreciation: 50% Green cards, 50% yellow, no red cards

Timmo Gaasbeek: Do we find it useful to share more examples of complexity and how we try to make sense of it in our evaluative practice?
Appreciation: 90% green cards, 10 % yellow, no red cards

Giel Ton: Are we more aware of our limitations to deal with complexity?
Appreciation: 45% green cards, 55 % yellow cards, 3 red cards

2. Evaluation after the Conference - Survey Monkey

In addition an online evaluation (Survey Monkey) was carried out several weeks after the conference was held. Forty participants filled this online survey, giving a healthy 26% of respondents – a relatively high score for this type of survey. Survey results at:

http://www.surveymonkey.com/sr.aspx?sm=LP_2fvme44eRlo3WqXn3woRHgg3NUdFxdAgJyhVhsl5T8_3d.

Insights gained by participants were related to complexity, values, trends, different approaches for rigorous evaluation practice, linking insights in various areas of work (e.g. OD, social science, evaluative practice), the need to combine forces rather than focusing on the differences (e.g. between different schools of thought), the importance of and (often different) understanding of 'rigor', learning, etc. The insights were important to these respondents mainly as they were relevant to people's work practice and could help improve these.

In terms of action (to be) taken as a result of the Conference, the majority indicated to share their knowledge and experience informally in their organization (76.7%) or informally externally (58.1%). Formal sharing is intended by a smaller group of people: 39.5% within the organization and 20.9% externally. Nearly half of the respondents indicated to help adapt evaluative practice: 44.2% in their organization, 46.5% their own evaluative practice and 48.8% in supporting capacity development on evaluative practice. About one quarter (25.6%) intend to use the information for lobby and advocacy purposes.

About half (51.2%) already had a lot of knowledge around the core question of the Conference ('what constitutes rigor in evaluative practice that embraces complexity?') but indicated to have learned a lot. Only a small group of people (11.6%) with a lot of understanding hardly increased their understanding. The others who had very little understanding increased this a lot (18.6%) or a little (18.6%).

The elements of the Conference that were rated most relevant varied for the different respondents but generally tended towards to be more positive. Case clinics, the presentation by Maarten Brouwer and methods and more markets were about equally rated (some 30%) as the most relevant. Patricia Rogers' presentation scored highest when looking at the scores related for most and very relevant (together nearly 75% as compared to the other key note speeches, the case clinics and the methods and more markets that all ranged between 50-60%). Ritual dissent and panel discussion rated lower in these categories (total 44.2% and 39% respectively).

In terms of new information and looking at the two categories for highest relevance a somewhat similar picture can be found also in terms of inspiration although the keynote presentations by Patricia Rogers and Sheela Patel were scored as most inspiring (65.2% and 68.3% respectively in the two categories of highest relevance).

The case clinics were generally rated as relevant to very relevant although it is difficult to compare as the respondents per case clinic vary greatly. New insights or knowledge were generally gained with all case clinics. The clinic given by Jess Dart particularly stimulated new insights. Presentation of case clinics also generally was good. The relevance, insights and presentation for the 'Methods and More Market' sessions received highly variable scores.

For respondents, the conference has been a good use of their time and resources, with the majority 88.4% strongly agreeing with this statement and the rest slightly agreeing. Generally the pre-event information was also useful (67.5% strongly agreeing and 27.2% slight agreeing) but 3 people disagree (1 strongly).

General comments made include generally very positive feedback about the process and approach of the conference, the mix of methods and keeping a red thread, inspirational, good for networking, with a need for documentation and follow up of the conference. Only a few minor suggestions for improvement were made, e.g. around logistics.

This report summarises the discussions and presentations of the Conference 'Evaluation Revisited: Improving the Quality of Evaluative Practice by Embracing Complexity', which took place on May 20-21, 2010. It positions these discussions within international debates on measuring development results, through additional insights and observations from several key evaluation events in 2010/2011 during which the May conference discussions were shared.

