



Impact Evaluation: Taking stock and looking ahead

Irene Visser (documentation)
Irene Guijt (organiser, editor)
Cecile Kusters (organiser, editor)
Marlene Roefs (organiser)
Marieke de Ruyter de Wildt (organiser)
Anja Wolsky (organiser)

Conference report



EXCUSE ME, DIDN'T YOU FORGET SOMETHING?



© 2013 www.aukeherrema.nl

The Centre for Development Innovation, part of Wageningen UR (Wageningen University & Research centre) works on processes of innovation and change in the areas of food and nutrition security, adaptive agriculture, sustainable markets, ecosystem governance, and conflict, disaster and reconstruction. It is an interdisciplinary and internationally focused unit of Wageningen UR 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. The Centre for Development Innovation works to inspire new forms of learning and collaboration between citizens, governments, businesses, NGOs and the scientific community.

More information: www.wageningenur.nl/cdi



Innovation & Change



Ecosystem Governance



Adaptive Agriculture



Sustainable Markets



Food & Nutrition Security



Conflict, Disaster & Reconstruction

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.

Vis-à-vis advies is the company name owned by Irene Visser. She works as an independent consultant in the management of natural resources, climate change, landscape architecture and Monitoring and Evaluation.



Impact Evaluation:

Taking stock and looking ahead

Irene Visser (documentation)

Irene Guijt (organiser, editor)

Cecile Kusters (organiser, editor)

Marlene Roefs (organiser)

Marieke de Ruyter de Wildt (organiser)

Anja Wolsky (organiser)

Conference Report

June 2013

Project code 8141200503

Report number CDI-13-015

Centre for Development Innovation, Wageningen UR

Impact Evaluation:

Taking stock and looking ahead

Visser, I.G.

Guijt, I.M.

Kusters, C.S.L.

Roefs, M.M.I.

De Ruyter de Wildt, M.A.M.

Wolsky, A.

June 2013

Centre for Development Innovation, Wageningen UR (University & Research centre)

Impact evaluation is crucial in current times of change, with increased attention for evaluation and an increasingly complex development sector. The aid architecture is changing, and power relations are shifting, both in the development world as in the professional evaluators world. Despite the increased attention for impact evaluation, its utilisation is yet to be improved.

This report summarises the presentations and discussions of the Conference 'Impact evaluation. Taking stock and looking ahead', which took place in Wageningen on March 25 and 26, 2013. The Conference was organised and funded by the Centre for Development Innovation (CDI), Wageningen UR in collaboration with Learning by Design; the Agricultural Economics Institute (LEI), Wageningen UR; and with funding from the Netherlands Ministry of Foreign Affairs (DGIS) and Hivos. The report describes current issues around impact evaluation (IE) and focuses on the following question that was central during the conference: 'What can be done in design and communication to enhance utilisation of IE findings?'

Orders

+ 31 (0) 317 486800

info.cdi@wur.nl

Preface

Impact Evaluation (IE) has been at the centre of attention within the evaluation field for several years. The demand for IE is increasing, but *“the utility of IE remains to be proven”* and there is *“very little evidence of the use of evidence”* according to keynote speaker Dr **Irene Guijt**. Impact evaluation and utility were at the heart of this two-day conference, organised on 25-26 March 2013 by the Centre for Development Innovation (CDI), Wageningen UR, in collaboration with Learning by Design and the Agricultural Economics Institute (LEI), Wageningen UR and with some 160 participants. Questions were framed on what influences design and communication of IE findings and how these then influence the utilization of IE.

Keynote speaker Professor **Elliot Stern**, team leader of the important publication ‘Broadening the range of designs and methods for impact evaluation’ (DFID working paper 38, April 2012) indicated that *“the ‘IE brand’ (as if it is new) is now too narrowly focused, methods led rather than content led and ignoring major developments in policy and practice in international development”*. Prof. Stern indicated that there are *4 main designs: 1. Experimental (RCTs) and quasi-experimental; 2. Statistical; 3. Theory-based; 4. Case-based. They all support causation and each have their strengths and weaknesses. There is a need to combine different designs at different levels.* Furthermore, Prof Stern explained an organising framework for IE design: start from the *evaluation questions*, look at the *program attributes* and balance this with *available designs*. This framework was further elaborated during the conference in the **case group sessions** and additional **working group sessions** around the core questions of the conference. Workshop participants also framed key messages for different stakeholders that were responded to in a **panel discussion**. In addition, **reference review** has also enriched this report.

This report summarizes the key issues discussed at the conference and aims to provide an answer to the key framing questions for the conference. Of course the discussion doesn’t end here, and there is more work to be done. We hope however that this report will stimulate your thinking and actions in impact evaluation to making these more useful for primary stakeholders.

Happy reading!

Co Verdaas
Director Centre for Development Innovation, Wageningen UR

Acknowledgements

The conference on impact evaluation was held in Wageningen, the Netherlands on 25-26 March 2013. It was organised and funded by the Centre for Development Innovation (CDI), Wageningen UR in collaboration with Learning by Design; the Agricultural Economics Institute (LEI), Wageningen UR; the Netherlands Ministry of Foreign Affairs (DGIS); and Hivos.

We are grateful to the many people that helped to make this conference a success, first of all the organising committee with CDI (Cecile Kusters – coordinator; Marlene Roefs; Anja Wolsky); LEI (Marieke de Ruyter de Wildt); Learning by Design (Dr. Irene Guijt) with financial support from DGIS (Henk van Trigt) and Hivos (Karel Chambille). Without their support this conference would not have taken place.

We are also grateful to the keynote speakers Dr. Irene Guijt and Professor Elliot Stern for their stimulating ideas, experiences and concepts that helped to frame the conference from the start.

The case studies gave a crucial focus to the group discussions. We are grateful for the case study owners for their willingness and courage to openly share their experiences, which the participants then scrutinised in detail. Many thanks to: Adinda van Hemelrijck, Claudia Kornahrens, Dieneke de Groot, Ferko Bodnar, Gideon Kruseman, Inka Pibilova, Pam Baatsen, Shovan Chakraborty, Ton Dietz, Verina Ingram, and Willem Cornelissen.

We are also grateful to the facilitators of the various group sessions: Cecile Kusters (CDI); Elias Zerfu, IFPRI, Ethiopia; Jan Brouwers (CDI); Marieke de Ruyter de Wildt (LEI); Marlene Roefs (CDI); Seerp Wigboldus (CDI); Simone van Vugt (CDI); Sylvester Dickson Baguma, NARO, Uganda; Zaina Maimu, freelance consultant.

We also thank the documenters of the process: Irene Visser (key reporter/documenter, Vis-à-vis Advies); Annemiek Schrijvers (Wageningen UR student); Anja Wolsky (CDI); Bram Peters (CDI intern); Hilde-Marije Dorresteijn (Wageningen UR student); Inge Janssen (Wageningen UR student); Marco Dekker (private consultant); Modupe Osharode (Wageningen UR student); Mundie Salm (CDI); and Pauline Chibvuma (Wageningen UR student).

Of course most of all we would like to thank all the participants of the conference, as without their active participation this conference would not have been a success.

The conference organisers

Table of contents

Preface	iii
Acknowledgements	iv
Executive summary	vi
List of abbreviations and acronyms	vii
1 About the conference	1
1.1 Conference questions.....	1
1.2 Conference approach and process	1
1.3 About the report	2
2 Impact Evaluation	3
2.1 The current context of impact evaluation in international development.....	3
2.2 Defining impact and impact evaluation	3
2.3 Contestations and Considerations.....	4
3 Utilisation	7
4 Design	11
4.1 Four types of causal inference.....	11
4.2 Organising framework for IE design choices	12
4.3 Mixed methods and nested design.....	14
4.4 Emergent alternatives.....	15
4.5 Outputs of the group discussions	188
5 Communication	23
6 Conclusions	27
References and resources	29
Appendix 1 – Biographies keynote speakers	37
Appendix 2 – Summaries of case studies	41
Appendix 3 – Panel members	79
Appendix 4 – List of participants	81
Appendix 5 – Evaluation of the conference	87

Executive summary

Impact evaluation is crucial in current times of change, with increased attention for evaluation and an increasingly complex development sector. The aid architecture is changing, and power relations are shifting, both in the development world as in the professional evaluators world. Despite the increased attention for impact evaluation, its utilisation is yet to be improved.

This report summarises the presentations and discussions of the Conference 'Impact evaluation. Taking stock and looking ahead', which took place in Wageningen on March 25 and 26, 2013. The Conference was organised and funded by the Centre for Development Innovation (CDI), Wageningen UR in collaboration with Learning by Design; the Agricultural Economics Institute (LEI), Wageningen UR; and with funding from the Netherlands Ministry of Foreign Affairs (DGIS) and Hivos.

Keynote speeches by Dr. Irene Guijt, and Professor Elliot Stern set the stage. Nine case studies of different IE applications were presented. The key questions were discussed in profound group discussions. Finally, in a panel discussion, panel members were asked to respond to key issues.

The report describes current issues around impact evaluation (IE) and focuses on the following question that was central during the conference: 'What can be done in design and communication to enhance utilisation of IE findings?'

Whilst impact evaluation is a heavily contested topic, one can identify four key questions: 1. To what extent can a specific (net) impact be attributed to the intervention in this setting? 2. Did the intervention make a difference on its own? 3. How has the intervention made a difference? 4. Will the intervention work elsewhere?

The report discusses assumptions about causal inference in social theory and the implications of those assumptions for the design of impact evaluations. It argues for the importance of alignment of evaluation questions and programme or policy attributes with a repertoire of available designs of impact evaluation. A good impact evaluation design should also include clear communication and utilisation strategies for different stakeholders. The utility of the IE process can be improved by joint development of evaluation questions and joint sense making.

Utilisation is the first programme evaluation standard, yet it receives little attention in IEs. Uptake of results needs explicit thought from the onset and only then can IEs be useful. To avoid the waste of time and resources, the conference called for a responsible use of evaluation findings for those who have the capacity to effect change.

List of abbreviations and acronyms

3IE	International Initiative on Impact Evaluation
AEA	American Evaluation Association
BMGF	Bill and Melinda Gates Foundation
CDI	Centre for Development Innovation, Wageningen UR
CDI	Centre for Development Impact (partnership between IDS and ITAD)
DAC	Development Assistance Committee
DFID	Department for International Development, U.K.
DGIS	Dutch Ministry of Foreign Affairs
EBP	Evidence Based Policy
EES	European Evaluation Society
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit
IDRC	International Development Research Centre, Canada
IE	Impact Evaluation
LEI	Agricultural Economics Institute (LEI), Wageningen UR
IOB	Policy and Operations Evaluation Department (IOB) at the Netherlands Ministry of Foreign Affairs.
JCSEE	Joint Committee on Standards for Educational Evaluation
J-PAL	Abdul Latif Jameel Poverty Action Lab
NGO	Non-Governmental Organisations
NONIE	Network of Networks on Impact Evaluation
OECD	Organisation for Economic Co-operation and Development
PD	Paris Declaration
QCA	Qualitative Comparative Analysis
RBA	Results Based Aid
RCT	Randomised Control Trials
TBE	'Theory Based' Evaluations
ToC	Theory of Change
TOR	Terms of Reference
VFM	Value for Money

1 About the conference

1.1 Conference questions

Impact Evaluation (IE) has been at the centre of attention within the evaluation field for several years. Governments, civil society organisations and companies want to know if their money has contributed to positive changes in peoples' lives, and if so, how and why. Evaluation is not – and never has been neutral. So too within the specific area of IE, different perspectives and interests vie for centre stage and power relations shape what transpires.

IE addresses the fundamental question 'did the programme or policy in some way cause the expected effect?' Prof. Stern (keynote speaker) indicated there are four key questions for IE:

1. To what extent can a specific (net) impact be attributed to the intervention in this setting?
2. Did the intervention make a difference on its own?
3. How has the intervention made a difference?
4. Will the intervention work elsewhere?

In addition, utilisation of evaluation lies at the heart of the evaluation profession. The first and most important internationally accepted programme evaluation standard is 'utility' (see also Box 1 below). Dr. Patton defines 'Utilisation-focused evaluation' as 'evaluation done for and with specific intended primary users for specific, intended uses' (Patton, 2008).

The focus on utility was central during the two day conference 'Impact evaluation. Taking stock and looking ahead', which took place in Wageningen on March 25 and 26, 2013. Using cases and in separate group sessions, participants explored the following underlying questions:

- What can be done in design and communication to enhance utilisation of IE findings?
- What influences IE design?
- What influences communication of IE findings?
- How is utilisation of IE findings influenced by IE design and communication?

1.2 Conference approach and process

The participants of the conference included policy makers, evaluators, M&E professionals, and managers with M&E or RBM responsibilities. They came from 36 different countries. All were actively engaged in reviewing a range of IE cases with different IE methodologies and exploring the core questions.

Two keynote speeches by Dr. Irene Guijt, Learning by Design/Overseas Development Institute, and Professor Elliot Stern (Lancaster University/University of Bristol/editor 'Evaluation') framed the conference and set the stage for exploration (see Appendix 1 for their biographies). Case group sessions were then held around different IE applications during which the key questions of the conference were discussed (see appendix 2 for the case descriptions).

On the second day, participants debated the core conference questions, pooling the case experiences and all other experiences. Each group generated and presented key messages to a of different stakeholders involved in impact evaluation (see Appendix 3). In a panel discussion, each group shared key reflections and ideas for the future. Panel members, requested to represent different stakeholders important for IE, were asked to respond to key issues.

1.3 About the report

The report is structured around the four central concepts that shaped the conference: impact evaluation, utilisation, design and communication. It concludes with messages for key stakeholders. As the main text is succinct, information can be found in the appendices. Please note that the text uses 'policy' and 'programme' interchangeably.

All relevant information related to the conference can be found (register first!) at:

www.managingforimpact.org/group/conference-impact-evaluation-2013

2 Impact Evaluation

2.1 The current context of impact evaluation in international development

The international development agenda and its players are changing dramatically. Economic relationships are shifting and countries like India, China and Brazil are new forces affecting the development scene. Philanthropic foundations, such as the North American foundations, have come into play and set new agendas. Businesses play an increasingly important role. The financial crisis is putting extra pressure on the already reduced budgets for international aid and this affects the international NGOs, bilateral organisations, and multilaterals in what they can do.

These shifts and related paradigms influence how aid is being framed, implemented and evaluated. Overall, the effect has been one of drawing increasing attention to 'results' and 'evidence' as critical for justifying expenditure and scaling up programmes or policies. A much stronger focus on quantifiable measurements of impact exists, accompanied by the hotly contested topic of the 'gold standard' for method choice.

2.2 Defining impact and impact evaluation

There are different definitions of 'impact' and of 'impact evaluation'. Existing definitions of 'impact' fall into two main groups – method-led and substantive or content-led. Definitions of 'impact' indicate different positions around impact and how to evaluate it.

The most commonly used and substantive definition is that of the OECD-DAC Glossary (2002), which starts from the content of international development and interventions funded through aid. It defines impact as: *“positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended”*. The OECD definition calls for theories about cause and effect.

The World Bank and 3ie define 'impact' as: *“the difference in the indicator of interest (Y) with the intervention (Y1) and without the intervention (Y0). That is, impact = Y1 - Y0. An impact evaluation is a study which tackles the issue of attribution by identifying the counterfactual value of Y (Y0) in a rigorous manner”* (Ravallion in White 2009).

According to J-PAL¹, *“impact evaluations estimate program effectiveness usually by comparing outcomes of those (individuals, communities, schools, etc) who participated in the program against those who did not participate”*. They also say that *“the primary purpose of impact evaluation is to determine whether a program has an impact (on a few key outcomes), and more specifically, to quantify how large that impact is”*². The World Bank/3ie and JPAL definitions tend to be more methods-led than content-led.

¹ Abdul Latif Jameel Poverty Action Lab

² Source: <http://www.povertyactionlab.org/methodology/what-evaluation/impact-evaluation>

In a recent study for DFID³, Prof. Stern's team uses the following working definition (inspired by the OECD/DAC initiated Paris Declaration on Aid Effectiveness):

Evaluating the positive and negative, primary and secondary long-term effects on final beneficiaries that result from a development intervention;

Assessing the direct and indirect causal contribution claims of these interventions to such effects especially for the poor whether intended or unintended;

Explaining how policy/programme interventions contribute to an effect so that lessons can be learnt.

The European Evaluation Society has taken the position that impact evaluation must not be defined by a particular method (EES 2007). Prof. Stern said that impact evaluation should not be confused with any one design, methodology or philosophy: "The core of what IE must be about is: did the programme or policy in some way cause the effect?"

2.3 Contestations and Considerations

As noted above, the increased attention to results and evidence have led to stronger emphasis on quantifiable measurements of impact, accompanied by a surge of interest in (quasi) experimental methods, irrespective of their appropriateness.

Narrowing focus. There is a narrowing of focus on results, which can be oppressing if the ways to measure those results are also reduced. Fresh takes on the subject are needed. See Eyben's (2013) excellent contribution about the dominant role of the results-and-evidence discourses. Some people and organisations uphold the position that there is only one interpretation of scientific rigour, namely the gold standard of (quasi)experimental designs. Organisations such as 3IE and JPAL have a strong focus on these approaches. In one of [blogs](#) Duncan Green (2013) refers to a [brilliant new post](#) by Ed Car about the clash between quants and quals, arguing that while quants can establish causation, only quals can explain *how* that causation occurs.

Degrees of broadmindedness. All those engaged in IE are emerging from a period in which positions were taken and defended around definitions, methods and utility. Currently, the climate is one of seemingly more open discussion with some (very modest) investments being made in developing robust alternatives that complement (quasi-)experimental approaches. The Stern et al paper (2012) is one such investment, as is IFAD and BMGF's support of participatory impact assessment and learning (currently underway) and many other initiatives. However, to date, these investments are very small compared to the hundreds of millions of dollars going to (quasi-)experimental evaluations. Furthermore, claims to being open to mixed methods are not always matched by practice, as can be seen by looking at the 3ie database and grant allocations. These remain narrowly focused on (quasi-)experimental methods. JPAL only undertakes RCTs – but does not claim to be multi-method oriented.

From 'how well' to 'what' to 'why'. Only looking back may not give us what is needed to move forward with confidence. Dr. Guijt challenged the utility of this way of using evidence: 'Will the past always be a good predictor of the future? What about when what works, doesn't work? And when what doesn't work, works?' Waiting for enough replication studies to give more confidence to an IE is a luxury some cannot wait for –

³ Stern, E.; Stame, N.; Mayne, J.; Forss, K.; Davies, R.; Befani, B. DFID Working Paper 38. 'Broadening the range of designs and methods for impact evaluations'. DFID, London, UK (2012) vi + 24 pp.

nor can it cover all the variation that the future has in store. The challenge is to broaden questions from results and targets – or ‘how well has this project worked’ towards curiosity about *‘Why do or don’t things work?’* Framing it in terms of ‘what works or what doesn’t work and what will work?’ can help bypass the term ‘failure’ often loaded with emotions.

Capacity building for options. The large demand in IEs has led to a great surge of requests for matching capacities. The Regional Centers for Learning on Evaluation and Results (CLEAR network) is an interesting example of this. Other initiatives include 3ie’s training courses, while [BetterEvaluation](#) and other web platforms such as MyMandE help guide evaluators through the many methodological options. These capacity building processes, depending on their origins, will be more or less inclusive of different approaches to IE.

Public debate on future and politics. Recent conferences are another source of insights. An example is the IDS conference (26-27 March 2013) “[Impact, Innovation and Learning: Towards a Research and Practice Agenda for the Future](#)” This also launched the Centre for Development Impact (CDI). CDI is a partnership between IDS and ITAD; both organisations share the view that impact evaluation should draw on a broad range of methods and be appropriate for specific requirements, questions and constraints. Another example is the [Big Push Forward Conference](#), held in April 2013 that dealt with the politics of evidence. One of the organisers blogged: “I reassure myself that policy is never evidence-based, but hopefully can be evidence informed – and that we all value evidence.” ([blog Guijt Big Push Forward, April 2013](#)).

Diversifying approaches. Within that changing context, both keynote speakers pointed to the need to open up space for different and appropriate kinds of approaches to impact evaluation beyond the focus on (quasi-)experimental approaches. There are a number of mainstream research and scientific designs that need to be brought in to broaden the IE portfolio (Stern 2013). Different valid definitions of impact and approaches to impact evaluation exist, and different questions need to be answered in impact evaluation than can not be dealt with by (quasi-)experimental approaches alone. Also the attributes of programmes and policies vary. This calls for an *appropriate design*, for example interventions that address complex issues need a mixed method or nested approach. Theory-based and case-based approaches might also be suitable. (See also the chapter on ‘design’).

Cost considerations. Undertaking IE is often costly. With overall budgets for development policy and programmes reducing, budgets for evaluation need to be considered wisely and ethically – from the perspective of their overall utility. Keynote speakers stressed the importance of careful consideration about when an IE is useful or not – and of which kind. The following criteria are increasingly considered useful guides for deciding to carry out an IE: innovative programmes that may be scaled up, programmes in a risky context; and high investment programmes. A Harvard Business Review blog by Ebrahim (2013) ‘Let’s be realistic about measuring impact’, shows that three large organisations that are seriously investing in impact evaluation, do it by making links between actions, outcomes and impact, but have consciously opted not to undertake IEs of the narrower, statistical sort. So alternatives exist that are useful to organisations.

Small ‘e’, big ‘E’. Amidst the focus on IEs – of whatever kind – we lose sight of the vast amounts of data, little ‘e’, that are produced by standard M&E processes, the results requirements of everyday implementation. These could be an important source of insights, if analysed well, for impact-oriented monitoring, for example. “*The first observation is that the vast majority of the snippets concern the day-to-day practice of small-e evidence – results and targets in management – rather than large-E evidence of establishing theories of change or development approaches. The stories are about concerns and opportunities in the nuts and bolts of the practice of evidence – what ‘e’ is being collected, how it is used, and to what effect. They show that the devil is in the detail, the practice.*” ([Brendan Whitty, 2013](#)).

Philosophical perspectives. A broad consensus exists in the philosophy of science that causality (from a realist perspective) is multi-causal and non-linear. Systems are open to 'external' influences; causes interact with each other; context changes everything; and you need a theory to understand how when and where causes make a difference. Dr. M.Q. Patton, one of the world's leading evaluation experts, has argued that that the multitude of influences on any initiative are so vast, changing each minute, that the notion of counterfactuals in the narrow sense – and any method that depends on this – is theoretically dead. The same input and action can have quite different consequences in a different context. Prof. Stern referred to Cartwright and Hardie (2012) in their publication about using evidence for reliable policy. Professor Stern: "Just because we know something has 'worked' somewhere doesn't prove it will work here or anywhere. We need to understand: 'what causal role your policy might play, and what support factors it will need in your specific context'".

3 Utilisation⁴

How much evidence exists that evidence is used for policy and programming? Not much, to date. Some are relieved by this, commenting on the bad quality of IEs. Utility is essential, otherwise the money spent on evaluations is a waste of time and money. Therefore, the quality of impact evaluations needs to improve, as well as the thinking about utilisation.

A recent process evaluation of 3IE (Morton et al 2012) reports that there is little evidence for the uptake of evaluation by policy makers: “To have an impact on quality, 3ie will need the concerted support of the users of impact evaluation: policymakers and donors. But quality is the true public good in impact evaluation for development.” (ibid: 132)

Therefore, a utilisation focus in the design and communication of IE is crucial: “Not only the outcomes, but also the process can influence the way people make decisions and have an impact on the lives of the people it was intended for” (Kusters, et al 2011). So it is important to think about utilisation from the onset, what the right questions are, for whom, what the commitment is to using the findings, and what possible resistance to change might be.

Dr. M.Q. Patton (2008) stated: “Therefore, the focus in utilisation-focused evaluation is on intended use by intended users. Since no evaluation can be value-free, utilisation-focused evaluation answers the question of whose values will frame the evaluation by working with clearly identified, primary intended users who have responsibility to apply evaluation findings and implement recommendations.”

Internationally agreed standards for programme evaluation put utility on top of the list.

The five agreed programme evaluation standards of the AEA/JCSEE* are: utility, feasibility, propriety, accuracy and evaluation accountability (see Box 1). The sequence is not accidental. Utility is the first on the list. Important to note also is that there is a difference between research and evaluation: researchers are only measured against accuracy standards whilst for evaluation the five evaluation standards are important (Guijt et al, CDI report 2012).

Box 1. The 5 internationally agreed standards for program evaluation⁵:

1. *Utility*. The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.
2. *Feasibility*. The feasibility standards are intended to increase evaluation effectiveness and efficiency.
3. *Propriety*. The propriety standards support what is proper, fair, legal, right and just in evaluations.
4. *Accuracy*. The accuracy standards are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality.
5. *Evaluation accountability*. The evaluation accountability standards encourage adequate documentation of evaluations and a meta-evaluative perspective focused on improvement and accountability for evaluation processes and products.

⁴ This section is based on some references and the discussions. While the questions were geared towards the utilisation of impact evaluation, the discussions also were based on utilisation of other kinds of evaluations. This is not a problem, because probably challenges are similar

⁵ ‘The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users’. Donald B. Yarbrough, Lyn M. Shulha, Rodney K. Hopson, Flora A. Caruthers, 2011. 3rd Edition, Sage Publications, USA.

In the case study discussions (appendix 2) one of the questions was: To what extent has the IE been utilised and why/why not? What influenced this? Whatever you do, think about utilisation was the key message of the conversations.

Groups highlighted the following four issues.

1. *Rigorous thought beyond methods.* “Rigour is not only about rigorous (scientific) methods, it covers the whole evaluation process: design, conducting and *using it*. Evaluation should be useful and used and has to produce meaningful evidence, accessible to primary stakeholders” (Patton cited in presentation van Hemelrijck)
2. *Common sensemaking.* Sensemaking and common understanding of impact/results improved the utilisation. This includes the informal sharing. An example of use, shared by one of the participants is that communities were empowered to assess impact; they then used findings to develop their work plan.
3. *Readiness and timing.* The timing of an evaluation is important, so that there is an opportunity for using the findings at an appropriate time. Political will and the capacity to make use of findings or to make changes is essential. One participant shared the example where a report for evidence planning was not used, since regional political bodies were not ready to use it due to political sensitivities.
4. *Clarity about use.* Clarity about how the evaluation is going to be used is important. Kusters et al (2012) distinguish different types of use: direct, conceptual, symbolic process, relational, value, and external. Some of the case studies of this conference were still in too early development to be used fully, while others had more ‘frustrations’ about not reaching the potential use. The example of PADev (Box 5, next chapter) mentions that the methodology has been used by other universities and institutes. The case study on HIV/AIDS (Box 11 next chapter) mentions that the institute has used the findings to advocate for the need to give more focus to younger children infected and affected by HIV in different meetings.

Box 2. illustrates a narrow focus, in this case a focus on ‘success’ could impede wider theory building around the notion of scaling up, on which the IE is focused. It also shows that through learning cafes, the user group is amplified to a wider group in the organisation.

In the next chapters we will see how design and how communication link to utilisation.

Box 2. Example from learning across scales

'Scaling up / Impact at Scale – an Independent Evaluation of GiZ' Claudia Kornahrens

The quality and effectiveness of GiZ in individual projects are widely acknowledged, but measurable scaling-up effects are sometimes limited. How can GiZ better roll out the results of successful programme approaches to cover a wider area? What steps must be taken to make national laws and strategies effective nationwide? The 'Corporate Strategy Evaluation on Broad Impacts/ Scaling Up' seeks to learn about successful cases across GiZ of scaling up to get to impact at scale. 'Success' is how the design has been framed and projects were randomly sampled from GiZ's portfolio. However, much can be learned from those projects that took risks but were not successful in achieving impact at scale. The design could have been more inclusive of 'what works and what doesn't work', leading to insights based on learning across a more diverse range of projects funded by GiZ. This broader focus might have led to a purposive sampling strategy that looked at successful and unsuccessful cases.

Learning cafés which are attached to a management response system can be a useful strategy to influence the utilization of evaluation findings. A learning café at GiZ can be understood as a dialogue-oriented format in which evaluators present their recommendations based on lessons learned as well as findings on what has worked, what not, and why. Afterwards the presented recommendations are discussed between mixed groups of management and staff who have a stake in the evaluations. Emphasis is placed on dialogue between evaluators and (operative) personnel in order to identify specific conclusions upon the findings and recommendations. Only selected persons (learning addressees) are invited to join the learning café in order to promote learning in a "secure environment" in which also critical experiences can be discussed without putting personnel under pressure to justify negative outcomes. However, learning addressees are not only those responsible/involved in the evaluated objects, but also those who may benefit from the evaluation results, such as expert colleagues and executive managers from other areas within the company. The purpose is to stream the learning effect up to the sector and enterprise level. At the end of the learning café, the conclusions discussed between evaluators and participants are agreed on and adhered to in a management response system.

See appendix 2 for more details.

4 Design

Prof. Stern made a theoretical contribution to the conference to explain the current thinking about designing impact evaluations, by explaining causal inference (section 4.1), an organising framework for design choices (section 4.2), the possibility of nested approaches (section 4.3) and emergent alternatives, including some cases (section 4.4). The conference discussions around the cases, and participants' experiences further enriched the framework (section 4.5).

4.1 Four types of causal inference

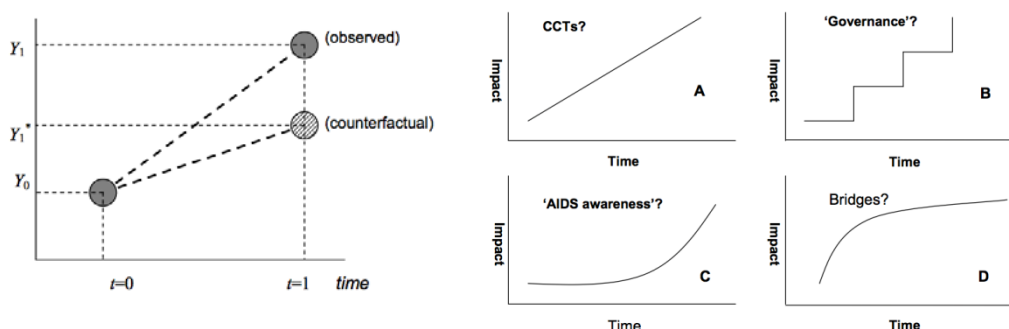
Since IE is about the question: 'did the programme or policy in some way cause the effect?', social theory is needed to research/evaluate this.

Within the social sciences, there are four types of causal inference, i.e. ways of looking at causality (Stern et al 2012):

- **Regularity frameworks** that depend on the frequency of association between cause and effect - the inference basis for statistical approaches to IE;
- **Counterfactual frameworks** that depend on the difference between two otherwise identical cases – the inference basis for experimental and quasi-experimental approaches to IE;
- **Multiple causation** that depends on combinations of causes that lead to an effect - the inference basis for 'configurational' approaches to IE;
- **Generative causation** that depends on identifying the 'mechanisms' that explain effects – the inference basis for 'theory based' and 'realist' approaches to IE.

Crucial for any of the four types is to consider the shape of change. Woolcock (2009) shows the bias and distortion that causes errors in understanding change, if only two points in time are being studied (see Box 3).

Box 3. The logic of the comparison/ non-linear impact pathways (Woolcock, 2009)



"Understanding the efficacy of development projects requires not only a plausible counterfactual but also an appropriate match between the shape of impact trajectory over time and the deployment of a corresponding array of research tools capable of empirically discerning such a trajectory. At present, however, the development community knows very little, other than by implicit assumption, about the expected shape of the impact trajectory from any given sector or project type, and as such is prone to routinely making attribution errors. Randomisation per se does not solve this problem. The sources and manifestations of these problems are considered, along with some constructive suggestions for responding to them."
Abstract Woolcock 2009

4.2 Organising framework for IE design choices

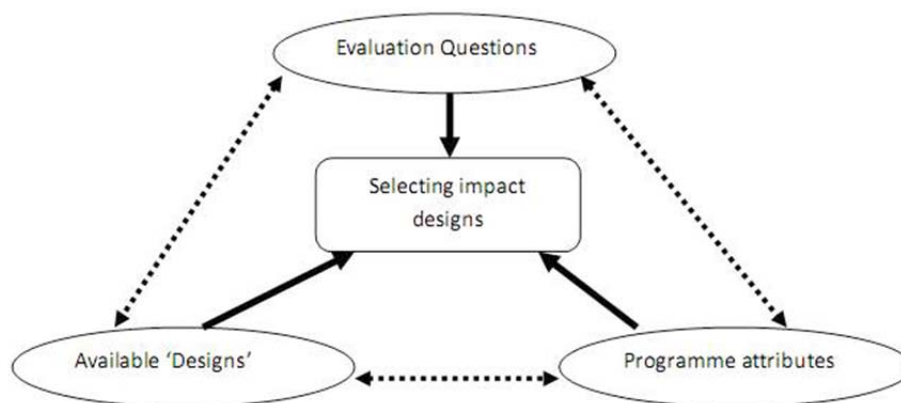


Figure 1 Organising framework for IE design choices

Prof. Stern presented an organising framework for IE design (see figure 1), the 'IE stool' with three legs (Stern et al 2012)⁶. The framework shows that the determinants of IE design choice must be: (1) the evaluation questions we want to address, (2) the attributes of a programme or policy, and (3) the available designs.

Leg 1. Evaluation questions

Prof. Stern refers to four key questions in IE:

1. To what extent can a specific (net) impact be attributed to the intervention in this setting?
Conditions that suit experiments and statistical models.
2. Did the intervention make a difference on its own?
Contributory causes and causal packages
3. How has the intervention made a difference?
Explanation and importance of theory
4. Will the intervention work elsewhere?
External validity, transferability and generalisation

Not all questions may be equally useful or necessary for each IE. So consider which questions make most sense for the task at hand. This depends on one's world view: how directly influential and powerful does one think that development interventions are? It also depends on what one thinks policy makers or programme staff are interested in: looking backwards or forwards? Duplication or replication? In one place or many places? These questions are the starting point for making methodological choices for IE.

⁶ ibid footnote 5.

Leg 2. Programme attributes

The Department for International Development (DFID) that commissioned the Stern paper asked the team to look at a portfolio of 'difficult to evaluate programmes.' This included complex development programmes, like climate change mitigation, strengthening governance, post conflict reconstruction, gender equality and natural resource management. For these types of programmes, IE is recognised to be more difficult and yet they comprise an important part of aid budgets and efforts.

In designing IE, clarity needs to exist about the attributes of the programme or policy being evaluated. Drawing on systems thinking and policy research, specific attributes of programmes that affect what should and can be assessed include⁷:

- Linear causality and 'trajectories';
- Timing and long-term effects;
- Standardisation and customization;
- Hierarchies and 'nested' systems;
- Comprehensive and targeted interventions; and
- Uncertainty and risk.

If, given the above generic attributes, the programme has the following features, then the first two frameworks mentioned in 4.1 are not suitable because one is dealing with:

- Overlap with other interventions with similar aims;
- Multiple and diverse activities and projects;
- Customised non-standard activities/ interventions in diverse contexts;
- Programmes working 'indirectly' through 'agents' and often at different levels and stages;
- The likely impacts of programmes were long term;
- Working in areas of limited understanding/experience;
- Intended impacts are difficult to measure and often intangible.

See Stern et al 2012 for further ideas about design implications of complex programme attributes.

Leg 3. Available 'Designs'

Different questions require different methods. All designs have requirements, i.e. conditions under which they do and do not apply, strengths and weaknesses (see Table 1), and so do different methods. The table illustrate the logic of using a mix of methods to address the specific evaluation questions and deal with the specific program attributes.

⁷ *ibid* footnote 5.

Table 1: Requirements, strengths and weaknesses of four approaches to causal inference (Stern et al. 2012)

	Regularity	Experiments/ Counterfactuals	Multiple causation	Generative/ Mechanisms
Requirements	Many (or highly diverse) cases Independent causes	Two identical cases for comparison Ability to 'control' the intervention	Sufficient number of cases Availability of cases with comparable characteristics	One case with good access to multiple data sources Theory to identify 'supporting factors'
Potential strengths	Uncovering "laws"	Avoiding some kinds of bias	Discovery of typologies Dealing with limited complexity	In-depth understanding Focus on the role of contexts 'Fine-grained' explanation
Potential weaknesses	Difficulties explaining 'how' and 'why' Construct validity	Generalisation Role of contexts	Difficulties interpreting <u>highly</u> complex combinations	Estimating extent Risks of loss of evaluators' independence

4.3 Mixed methods and nested design

The challenge of choosing an appropriate methodology is not a choice between quantitative or qualitative methods. A mixed methods approach seeks to *integrate* social science disciplines with predominantly quantitative and predominantly qualitative approaches to theory, data collection, data analysis and interpretation (Bamberger 2012). An integrated approach strengthens the reliability of data, validity of the findings and recommendations. It also broadens and deepens the understanding of the processes through which program outcomes and impacts are achieved, and how these are affected by the context within which the program is implemented.

Nested designs can be seen as a Russian doll (Stern 2013) where different wooden dolls are inserted into each other (see figure 2). Each doll represents a slice or level of a complex programme, which entails a suitable mixed methods approach. The overall impact is 'covered' by the integrated set of dolls, i.e. methods.



<http://beingzhenya.com/2011/08/23/russian-dolls/>

Figure 2: Russian dolls as metaphor for nested approach

Prof. Stern stressed two critical design challenges when it comes to complex programmes:

1. How to 'slice' complex programmes made up of multiple and interdependent sub-programmes for evaluation; and
2. How to sequence the evaluation of long-term programmes or emerging programmes – where learning is vital.

Where different initiatives are concurrent and known as 'the programme', we can look at different levels and isolate sub-initiatives as well as looking at wider context, in order to show results and how they are linked. Prior to this we need to consider carefully how far and under which circumstances it is reasonable to break complex programmes down and evaluate their sub-parts separately.

4.4 Emergent alternatives

The Stern paper was commissioned due to a growing recognition of the need to broaden the range of IE designs. Although evaluators have become more interested in IE, Prof. Stern argued that the IE 'brand' is now too narrowly focussed – methods-led rather than content-led and ignoring major developments in the policy and practice of international development. Due to this shift in the impact evaluation world, new initiatives are emerging that explore the possibilities that mixed methods or complexity-sensitive approaches have to offer.

Some of the emerging alternatives are:

1. *Realist/realistic evaluation* focuses on the question of 'what works for whom under which conditions', and therefore seeks to disentangle outcomes, mechanisms and context (Pawson and Tilley, 1997);
2. *Contribution analysis* is an analytical framing for assessing causal questions and inferring causality in real-life program evaluations;
3. *Process tracing* is an approach for causal inference that focuses on the use of clues (causal-process observations) to adjudicate between alternative possible explanations;
4. People's *narratives* for attitudes/behaviour shifts like Sensemaker (<http://www.sensemaker-suite.com/smsite/index.gsp> or Most Significant Change Technique (Davies and Dart, 2005) see Box 4);
5. *PADev* (see box 5);
6. *Configurational Comparative Methods* pave the way for an innovative approach to empirical scientific work through a strategy that integrates key strengths of both qualitative (case-oriented) and quantitative (variable-oriented) approaches. It is ideally suited for "small-N" or "intermediate-N" research situations, which both mainstream qualitative and quantitative methods find difficult to address⁸.

1.1 ⁸ Benoît Rihoux and Charles Ragin (2009). 'Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques'. Sage publications

Other options include PIALA (Participatory Impact Assessment and Learning Approach), currently being piloted and complexity-sensitive evaluation (see box 6).

Box 4. Example of a case study to discover unexpected significance

'The Most Significant Change technique for assessing impact with community and children (India) – Inka Pibilova

In order to collect unexpected (positive and negative) impacts, the Most Significant Change (MSC) technique was used to assess even “soft” changes, such as changes in attitudes of children, without having a baseline study. The Most Significant Change (MSC) technique is explained in the free MSC Guide at <http://www.mande.co.uk/docs/MSCGuide.pdf>. The method was triangulated with other quantitative and qualitative methods – data review, key informant interviews, focus groups, community group discussions, a survey among schools and site visits.

Community volunteers were trained and sent to different villages to collect MSC stories or drawings, depending on the age of the children. Apart from drawings of the building, the sports and learning materials, they mentioned Child Clubs (established during the project), where they discussed school problems like lack of potable water or dirty toilets. Others drew their friends, homes, flowers, local animals. All stories and drawings were debated jointly with project staff and community volunteers. The selected top MSC story concerned the Child Clubs, which were not originally planned in the project logical framework and were just a “side activity” of the project manager. This was rather surprising for the project staff as well as for donors, who spent most of the resources on enhancing the school infrastructure.

See appendix 2 for more details.

Box 5. Example of a case study of sense making with intended beneficiaries

'A beneficiary picture of development and change (PADev: holistic and participatory approach to development assessment) – Ton Dietz and Dienneke de Groot

PADev is an innovative toolbox of techniques and methods, which aims to achieve a bottom-up assessment of development and change in a particular area over a period of time based on the value systems of the population. PAdDev is a holistic and participatory approach to development assessment (see for the toolkit: www.paddev.nl).

A toppled perspective – from those participating in programmes and projects focussing on their development in various ways- was deemed necessary to be able to learn on the why questions and to assess the long term impact of inducing change in a situation of multiple stakeholders.

PADev can be differentiated from conventional assessment methodologies which often focus on a single agency, project/programme or sector with quite a “technical” expert-driven approach to evaluation of output, effect and sometimes impact. In PAdDev, participants assess a wide range of changes, projects and agencies based on their perceptions. Further, PAdDev assessments typically look back at development and change over the past twenty to thirty years. This yields extremely valuable information for NGOs and other agencies, e.g. the government and foreign donor agencies' in the area: they learn about their own impact vis-à-vis other actors, and in addition, they find out which types of projects have been regarded as most effective and relevant and more importantly: for whom and why. This can be an important lesson for future interventions. PAdDev should not be thought of as a replacement for conventional approaches, but rather as an alternative approach that enables a “big picture” of development and change to be constructed for an area.

One of the main strengths of the approach – its ability to compare the contribution made by particular projects with other agents of change – also holds a limitation. Its broad scope inevitably reduces the focus on specific projects or interventions. Sense making meetings are important for the validation of findings and hopefully the increase of utility. What remains a problem, also in PAdDev, is reaching the ultra-poor as participants and beneficiaries of this method. However, attempts have been made to focus PAdDev specifically on these ultra-poor, and experiments continue to develop that further.

See appendix 2 for more info

Box 6. Example of complexity sensitive evaluation in Ethiopia

'Impact measurement and learning in complex collaborative settings, in the case of smallholders' productive water rights in Ethiopia – Adinda van Hemelrijck

Oxfam America commissioned the Impact Evaluation to IDS. The IE deals with ongoing impact measurement and learning, to support transformational development. The purpose of this approach is to generate a collective understanding among the program stakeholders of institutional bottlenecks, technological and infrastructural challenges, conflict and sustainability issues, and community-level organisational and social aspects of water resource development for smallholder productivity, with a view to the complexity of the programme and context. The desire to promote dialogue among stakeholders around evidence that may contribute to generating changes in interactions and behaviours.

The Ethiopian water program aims at enabling vulnerable agriculturalist and agro-pastoralist communities in moisture-stressed areas to access, use and manage productive water resources.

Impact is defined and measured, not as a change in a single dependent variable, but as a change pattern or the interactive changes in a core set of interrelated indicators that *together* measure complex systemic change. The identification and analysis of causing mechanisms help explain why observed change patterns occur as they do, and how they are generated, in a particular context. (Befani, 2012, p. 19; Pawson, 2008, pp. 14–15; Stern et al., 2012, pp. 24, 48).

The program's Theory of Change forms the basis for identifying and measuring the core mechanisms of change in each particular time frame. In the water program's first phase (2010-2013), the core mechanism is the core proposition of 'co-investment' that involves the development of individual capacity and collective responsibility (based on a tri-partite agreement) of communities, NGOs and local government for the sustainable design, construction and management of small-scale productive watering systems. The ToC visualizes the investments and institutional changes needed to support this. Improvements in smallholders' capacity, awareness and organization, and the expansion of women's influence and leadership, presumably will lead to better governance of the water systems, increased productivity, better access prices in the markets, and thus higher revenues. This will form the basis for reinvestments in the water systems and cooperative business enterprises, which over time will contribute to strengthening livelihoods and food security (Oxfam America, 2009; Van Hemelrijck, 2013).

- The complexity of the problem that the program is trying to tackle = WRD/M = "wicked problem" requiring complex systemic change.
- The degree of complexity of the program and its context (chronicle drought and food insecurity; high vulnerability and low capacity; lack of infrastructure, access to markets and services; high degree of agro-ecological, socio-cultural and political-economic variability; relatively high risk and unpredictable outcomes of investments; multiple interests and potential conflicts from local up to national and even international levels).

"Rigorous thinking" in IE (Patton, 2012) acknowledges and deals with the politics of impact knowledge and its use, by:
1. critically analysing stakeholder power relations
2. facilitating inclusive IE design and validation (with spaces for meaningful and equal participation)
3. combining/mixing and nesting methods (incl. sampling) in ways that rigorously produce and triangulate various kinds of evidence that are meaningful, rigorous and coherent Ethiopia case: 1, 2 & 3 are built-in into the ongoing impact measurement approach.

Appendix 2 shows more details.

4.5 Outputs of the group discussions

4.5.1 Output group discussions I: Case discussions enriched the design framework

The following questions were discussed in groups: 'what choices were made – explicitly or implicitly – in design and how do these influence utility'? The sessions and discussions led to an elaboration of the organising framework for IE design offered by the Stern et al study. Figure 3 shows the additions to the framework in red, and the explanation in the text below.

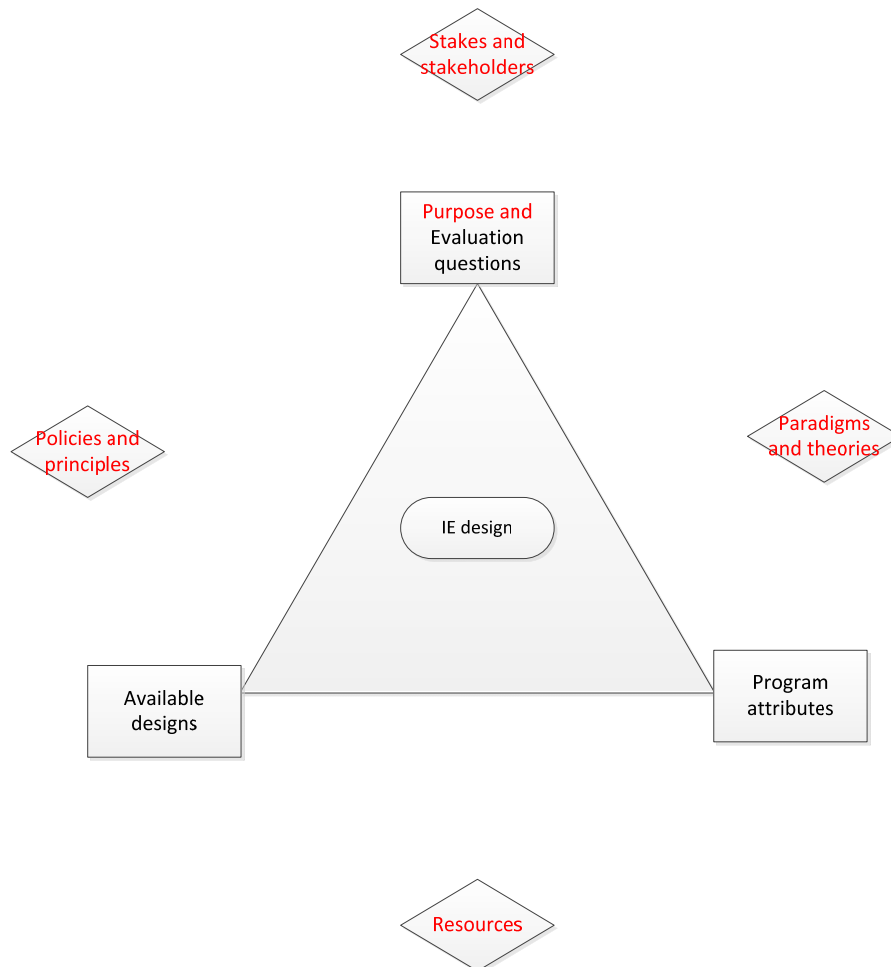


Figure 3: Additions to the IE framework by participants

Stakes and stakeholders: To improve design and enhance utilisation, it is important to analyse stakeholder engagement early on in the design phase. Analyse which stakeholders to involve as primary intended users. All kinds of issues need to be negotiated since stakes around resources and power are present. Participants noticed the importance of making decisions that are based on the programme itself and not on other interests.

Policies and principles: Participants also thought that (in)formal policies for the evaluation of policies, and the space for innovation are crucial. An evaluation policy should not only have principles around disclosure and transparency, but also demand clarity about the use of the evaluation.

Paradigms and theories: Paradigms and theories are guiding choices in evaluation methodology. Evaluators need to make explicit what ideas and theories are chosen and why. This determines methods and approaches, and whether concepts are understood similarly. E.g. the word 'rigorous' might mean different things depending on whether you are an anthropologist or statistician.

Purpose of IE and evaluation questions: The purpose of evaluations influences impact evaluation questions. It is crucial to be clear why you do IE and to be clear about the evaluation questions and how methodological designs can answer these.

Design availability: Participants said that "We tend to use what we know, not what is available (and some of it we don't know)". Therefore it remains important to share knowledge around available designs and their advantages and disadvantages.

Programme attributes: All kinds of programme attributes determine the design. These include the subject of the evaluation, organisational characteristics, the stage of the intervention, the lifespan, the complexity etc. Also the availability of baseline data or predefined indicators or a control group is important. Naturally, context and situational analysis are important. In specific cases, it was difficult to find an evaluator willing to go into conflict areas.

Financial and human resources: Available resources are influential. This not only covers the funding, but also the expertise and capacities within and outside organisations. The availability of data is a key issue to consider, as well as access to key stakeholders and information sources.

Given the variation in all these variables, it is inevitable that each IE design is tailor-made to the situation. Below are three examples (see appendix 2 for more detailed versions of these cases). Box 7 shows how bias is reduced by using a very transparent system to engaging stakeholders. Box 8 illustrates one way of working with value chain interventions by using mixed methods, while Box 9 shows a quasi-experimental design on the topic of local energy.

Box 7. Example of reducing bias and systematic transparency

'Systematic analysis and synthesis in qualitative evaluation. Evaluation of the Oxfam Novib programme in Burundi (2005-2009) – Ferko Bodnar.

In this programme evaluation the search was for generic programme outcomes for each of the strategic objectives (SO) of Oxfam Novib. For each SO a results chain from intervention to impact was reconstructed. This resulted in an evaluation matrix with evaluation questions before going into the field. Information was sourced wider than the involved project staff and targeted beneficiaries. Besides focus group discussions, a survey was done among 293 randomly selected households (targeted and (many) non-targeted households). In discussions and survey, we first asked for changes per theme at impact level before discussing the causes and the possible contribution by projects.

Since there were not a lot of hard data around results, one of the team's concerns was how to deal with different opinions in a systematic and transparent way, so that opinion holders recognise their own opinion, understand the opinions of others, and understand how an overall conclusion is drawn from these different opinions. The different opinions were presented in disaggregated form in tables, based on the 'evaluation matrix'. One table was made per evaluation question, with a few judgment criteria. The different statements and opinions were inserted, indicating from which group these opinions came. These indications summarized in a way the findings: + for positive – confirming the judgment criterion, - for negative, in 5 narrow columns next to the statements. In my case I distinguished targeted people (rural households), non-targeted people, involved project staff, other staff not involved, and government. This overview showed where opinions converged or contradicted. This way, the answer to the evaluation question, at the bottom of the table, becomes a logical, almost unavoidable conclusion. In the restitution meeting, people could still add or reinterpret statements and opinions, but the logic of how to come to conclusions could not be dominated by one particular stakeholder.

The systematic presentation of different opinions facilitated drawing (tentative) conclusions and reduced evaluator bias. This was done just before the restitution / synthesis workshop.

The expectations underpinning this communication were: The transparent presentation and synthesis makes the conclusions more acceptable, and thus more useful, for the different opinion holders. Also, partner organisations, often involved in just one project, appreciated the overview and focus on higher level programme outcome and impact. Some mentioned that they would like to repeat such an exercise at the programme planning stage.

The design gave room for a critical view on the programme achievements by different forms of validation and triangulation. This in turn resulted in stronger recommendations, and meeting the IOB evaluation assessment criteria. Unfortunately, ON did not link the planning of a new 5-year plan for Burundi 2011-2015 to the outcomes of the evaluation. It would also have been good to invite other donors and external experts in the planning process. See more info in appendix 2.

Box 8. Example of mixed methods

'How to evaluate the ultimate poverty impact of value chain interventions? Mixed methods design for attributing indirect interventions to farmers' income' – Gideon Krusemann and Shovan Chakraborty

In Bangladesh, Swisscontact Katalyst introduced maize cultivation in 2004. Maize was expected to provide additional incomes for particularly marginalized farmers. Katalyst wanted to record and understand its impact properly. The first IE was conducted in 2011. Additionally, Katalyst commissioned an external study to experts of LEI and Opportunities Unlimited.

The central methodological challenge of this IE is typical for indirect, value chain intervention. Donors want to see impact at poverty level but interventions are geared to change market systems around poor people. Hence, poverty impacts are indirect and often take 5 to 10 years to become visible in indicators at household level.

New methods need to be developed that can balance between (A) a realistic understanding of how indirect interventions contribute to longer-term impacts pathways and (B) robust conclusions on impact, right after interventions end. This IE developed such new method.

To assess the impact of these activities, Katalyst maintains an M&E system in compliance with the DCED-standard, which is based on results chains. In the course of and after an intervention, indicators for each box of the results chain are assessed by enquiries of service/input providers and beneficiaries. It mixes small n with large n approaches. The IE was expected to identify the people that benefitted from by Katalyst's interventions, how they benefitted, by how much and why.

The standard methods for this IE were found inappropriate. RCT, for example, could not be applied because interventions were in a final stage and ex-post 'treated' and 'untreated' farmers could not be singled out. A mixed methods approach was designed, including surveys (large n) and in-depth interviews with treated and untreated respondents from various levels of the value chains (small n). Validation of information collected through multiple sources was a primary duty. The design of the IE was based on rigorous impact logic, linking the various interventions to the expected ultimate impact.

To communicate results and findings different communication methods were used: a workshop with local stakeholders, publication of the impact pathways, publication of a paper and sharing of methodological choices in conferences.

The IE is being utilised ex-ante and ex-post. Ex-post for the project and donors, to validate impact claims. Ex-ante to improve on future intervention design (improve aid effectiveness), the M&E system in place (get better utilization of information generated) and on impact expectations by donors (enhance more realistic expectations on what impact can be expected).

See appendix 2 for more information

Box 9. Example of quasi-experimental IE

'Quasi-experimental IE of Rwanda's National Domestic Biogas' – Willem Heemskerk

The impact evaluation was commissioned by the Dutch Policy and Operations Evaluation and Netherlands Development Organisation SNV. The field studies were implemented by the International Institute of Social Studies, Erasmus University Rotterdam.

The main purpose of the study was to identify the impact of access to biogas provided through a Dutch funded Rwandan

government programme about biogas, within a larger group of projects on renewable energy. Scientific evaluations on biogas are extremely scarce.

The evaluation design was based on the results chain. The impacts have been determined based on a results-chain, of input to impact. The final impact: household convenience and feelings of security; slow down of deforestation; savings to household and income generation; gender equity; improved health; less eutrophication; improved agricultural yields.

The evaluation design was also based on conceptual requirements for applying rigorous methods, using mixed methods. Financial restrictions had an impact on i) the sample size, ii) the choice of a cross-sectional method (and not a double survey) and iii) the use of surveys only (and not additional objectively verifiable methods). In the biogas study, the survey was sufficiently large (a sample of 600) to produce robust data. The evaluation team in this case selected 300 households with digester, and 300 similar households without digester to evaluate effects (propensity score matching). See appendix 2 for more info

4.5.2 Output group discussions II: Implications for utilisation

Use of results and or process is influenced by multiple factors other than design, including the match between the right approach for the right stakeholders asking the right questions. Design-related factors also influence utilisation. As seen in chapter 3, a key concern is how to improve the quality of impact evaluation and the use of impact evaluations – unused evaluations are lower quality than those that are used.

Engage stakeholders in the IE process

The involvement of stakeholders in the design leads to more empowerment. Evaluation is learning by itself. During the process of evaluation, learning already takes place. In one case findings were rejected because stakeholders were not involved in the process so it did not address their interest. Another case mentioned how important it is to involve funders in design, as they may reject findings based on a design (e.g. without control groups when that might be a valid option) One of the groups encouraged thinking about 'how to engage stakeholders in the data collection, so they become 'champions in change'.

Engage beneficiaries in design

Although it might not be easy to access the IE team, intended beneficiaries are often not involved in the design of the evaluations, whilst this could enhance utility of the IE. Box 5 shows an example of how joint sense making can be part of the design.

Deal with resistance to change

In spite of good preparation and design, resistance to change is human. One of the participants said: "It is very difficult to implement recommendations though, people do not always want to learn because they may learn that mistakes were made. They do not wish to admit failure and do not want to change things." Incentives to use results might counteract the resistance.

Increase credibility and use by design

Make sure that the design focuses on the key questions stakeholders have. Participation may increase use in a broader group - stakeholders need to identify themselves within the process.

During the case study sessions and the subsequent group discussions, the following question was discussed: 'what choices were made – explicitly or implicitly – in communication and how do these influence utility'? Several key aspects were identified by the groups on both days as important when considering communication, which are discussed further in the sections below.

Communication from the onset.

Communication should be an integrated part of the design of an IE process. Communicating the results of an impact evaluation is not the last step in an evaluation process. Nor is it only about 'communicating' one-way of reported findings. From the start, expectations around impacts need to be articulated and managed, instead of simply offloading results on the client/user. At the onset, discuss how different groups expect the results to be shared (or even generated) and how they intend to use the findings. Clarity around audiences is key, especially identifying the primary intended users at different levels, with their specific information needs and preferred communication style and mode.

Audience-specific reporting.

The level of utilization will depend on the extent to which the communication is tailored to the specific audience(s). This requires clarity about who the audiences are. It might also be (re)commendable to 'modify' information according to the group that is meant to be informed ('tailor-made'). Obviously language should be understandable for the audience. Different communication methods may be needed for different stakeholders. Apart from publishing a final evaluation report there are other ways to distribute findings, like newsletters or websites and stories and visuals can greatly enhance understanding. Organising seminars on lessons learned and writeshops can also be useful.

Often, an evaluator is not involved in all communication. However, discussing the IE with the evaluator helps to understand and own the findings, conclusions and recommendations. It also helps the evaluator to understand the context in a deeper way. Furthermore, it is important to decide who disseminates results to different levels. For example, a partner organization may communicate to their end users, while an evaluator may only communicate with the partner organization. Exceptions exist like having organisation stakeholder meetings at the end of an IE to validate and communicate key findings. Utilisation might improve by reducing the evaluation report to policy briefs for policy makers, while keeping the full evaluation report as a resource. The ToR negotiation process should cover decisions around audience-specific reporting.

Accessibility of data and knowledge brokering.

IE requires highly skilled evaluators and the utilisation of IE requires access to findings, both in terms of logistics as well as in understanding. Interpretation of findings for different audiences in many ways emerged from the discussion as a critical contribution to utilization.

One of the basic issues is the language, especially in situations where native speakers and donors do not speak the same language, and reports are required in different languages, including visualisations. It is not only specialised, technical language, but also the way in which language is used. Different ways of writing and the level of detail of reporting influences use by different people/organisations. In the panel discussion, Don Seville indicated a need to have knowledge brokering so as to interpret evaluation reports. Some of the working groups indicated that the evaluation capacity like understanding methodological discussions to interpret data is often weak in all parts of organisations.

⁹ The information in this section is based on the group discussions during the two days of the conference.

Confidentiality versus transparency.

Every evaluation requires negotiating the politics around explaining and sharing findings. For example, when stakes around future funding are high, there is a potential conflict of interests between implementer and commissioner and therefore other information and communication needs. Even when the intention to inform beneficiaries/end users is present, results might be considered sensitive information and evaluators might be tied to confidentiality clauses. Box 10 shows an example of impact evaluation in business settings. The notion of 'failure' or 'lack of success' is difficult for most organisations, so it is important to handle transparency diplomatically. Also, an easy defence when findings are not positive is to blame the methodology. Therefore, investing in clear explanations of the methodology is worth the effort. Implementers may not like to share negative effects to different funders. Even if the evaluation was done very well. Box 11 shows also that in spite of a protocol, when the time has come, transparency is not always wanted for.

Box 10. Example of impact evaluation in business

'Methods for impact assessment of certification schemes' -Verina Ingram, Giel Ton and Yuca de Waardt

The case is based on learning about impact studies of corporate sustainability schemes, particularly the certification of tropical commodities (tea, coffee, cocoa, cotton). It involves a number of studies with and for: UTZ, ISEAL Alliance, Dutch Ministry of Foreign Affairs, Dutch Ministry of Economic Affairs, Agriculture and Innovation, commodity processing companies, Solidaridad/WAFF and the Dutch Sustainable Trade Initiative.

The design constraints were significant: too short a timeframe to be feasible, no baseline data, reducing the initial list of required 61 indicators to a more manageable 15-20, multiple participating organisations with differing and sometimes conflicting ideas entering at different stages of the IE process, and more. The study had to be realistic about attribution claims. They learned the importance of building a plausible story line for the contribution to the ultimate outcomes. The researchers made use of existing quantitative information to reflect on the impact logic, and also collected qualitative data to both support and challenge the overall impact storyline.

In terms of communication, the researchers faced a series of challenges, including the tricky politics of conveying 'bad news' to a broader audience for different certifying agencies. Some of these agencies vie for the same markets. So how does one deal with confrontational results? Confidentiality vies with transparency – both of which are important to create a safe learning process. Although there was no communication plan, the engagement of the client organisations in discussions en route has had 'process results', enabling some of the organisations to reconsider aspects of their theory of change, for example.

See appendix 2 for more information.

Multiple accountability

Groups identified upward, downward, vertical and horizontal communication key to being accountable. Participants considered it problematic that the results are often communicated mainly to funders, to a lesser extent to partner organisations, but not communicated to intended beneficiaries. One of the participants in the Burundi case discussions said: "It's becoming a concern that findings do not get communicated to the villages – this must happen more."

Box 11. Example of an evaluation protocol on sensitive issues

'Balas "Are the children better off" - Balasahyoga Concurrent Evaluation (India) – Pam Baatsen

The evaluation executed by the Dutch Royal Tropical Institute (KIT) centred around a programme with an intervention model to improve the quality of life of children infected and affected by HIV, and their households. Appendix III shows the details of the design. What made the evaluation also very interesting was the establishment of an evaluation protocol and an evaluation core group. The evaluation protocol mentioned a number of guiding principles, including that credit would be given to all those involved in the evaluation, and that baseline and end-line evaluation findings would be shared with the project as soon as they would become available so that these findings could help strengthen the programme. This evaluation core group was composed of representatives of relevant stakeholders. This all to facilitate communication, and increase buy-in to the evaluation and its results.

On the basis of the first round of evaluation findings, the evaluators, besides an elaborate evaluation report, prepared policy briefs that could be shared with stakeholders by the intervention INGOs. However, these were not shared, as the INGOs were disappointed with the nil-finding.

The findings were shared with everyone who had been involved in the project, also at district level through a Reflection, learning and action workshop at the end of the evaluation.

The IE has influenced the programme itself, to for instance better define the role of the outreach workers. However, the final results, have not been shared with a large audience, due to the sensitivity of the findings, in spite of earlier agreements on the dissemination of the results.

For more details see appendix 2.

Timing is important

Users will decide what to use, and how to use the findings. Of course, timing of relevant information is key. Timing of when IE results come out in relation to planning for next or other interventions where results can be applied.

Credibility of evaluations

Groups discussed that utilisation is influenced by the credibility of evaluations. And that includes: the reputation of the evaluator/institution; personal and professional attitude and involvement with the project; the audience and the presenter matter; understanding of the context and owning the recommendations; data analysis together with the stakeholders (ownership and validation of findings).

Translation of findings to actionable recommendations

The extent to which the IE connects to real issues and addresses what is relevant and important to stakeholders requires an ability to connect to the on-the-ground realities. Having well-targeted findings for those who need to act upon the findings as well as thinking through the incentives to act on these findings is crucial.

All in all thinking through the communication process, what to communicate and to who and how this translates into action is crucial in enhancing the utility of an IE.

6 Conclusions

So, what can we conclude about what we can do to improve utilisation of impact evaluations by means of design and communication?

Impact evaluation

There are different definitions of impact and impact evaluation (IE), and these have different implications for design, implementation and utilisation of IE. Impact evaluation should not be confused with any single design, methodology or philosophy. Good quality is a precondition for the use of evaluations. Quality of impact evaluations can for example be improved by broadening the range of designs. There are a number of mainstream research and scientific designs that need to be brought in to broaden the IE portfolio.

Design

One way to improve the quality of IEs is to improve the design. This starts with readiness to invest resources in the design of IEs. The determinants of design choice must be the questions we want answers to and the types/attributes of programmes and policies we should learn from, aligned with an available repertoire of designs. It is important to understand the different types of causal inference, i.e. ways of looking at causality so as to make the right choices in design. A good impact evaluation design should include clear communication and utilisation strategies for different stakeholders.

Communication

A thorough communication strategy makes explicit the different users of the evaluation and caters for their preferred communication method and style. When done well, communication can serve as capacity development and enhance use. 'Translating' and interpreting the findings for the user might be an extra role for knowledge brokers. Multiple accountability is needed, and special attention for downward accountability. Communication is not a neutral activity: balancing rigorous evidence with politics is an art in itself.

Utilisation

Utilisation is the first programme evaluation standard, yet it receives little attention in IEs. Uptake needs explicit thought from the onset and only then can IEs be useful. This conference would like to encourage those involved in IE to more thoroughly consider design and communication as central for utilisation.

The conference recommends the joint development of evaluation questions and joint sense making by evaluators, commissioners and implementing agencies. This will increase relevance and usefulness of the IE process and findings for different users of the IE. It is also advisable to translate the findings into actionable recommendations targeted at these different audiences.

Finally, the conference also called for responsible use of evaluation findings for those who have the capacity to effect change. Those for whom IEs are intended, including policymakers and programme managers, have a responsibility to follow up promptly on lessons learned otherwise the IE is a waste of time and resources.

References and resources

- Appleton-Dyer, S., Clinton, J., Carswell, P., & McNeill, R. (2012). Understanding evaluation influence within public sector partnerships: A conceptual model. *American Journal of Evaluation*, 33(4), 532-546. doi: 10.1177/1098214012447672
- Bamberger, M. (2012). Introduction to mixed methods in impact evaluation. *InterAction Impact Evaluation Guidance Notes*. Retrieved from <http://www.interaction.org/document/guidance-note-3-introduction-mixed-methods-impact-evaluation>
- Bamberger, M., Rao, V., & Woolcock, M. (2010). Using mixed methods in monitoring and evaluation: Experiences from international development. *World Bank Policy Research Working Paper* (5245).
- Banerjee, A. V., & Duflo, E. (2011). *Poor economics. A radical rethinking of the way to fight global poverty*. Penguin Group.
- Barrett, C. B., & Carter, M. R. (2010). The power and pitfalls of experiments in development economics: Some non-random reflections. *Applied Economic Perspectives and Policy*, 32(4), 515-548. doi: 10.1093/aapp/ppq023
- Befani, B. (2012). Models of causality and causal inference. Retrieved from <http://mande.co.uk/blog/wp-content/uploads/2012/07/2012-Causal-Inference-BB-February-26.pdf>
- Big Push Forward (2013) Programme 'The Politics of Evidence' Conference 23-24 April, 2013
Institute of Development Studies, Brighton, UK <http://bigpushforward.net/the-politics-of-evidence>
- Bonbright, D. (2012). Use of impact evaluation results. *InterAction Impact Evaluation Guidance Notes*. Retrieved from <http://www.interaction.org/document/guidance-note-4-use-impact-evaluation-results>
- Campos, F., Coville, A., Fernandes, A. M., Goldstein, M., & McKenzie, D. (2012). Learning from the experiments that never happened. Lessons from trying to conduct randomized evaluations of matching grant programs in Africa. [working paper]. *World Bank Policy Research Working Paper*, 6296.
- Car, E. (2013) Causality isn't the same as explanation: why development needs more of the latter. Blog <http://www.edwardrcarr.com/opentheechochamber/2013/02/04/causality-isnt-the-same-as-causal-mechanisms-why-development-needs-more-research-into-the-latter/>
- Cartwright, N and J. Hardie (2012) Evidence-based policy: a practical guide to doing it better. Oxford University Press 224pp, Published 27 September 2012
- Chambers, R. (2007). Who counts? The quiet revolution of participation and numbers IDS Working Paper No. 296 Retrieved from <http://www.ids.ac.uk/files/Wp296.pdf>
- Chambers, R. (2009). Making the poor count: Using participatory methods for impact evaluation. In: Designing impact evaluations: Different perspectives, 3ie Working paper series 4 (pp. 4-7). Retrieved from <http://www.3ieimpact.org/en/evaluation/working-papers/working-paper-4/>.
- Chambers, R., Karlan, D., Ravallion, M., & Rogers, P. (2009). Designing impact evaluations: Different perspectives. *3ie Working paper series*(4).
- Collier, D. (2011). Understanding process tracing. *PS: Political Science & Politics*, 44(04), 823-830. doi: doi:10.1017/S1049096511001429
- Contandriopoulos, D., & Brousselle, A. (2012). Evaluation models and evaluation use. *Evaluation*, 18(1), 61-77. doi: 10.1177/1356389011430371

- Davies, R and J. Dart (2005) The 'Most Significant Change' (MSC) Technique. A Guide to Its Use. <http://www.mande.co.uk/docs/MSCGuide.pdf>
- Deaton, A. S. (2009). Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development. [working paper]. *NBER Working Paper Series*(14690).
- Delahais, T., & Toulemonde, J. (2012). Applying contribution analysis: Lessons from five years of practice. *Evaluation, 18*(3), 281-293. doi: 10.1177/1356389012450810
- Diaz, F. J. M. (2012). Small is useful: The empirical approach to improve Poor Economics. [Book review: Poor economics. A radical rethinking of the way to fight global poverty]. *European Journal of Development Research, 24*(5), 832-834. doi: 10.1057/ejdr.2012.28
- du Toit, A. (2012). Making sense of 'evidence'. Notes on discursive politics of research and pro-poor policy making. *PLAAS Working Paper 21*. Retrieved from <http://www.plaas.org.za/plaas-publication/dutoit-ebpm-2012>
- Ebrahim, A. S., & Rangan, V. K. (2010). The limits of nonprofit impact: A contingency framework for measuring social performance [working paper]. *Harvard Business School General Management Unit Working Paper No. 10-099*.
- Ebrahim, A.S. (2013). Let's be Realistic about measuring impact. Blog in Harvard Business Review, March 13, 2013 <http://blogs.hbr.org/hbsfaculty/2013/03/lets-be-realistic-about-measur.html>
- Eckerd, A., & Moulton, S. (2010). Heterogeneous roles and heterogeneous practices: Understanding the adoption and uses of nonprofit performance evaluations. *American Journal of Evaluation, 32*(1), 98-117. doi: 10.1177/1098214010381780
- EES (2007) European Evaluation Society Statement on methodological diversity
- Elbers, C., Gunning, J. W., & de Hoop, K. (2009). Assessing sector-wide programs with statistical impact evaluation: A methodological proposal. *World Development, 37*(2), 513-520.
- Eyben, R. (2013) Uncovering the politics of 'evidence' and 'results'. A framing paper for development practitioners. Paper prepared for the Big Push Forward Conference. <http://bigpushforward.net/wp-content/uploads/2011/01/Uncovering-the-Politics-of-Evidence-and-Results-by-Rosalind-Eyben.pdf>
- Forss, K., Marra, M., & Schwartz, R. (Eds.). (2011). *Evaluating the complex. Attribution, contribution, and beyond* (Vol. 18): Transaction Publishers.
- Garbarino, S., & Holland, J. (2009). Quantitative and qualitative methods in impact evaluation and measuring results *GSDRC Emerging Issues Papers*. Birmingham, UK: University of Birmingham.
- Gauck, J. (2011). Measuring impact on the immeasurable? Methodological challenges in evaluating democracy and governance aid. *Paper prepared for the 2011 Annual Meeting of the American Political Science Association Seattle, September 4, 2011*. Retrieved from <http://ssrn.com/abstract=1902105>
- Greeley, M., Lucas, H., & Chai, J. (Eds.). (2013). *Special issue: Real time monitoring for the most vulnerable* (Vol. 44-2): Blackwell Publishing Ltd.
- Greeley, M., Lucas, H., Chai, J., & Cummins, M. (2013). Introduction: Real time monitoring for the most vulnerable – Investing in common guidance for equity and quality. *IDS Bulletin, 44*(2), 1-14. doi: 10.1111/1759-5436.12012
- Green, D (2013) So what do I take away from the great evidence debate? Final thoughts (for now) , february 2013 www.oxfamblogs.org/fp2p/?p=13590

- Guijt, I., Brouwers, J., Kusters, C., Prins, E., & Zeynalova, B. (2010). Evaluation revisited. Conference report. Paper presented at the Conference: Evaluation revisited. Improving the quality of evaluative practice by embracing complexity, Wageningen, the Netherlands.
- Guijt, I. (2013) 'The politics of evidence' in my daily working life. Blog March 30, for the Big Push Forward conference. <http://bigpushforward.net/archives/2177>
- Hivos Knowledge Programme. Retrieved from Hivos Theory of Change (ToC) Resource Portal, website: <http://www.hivos.net/Hivos-Knowledge-Programme/Themes/Theory-of-Change>
- Holland, J. (Ed.). (2012). *Who counts? The power of participatory statistics*. Warwickshire, UK: Practical Action Publishing.
- Holma, K., & Kontinen, T. (2011). Realistic evaluation as an avenue to learning for development NGOs. *Evaluation, 17*(2), 181-192. doi: 10.1177/1356389011401613
- Hughes, K., & Hutchings, C. (2011). Can we obtain the required rigour without randomisation? Oxfam GB's non-experimental Global Performance Framework. [working paper]. *3ie Working paper series*(13).
- Hulme, D. (2000). Impact assessment methodologies for microfinance: Theory, experience and better practice. *World Development, 28*(1), 79-98. doi: doi: 10.1016/S0305-750X(99)00119-9
- IDS conference (26-27 March 2013) "Impact, Innovation and Learning: Towards a Research and Practice Agenda for the Future" <http://www.ids.ac.uk/events/impact-innovation-and-learning-towards-a-research-and-practice-agenda-for-the-future>
- Jackson, E. T. (2013). Interrogating the theory of change: Evaluating impact investing where it matters most. [Journal of Sustainable Finance & Investment]. *Journal of Sustainable Finance & Investment, 1*-16. doi: 10.1080/20430795.2013.776257
- JCSEE (2011) The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users'. [Donald B. Yarbrough](#), [Lyn M. Shulha](#), [Rodney K. Hopson](#), [Flora A. Caruthers](#), 2011. 3rd Edition, Sage Publications, USA.
- Karlan, D. (2009). Thoughts on randomized trials for evaluation of development: Presentation to the Cairo Evaluation Clinic. In: Designing impact evaluations: Different perspectives, 3ie Working paper series 4 (pp. 8-23). Retrieved from <http://www.3ieimpact.org/en/evaluation/working-papers/working-paper-4/>.
- Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). *Handbook on impact evaluation: Quantitative methods and practices*. Washington, DC: World Bank Publications.
- Kusters, C., & McGregor, C. (2011). *Managing for Impact*. Wageningen, The Netherlands: Wageningen UR Centre for Development Innovation.
- Kusters, C., van Vugt, S., Wigboldus, S., Williams, B., & Woddhill, J. (2011). *Making evaluations matter: A practical guide for evaluators*. Wageningen, The Netherlands: Centre for Development Innovation, Wageningen University & Research centre.
- Leeuw, F. L. (2012). Linking theory-based evaluation and contribution analysis: Three problems and a few solutions. *Evaluation, 18*(3), 348-363. doi: 10.1177/1356389012452051
- Lemire, S. T., Nielsen, S. B., & Dybdal, L. (2012). Making contribution analysis work: A practical framework for handling influencing factors and alternative explanations. *Evaluation, 18*(3), 294-309. doi: 10.1177/1356389012450654
- Lennie, J., Tacchi, J., Koirala, B., Wilmore, M., & Skuse, A. (2011). Equal access participatory monitoring and evaluation toolkit. Retrieved from http://betterevaluation.org/toolkits/equal_access_participatory_monitoring

- Levine, C. J., & Chapoy, C. (2013). Building on what works: Commitment to evaluation (c2e) indicator. [working paper]. *3ie Working paper series*(18).
- Loevinsohn, M. (2013). Natural experiments: An underappreciated evaluation resource? *Centre for Development Impact, Practice Paper. Innovation and learning in impact evaluation, 2*.
- Mansfield, W., & Grunewald, P. (2013). The use of indicators for the monitoring and evaluation of knowledge management and knowledge brokering in international development. Report of a workshop held at the Institute for Development Studies 8th March 2013: Institute for Development Studies (IDS).
- Mayne, J. (2007). Evaluation for accountability: Reality or myth? In M.-L. Bemelmans-Videc, J. Lonsdale & B. Perrin (Eds.), *Making Accountability Work: Dilemmas for Evaluation and for Audit* (pp. 63-84). New Brunswick, NJ Transaction Publishers.
- Mayne, J. (2008). Contribution analysis: An approach to exploring cause and effect. *ILAC Brief No. 16. The Institutional Learning and Change Initiative*.
- Mayne, J. (2011). Contribution analysis: Addressing cause and effect. In K. Forss, M. Marra & R. Schwartz (Eds.), *Evaluating the Complex. Attribution, contribution, and beyond* (Vol. 18, pp. 53-96): Transaction Publishers.
- Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation, 18*(3), 270-280. doi: 10.1177/1356389012451663
- Mayne, J., & Stern, E. (2012, 18 December, final draft). Position paper on impact evaluation of NRM Research Programmes: A broader view. Prepared for the NRM-Impact Community of Practice
- Morton, J, Shaxson, L and J. Greenland (and others) Process Evaluation of the International Initiative for Impact Evaluation (2008-11) 27 September 2012. Triple line consulting and ODI
- Murray, K. (2012). Influencing of policy and public management program: Impact assessment summary report. Reducing poverty by improving public policy and public administration in Bolivia. *Oxfam Evaluation Reports*. Retrieved from <http://policy-practice.oxfam.org.uk/publications/effectiveness-review-influencing-of-policy-and-public-management-bolivia-249816>
- Newman, D et al (2012) <http://www.thevalueweb.org/world-economic-forum/complex-systems/>
http://www3.weforum.org/docs/LA12/WEF_LA12_YGL_GuideInfluencingComplexSystems.pdf
- Ní Ógáin, E., Lumley, T., & Pritchard, D. (2012). Making an impact. Impact measurement among charities and social enterprises in the UK. London, UK: New Philanthropy Capital (NPC).
- Nutley, S., Powell, A., & Davies, H. (2012). What counts as good evidence? Retrieved from <http://www.nesta.org.uk/library/documents/A4UEprovocationpaper2.pdf>
- OECD-DAC (2002). Glossary of Key Terms in Evaluation and Results Based Management Paris: OECD. <http://www.oecd.org/dataoecd/29/21/2754804.pdf>.
- Oliveira, V., & Pinho, P. (2011). Bridging the gap between planning evaluation and programme evaluation: The contribution of the PPR methodology. *Evaluation, 17*(3), 293-307. doi: 10.1177/1356389011411686
- Patton, M. Q. (2008). Advocacy impact evaluation. *Journal of MultiDisciplinary Evaluation, 5*(9), 1-10.
- Patton, M. Q. (2010). *Developmental evaluation. Applying complexity concepts to enhance innovation and use*. New York, NY: The Guilford Press.
- Patton, M. Q. (2012). A utilization-focused approach to contribution analysis. *Evaluation, 18*(3), 364-377. doi: 10.1177/1356389012449523
- Pawson and Tilley (1997) Realistic evaluation. Sage publications

- Perrin, B. (2012). Linking monitoring and evaluation to impact evaluation. *InterAction Impact Evaluation Guidance Notes*. Retrieved from <http://www.interaction.org/document/guidance-note-2-linking-monitoring-and-evaluation-impact-evaluation>
- Picciotto, R. (2012). Experimentalism and development evaluation: Will the bubble burst? *Evaluation*, 18(2), 213-229. doi: 10.1177/1356389012440915
- Ravallion, M. (1999). The mystery of the vanishing benefits. Ms Speedy Analyst's introduction to evaluation. *World Bank Policy Research Working Paper*(2153).
- Ravallion, M. (2008), Evaluating Anti-Poverty Programs , T.P. Schultz and John Strauss Handbook of Development Economics Volume 4. In: White, Howard (2009), Some reflections on current debates in impact evaluation. International Initiative for Impact Evaluation, Working paper 1.
- Ravallion, M. (2009a). Evaluating three stylized interventions. In: Designing impact evaluations: Different perspectives, 3ie Working paper series 4 (pp. 14-23). Retrieved from <http://www.3ieimpact.org/en/evaluation/working-papers/working-paper-4/>.
- Ravallion, M. (2009b). Should the randomistas rule? *The Economists' Voice*, 6(2), 1-5. doi: 10.2202/1553-3832.1368
- Ravallion, M. (2011). On the implications of essential heterogeneity for estimating causal impacts using social experiments. *Policy Research Working Papers*, 5804. Retrieved from <http://elibrary.worldbank.org/content/workingpaper/10.1596/1813-9450-5804> doi:10.1596/1813-9450-5804
- Ravallion, M. (2012). Fighting poverty one experiment at a time: A review of Abhijit Banerjee and Esther Duflo's *Poor economics: A radical rethinking of the way to fight global poverty*. *Journal of Economic Literature*, 50(1), 103-114.
- Reddy, S. G. (2013). Randomise this! On Poor Economics. *Review of Agrarian Studies*, 2(2), 60-73.
- Richards, P. (2011). A systematic approach to cultural explanations of war: Tracing causal processes in two West African insurgencies. [Ethnicity and Ethnic Strife]. *World Development*, 39(2), 212-220. doi: <http://dx.doi.org/10.1016/j.worlddev.2009.11.030>
- Rihoux, B. and C. Ragin (2009). 'Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques'. Sage publications
- Rodrik, D. (2008). The new development economics: We shall experiment, but how shall we learn? *Faculty Research Working Papers Series*. Harvard University, John F. Kennedy School of Government.
- Rogers, P. (2007). Theory-based evaluation: Reflections ten years on. doi: DOI: 10.1002/ev.225
- Rogers, P. (2009). Matching impact evaluation design to the nature of the intervention and the purpose of the evaluation. In: Designing impact evaluations: Different perspectives, 3ie Working paper series 4 (pp. 24-33). Retrieved from <http://www.3ieimpact.org/en/evaluation/working-papers/working-paper-4/>.
- Rogers, P. (2012). Introduction to impact evaluation. *InterAction Impact Evaluation Guidance Notes*. Retrieved from <http://www.interaction.org/document/introduction-impact-evaluation>
- Rosenzweig, M. R. (2012). Thinking small: A review of *Poor economics: A radical rethinking of the way to fight global poverty* by Abhijit Banerjee and Esther Duflo. *Journal of Economic Literature*, 50(1), 115-127. doi: doi:10.1257/jel.50.1.115
- Saunders, M. (2012). The use and usability of evaluation outputs: A social practice approach. *Evaluation*, 18(4), 421-436. doi: 10.1177/1356389012459113

- Scheu, L., & Baldasare, A. (2013). Linda Scheu and Angela Baldasare on Using good presentation principles to increase potential impact. Retrieved from AEA365 | A Tip-a-Day by and for Evaluators, website: <http://aea365.org/blog/?p=8183>
- Schmidt, B. (2013). Divining a vision for Markets for Good. Retrieved from Alliance Magazine, website: http://www.alliancemagazine.org/en/content/divining-vision-markets-good?utm_source=PR+list&utm_campaign=4547eaecb0-1302PRemail2&utm_medium=email
- Shaffer, P. (2011). Against excessive rhetoric in impact assessment: Overstating the case for randomised controlled experiments. [The Journal of Development Studies]. *The Journal of Development Studies*, 47(11), 1619-1635. doi: 10.1080/00220388.2010.514331
- Shutt, C., & McGee, R. (2013a). Improving the evaluability of INGO empowerment and accountability programmes. *Centre for Development Impact, Practice Paper. Innovation and learning in impact evaluation, 1*.
- Shutt, C., & McGee, R. (2013b). Tools and guidelines for improving the evaluability of INGO empowerment and accountability programmes *Centre for Development Impact, Practice Paper Annex, 1*.
- Stame, N. (2010). What doesn't work? Three failures, many answers. *Evaluation, 16*(4), 371-387. doi: 10.1177/1356389010381914
- Stern, E., Altinger, L., Feinstein, O., Marañón, M., Schultz, N.-S., & Steen Nielsen, N. (2008). Thematic study on the Paris Declaration, aid effectiveness and development effectiveness. Evaluation of the Paris Declaration: Ministry of Foreign Affairs of Denmark.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). DFID Working Paper 38. Broadening the range of designs and methods for impact evaluations (pp. vi + 24 pp.). London, UK: DFID.
- Székely, M. (2013). Promoting commitment to evaluate. *3ie Working paper series, 19*.
- Ton, G. (2012). The mixing of methods: A three-step process for improving rigour in impact evaluations. *Evaluation, 18*(1), 5-25. doi: 10.1177/1356389011431506
- Unit, G. M. a. E. (2012). Measuring impacts using experimental designs – and their application at GIZ.
- Vaessen, J., & Raimondo, E. (2012). Making sense of impact: A methodological framework for assessing the impact of prizes. *Evaluation, 18*(3), 330-347. doi: 10.1177/1356389012450655
- Weyrauch, V., & Langou, G. D. (2011). Sound expectations: From impact evaluations to policy change. *3ie Working paper series*(12).
- White, H. (2006). Impact evaluation: The experience of the Independent Evaluation Group of the World Bank. Washington, DC: IEG/World Bank.
- White, H. (2009a). Some reflections on current debates in impact evaluation. *3ie Working paper series*(1).
- White, H. (2009b). Theory-based impact evaluation: Principles and practice. *3ie Working paper series*(3).
- White, H. (2013a). An introduction to the use of randomised control trials to evaluate development interventions. *Journal of Development Effectiveness, 5*(1), 30-49. doi: 10.1080/19439342.2013.764652
- White, H. (2013b). An introduction to the use of randomized control trials to evaluate development interventions. *3ie Working paper series*(9).
- White, H. (2013c) Public lecture What we have learned from 3ie's experience in evidence-based development. <http://www.3ieimpact.org/en/events/3ie-conferences-and-workshops/3ie-public-lecture/>

- White, H., & Philips, D. (2012). Addressing attributions of cause and effect in small n impact evaluations: towards an integrated framework. *3ie Working paper series*, (15). Retrieved from <http://www.3ieimpact.org/en/evaluation/working-papers/working-paper-15/>
- Whitty, B. (2013) Of devils, details and multiple perceptions. Blog <http://bigpushforward.net/archives/2130>
- Williams, B., & Imam, I. (Eds.). (2007). *Systems concepts in evaluation: An expert anthology*.
- Wimbush, E., Montague, S., & Mulherin, T. (2012). Applications of contribution analysis to outcome planning and impact evaluation. *Evaluation*, 18(3), 310-329. doi: 10.1177/1356389012452052
- Wood (2011) Evaluation of the Paris Declaration's impacts on aid effectiveness <http://www.oecd.org/development/evaluation/dcdndep/40888983.pdf>
- Woolcock, M. (2009). Toward a plurality of methods in project evaluation: a contextualised approach to understanding impact trajectories and efficacy. *Journal of Development Effectiveness*, 1(1), 1-14. doi: 10.1080/19439340902727719
- Zerfu, E., & Kebede, S. W. (2013). Filling the learning gap in program implementation using participatory monitoring and evaluation. Lessons from farmer field schools in Zanzibar. *IFPRI Discussion Paper 01256*.

Appendix 1 – Biographies keynote speakers

Keynote Dr. Irene Guijt

“Impact Evaluation for International Development: Why we need it and why it is hard.”

Dr. Irene Guijt is an irrigation engineer by training, specialised in organizational learning and evaluation for transformational development. She undertakes research, system design, facilitation and advisory work on learning-oriented knowledge processes in international development, in particular known for innovative thinking on monitoring, evaluation and organizational learning in rural development. Her PhD, *Seeking Surprise*, was on the grounded theory of monitoring for learning in ‘messy partnerships’ and involved a study of Brazilian efforts for systemic agricultural change. (Guijt, 2008)

In 2000, she coordinated the writing of IFAD’s ‘Managing for Impact’ guide (Guijt & Woodhill, 2002). She has been team leader on several large evaluations on civil society participation and North-South NGO relations, and coordinated the M&E processes of an 11 country action research network in Latin America funded by IDRC, New Zealand Aid and IFAD ([Rural Territorial Development Program](#)). Recent work includes a series of pilot processes with SenseMaker® in East Africa and Asia on issues ranging from girls’ empowerment to agricultural value chains and water service delivery, which she is writing up as part of a Visitorship at the Department of Anthropology at the Australian National University.

She is active in global capacity building through [BetterEvaluation](#) and in the politics of evaluation as co-convenor of the [Big Push Forward](#). Methodologically she has recently started work with IFAD on developing their Participatory Impact Assessment and Learning Approach, as well as with ODI on the MethodsLab which is a collaboration with AusAid. Her core interests focus on learning that aids transformational development and its complexity, and within that: how sensemaking occurs; citizen’s voice and how it can inform policy and practice; the theory of monitoring; values and how these are espoused and practiced; and narratives for evidence-informed decisions.

Publications

- Guijt, I. 2010. Exploding the Myth of Incompatibility between Accountability and Learning. In: Ubels, J., N. Acquaye-Baddoo and A. Fowler. *Capacity Development in Practice*. SNV, The Hague. pgs 277-291.
- Guijt, I. 2010. Rethinking monitoring in a complex messy partnership in Brazil. *Development in Practice* 20 (8).
- Guijt, I. 2008. *Seeking Surprise*. Rethinking monitoring for collective learning in rural resource management. Published PhD dissertation. Wageningen University, Wageningen, The Netherlands.
- Guijt, I. 2008. *Critical Readings on Assessing and Learning for Social Change: A Review*. IDS Development Bibliography 21. Brighton: IDS.
- Guijt, I (Editor). 2007. *Negotiated Learning: Collaborative Monitoring in Resource Management*. Resources for the Future, Washington DC.
- Guijt, I. 2007. *Assessing and Learning for Social Change: A Discussion Paper*. Brighton: IDS.
- Bawden, R., I. Guijt and J. Woodhill. 2007. The Critical Role of Civil Society in Fostering Societal Learning for a Sustainable World. In: *Social Learning Towards a More Sustainable World*. A. Wals (Ed).
- Guijt, I. 2007. The Myth of Community: Implications for Civil Society Organizations and Democratic Governance. In: *Learning Civil Societies: Shifting Contexts for Democratic Planning and Governance*, edited by Penny Gurstein and Leonora Angeles, University of Toronto Press.

Guijt, I. 2005. Synthesis Report of Dutch CFA Programme Evaluation. "Assessing Civil Society Participation as Supported In-Country by Cordaid, Hivos, Novib and Plan Netherlands 1999-2004". MFP Breed Netwerk, the Netherlands.

Guijt, I and J. Woodhill / IFAD. 2002. Managing for Impact in Rural Development: A Practical Guide for M&E. 350 p. In English, Spanish, French and Arabic.

Keynote Professor Elliot Stern

“The Problem with ‘Impact Evaluation’. Balancing precision and generalisation in a complex world.”

Prof. Elliot Stern argues that there is nothing wrong with trying to evaluate outcomes and impacts in international development but that we need to do so in a way that suits the complex world we live in. The dominant discourse over recent years has argued for an emphasis on precision – better measurement, greater certainty, less risk and better value for money. Prof. Stern argues that we need to recognise that the kinds of programmes we evaluate nowadays – devolved, partnership based, grappling with complex socio-cultural and material problems like gender equality, governance, peace-building, climate change mitigation - pose new challenges that the evaluator’s toolkit is ill-equipped to deal with. We have to broaden out the range of methods and designs that we use from the narrow range that have monopolised the impact evaluation territory in recent years. There is of course nothing wrong with much of the excellent work that is being done – it is however concentrating on too small a proportion of what we need to evaluate. One reason for this is that by trying to get it precisely right in particular settings we almost inevitably sacrifice the opportunity to generalise beyond that setting. Furthermore shifts in values that have accelerated with globalisation place an obligation on evaluators to work within new ethical and political parameters – after the Paris Declaration, Accra agenda for action and Busan. Who asks the questions inevitably influences the questions that are being asked – and donors no longer have it all their own way. If as appears likely the major users of evaluation in the coming decades will be governments and civil society in developing countries, the debate about evaluation utilisation as well as evaluation design will also inevitably change its character in future.

Prof. Elliot Stern’s presentation builds on a recent major project that he led for DFID, the UK international development department, on Impact Evaluation that resulted in the report:

Stern, E., N. Stame, J. Mayne, K. Forss, R. Davies and B. Befani (2012). Broadening the Range of Designs and Methods for Impact Evaluations, DFID Working Paper, 38. London: DFID. Available at <http://www.dfid.gov.uk/R4D/Output/189575/Default.aspx>

Prof. Elliot Stern is Visiting Professor at the University of Bristol, Emeritus Professor of Evaluation Research at Lancaster University and was previously a Director of the Tavistock Institute in London. He edits the journal *Evaluation: the international journal of research, theory and practice*, the leading European scholarly journal in the field of evaluation. He was a founding Academician of the UK Academy of Social Sciences serving on its Council from 2006-2011; is a past President of the UK and European Evaluation Societies; and was founding President of the International Organisation for Cooperation in Evaluation (IOCE).

Prof. Stern has specialised in three substantive areas in recent years: advising on and designing evaluation systems; methodological innovation in the evaluation of policy outcomes and impacts; and the role of evaluation in promoting partnership working in international development. His work draws heavily on philosophy of science, organisational theory and systems ideas. He is currently on a number of committees - for the UN; the EU (DG REGIO); and UK government - that advise on improving the quality of evidence in policy making.

In 2010/2011 Prof. Stern led the independent evaluation team reviewing the operations and strategy of UNESCO reporting to the agency’s intergovernmental Executive Board. In 2012 he led a team that reviewed potential methods for impact evaluation in international development for the UK Department for International Development. The report from this study has been widely disseminated. He is currently reviewing ‘good practices’ in OECD Committees that contribute to policy impact and learning among OECD member and non-member States

Selected publications:

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

http://www.dfid.gov.uk/r4d/pdf/outputs/misc_infocomm/DFIDWorkingPaper38.pdf

Stern, E.D: *Evaluation Policies in the European Union and its institutions*. Chapter in 'Evaluation policies'. New Directions in Evaluation. Wiley/Josey Bass (2009)

Stern E.D. *et al: Thematic Study on the Paris Declaration, Aid Effectiveness and Development Effectiveness*. (DIIS/OECD) 2008 <http://www.oecd.org/dataoecd/59/28/41807824.pdf>

Stern E.D; *Contextual Challenges for Evaluation Practice*. Chapter 13 in Sage Handbook of Evaluation, Eds. Ian Shaw, Jennifer Greene & Melvin Mark Sage UK (2006)

Stern E.D (editor) *Evaluation Research Methods* Volumes 1-4. Benchmark Series in Research Methods. Sage UK (2005)

Appendix 2 – Summaries of case studies¹⁰

Case 1 Impact measurement and learning in complex collaborative settings, in the case of smallholders' productive water rights in Ethiopia

Case 2 Corporate Strategy Evaluation “Scaling up / Delivering Impact at Scale”

Case 3 Balas “Are the children better off” Balasahyoga Concurrent Evaluation (India)

Case 4 Mixed Methods in Impact Assessment of Value Chain Interventions. A case study on the maize sector in Bangladesh

Case 5 A beneficiary picture of development and change (PADev: holistic and participatory approach to development assessment)

Case 6 Methods for impact assessment of certification schemes

Case 7 Impact Evaluation of Rwanda’s National Domestic Biogas Programme

Case 8 The Most Significant Change method for assessing impact with community and children (India)

Case 9 Systematic analysis and synthesis in qualitative evaluation. A case study evaluation of the Oxfam Novib programme in Burundi (2005-2009)

¹⁰ Powerpoints on the website

Appendix 2

Case 1: Impact measurement and learning in complex collaborative settings, in the case of smallholders' productive water rights in Ethiopia

PROFILE CASE OWNER

Name of case owner:

Adinda van Hemelrijck

Organisation:

Institute for Development Studies

Country:

England

Your current function:

PhD candidate

Independent Advisor

Your role in the Impact Evaluation case:

Evaluation advisor/manager

Other relevant experiences/background:

Advisory and research around systemic and empowering approaches to impact measurement & learning in collaborative settings.

Past 5 years of advisory work with Oxfam America, assisting local actors worldwide with building and managing frameworks/systems for assessing whether their development efforts make any difference in the lives of the poor. Focus on coaching and capacity-building of local stakeholders to collaboratively analyse and influence complex poverty and development issues using validated evidence.

Previous 10 years of research and consultancy work around performance management, evaluation and learning with international research institutes and governmental and non-governmental development and social integration agencies based in the EU.

Specialties: Design of impact measurement and learning systems for rights-based development programs and projects in the field of smallholder agriculture and pastoralism (Indonesia, Cambodia & Vietnam, Ethiopia); indigenous people's livelihoods rights (Central America); gender issues (El Salvador, Indonesia, Ethiopia); basic social services for migrants, nomads and people living in poverty (Belgium, EU).

CASE DESCRIPTION

Title case:

Impact measurement and learning in complex collaborative settings, in the case of smallholders' productive water rights in Ethiopia

Who commissioned the impact evaluation (IE):

OXFAM America

What was the main purpose of the IE:

The case presented is not about a single one-time-off impact evaluation, but an approach to ongoing impact measurement and learning. The purpose of this approach is to generate a collective understanding among the program stakeholders of institutional bottlenecks, technological and infrastructural challenges, conflict and sustainability issues, and community-level organisational and social aspects of water resource development for smallholder productivity. Impact measurement and learning of selective program pilots is expected to generate a greater commitment and responsibility for collective impact among these stakeholders.

Describe the impact that is being evaluated, the intervention that is expected to have led to this impact and the context of the intervention.

The Ethiopian water program aims at enabling vulnerable agriculturalist and agro-pastoralist communities in moisture-stressed areas in Tigray, Amhara and Oromya Regional States to access, use and manage productive water resources in an equitable and sustainable manner, in order to improve their food security and strengthen their rural livelihoods. (Oxfam America, 2009)

Impact is defined and measured, not as a change in a single dependent variable, but as a change pattern or the interactive changes in a core set of interrelated indicators that **together** measure complex systemic change. The identification and analysis of causing mechanisms help explain why observed change patterns occur as they do, and how they are generated, in a particular context. (Befani, 2012, p. 19; Pawson, 2008, pp. 14–15; Stern et al., 2012, pp. 24, 48)

The program's Theory of Change forms the basis for identifying and measuring the core mechanism of change in each particular time frame. In the water program's first phase (2010-2013), the core mechanism is the core proposition of 'co-investment' that involves the development of individual capacity and collective responsibility (based on a tri-partite agreement) of communities, NGOs and local government for the sustainable design, construction and management of small-scale productive watering systems. The ToC visualizes the investments and institutional changes needed to support this. Improvements in smallholders' capacity, awareness and organization, and the expansion of women's influence and leadership, presumably will lead to better governance of the water systems, increased productivity, better access prices in the markets, and thus higher revenues. This will form the basis for reinvestments in the water systems and cooperative business enterprises, which over time will contribute to strengthening livelihoods and food security (Oxfam America, 2009; Van Hemelrijck, 2013).

What expectations underpinned the choices for the design of the IE? Which of these expectations were more robust (valid) and which were problematic (less valid)?

- The question about how impact measurement could support (rather than undermine) transformational development (i.e. systemic change in existing patterns of behaviours and interactions, the power relationships, and the institutional structures) and thus support empowerment.
- The complexity of the problem that the program is trying to tackle = WRD/M = "wicked problem" requiring complex systemic change.
- The degree of complexity of the program and its context (chronicle drought and food insecurity; high vulnerability and low capacity; lack of infrastructure, access to markets and services; high degree of agro-ecological, socio-cultural and political-economic variability; relatively high risk and unpredictable outcomes of investments; multiple interests and potential conflicts from local up to national and even international levels).

What expectations underpinned the choices for the communication of IE findings? Which of these expectations more robust (valid) and which were problematic (less valid)?

The desire to promote dialogue among stakeholders around evidence that may contribute to generating changes in interactions and behaviours.

To what extent has the IE been utilized and why/why not? What influenced this?

(Contandriopoulos & Brousselle, 2012; Forss, Marra, & Schwartz, 2011; "Rigorous thinking" in IE (Patton, 2012)) acknowledges and deals with the politics of impact knowledge and its use, by: 1. critically analysing stakeholder power relations 2. facilitating inclusive IE design and validation (with spaces for meaningful and equal participation) 3. combining/mixing and nesting methods (incl. sampling) in ways that rigorously produce and triangulate various kinds of evidence that are meaningful, rigorous and coherent Ethiopia case: 1, 2 & 3 are built-in into the ongoing impact measurement approach. Rigorous thinking according to Michael Quinn Patton (2008) determines overall IE validity and utility, and implies 'methodological rigor' AND 'competences' (both scientific AND socio-cultural).

Still ongoing impact evaluation.

Appendix 2

Case 2: Corporate Strategy Evaluation “Scaling up / Delivering Impact at Scale”

PROFILE CASE OWNER

Name of case owner:

Claudia Kornahrens

Organisation:

Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ)

Country:

Germany

Your current function:

Monitoring and Evaluation Unit
Head of Section Central Evaluation

Your role in the Impact Evaluation case:

The Central Evaluation Section commissions and steers the Independent Evaluations of GIZ including Corporate Strategy Evaluations like “Scaling up / Impact at Scale”.

Other relevant experiences/background:

- Commissioning and steering final and ex-post evaluations of the GIZ portfolio in different sectors to provide a final or retrospective assessment of the success of development projects and programmes;
- Commissioning and steering meta-evaluations and syntheses of the various evaluations within each sector in order to obtain findings on organisational learning across programmes as well as on the quality of evaluations.

CASE DESCRIPTION

Title case:

Corporate Strategy Evaluation “Scaling up / Delivering Impact at Scale”

Who commissioned the impact evaluation (IE):

GIZ Monitoring and Evaluation Unit upon a decision of the GIZ Management Board

What was the main purpose of the IE:

Corporate-strategy evaluations address company-wide policies and strategies for service delivery. They are usually designed to span several business sectors and instruments. Apart from addressing the implementation or application of the adopted policies and strategies, their main aim is to examine aspects that serve to develop and refine the company’s policies and strategies.

The corporate strategy evaluation on scaling up and delivering impact at scale is intended to help GIZ further develop corporate policy as well as honing the competence of the company in its capacity as a learning organisation, and expanding business in both the public-benefit and commercial sectors.

Describe the impact that is being evaluated, the intervention that is expected to have led to this impact and the context of the intervention.

The evaluation was based on the following definitions:

- ‘Delivering impact at scale’ is taken to mean delivering needs-driven and qualitatively appropriate improvements that benefit a significant size of target group and that can be maintained beyond the end of the project term.
- ‘Scaling up’ is a consciously selected and targeted impetus during or following a project or programme designed to achieve impacts at scale or on a greater scale. The aim is to transfer concepts, approaches and methods either to a new level (vertically to higher and/or lower levels), horizontally to a new unit at the same level, and/or to a new context (functional scaling up).

After the merger the experience of all three predecessor organisations of GIZ (GTZ, InWEnt, DED) shall be evaluated and integrated into the new company’s spotlights for the future and innovations:

- The role of the former GTZ in scaling-up processes varied widely, and depended in no small way on the interests of the commissioning party, the resources available and the capacities of the partners. In many cases, roles even vary within one measure, sometimes in the course of implementation. In many programmes, GTZ was called on to advise the partner country on the development of appropriate, new concepts. The GIZ’s innovative capacities and the chance of conducting ‘selective experiments’ with it, are very much appreciated in emerging economies in particular, with specific requests for services of this sort being submitted.

In most cases the commission involves advising the partner country not only on the introduction of new concepts, but also on the processes involved in implementing activities at scale. Many measures in fact take this as their entry level, when it no longer appears necessary to pilot concepts because enough experience has already been gained in other contexts. Overall however, the charge has been levelled that too many projects and programmes fail to move beyond the status of successful pilot projects. A large number of the independent evaluations conducted in 2008/2009 point out that more attention must be paid to the imperatives of delivering impact at scale from the outset when measures are planned and implemented.

- With a view to achieving inclusive, structure-building impacts that could be taken as models, the former InWEnt addressed three levels – the individual, the organisation and the system – with its training, human resources, and organisational development measures, and interlinked the three. Regional, cross-border and global approaches enjoy a high status, because the challenges posed by our increasingly interconnected world demand supra-national discussions and approaches. The capacity building instruments used by the former InWEnt include training, dialogue, networking and advisory services for human resources development. These have been incorporated in GIZ's new toolbox as 'human capacity development', which embraces training for managers and specialists of partner organisations, leadership development, e-learning, dialogue platforms, alumni networks, global knowledge cooperation arrangements and networks and capacity building for training institutes.
- At the former DED, methods, approaches and measures implemented at one project location, within one development measure or one country programme, that proved suitable for scaling up were disseminated by the DED coordinators. A variety of methods and procedures were devised by DED to help pass on successful approaches, measures and methods. They included the documentation and discussion of tried and tested methods, approaches and measures in the form of reports on first-hand experience and lessons learned workshops for sector groups at country level and at supraregional level. Within the context of the multilevel approach of official German development cooperation, former GTZ sector coordinators were kept abreast of experience gained by DED at micro and meso that was felt to be suitable for scaling up.

What expectations underpinned the choices for the design of the IE? Which of these expectations were more robust (valid) and which were problematic (less valid)?

Rambøll Management Consulting was commissioned by GIZ to conduct the corporate strategy evaluation. First, a portfolio-screening was completed, in which (i) progress reports of ongoing GIZ-programmes as well as (ii) evaluation re-ports of the mostly completed programmes of the three former implementing organizations GTZ, InWEnt and DED were analysed, based on random samples of the GIZ. Additionally, (iii) a data collection was conducted through online-surveys that were filled out by responsible authorities of the ongoing programmes (270).

The portfolio-screening aimed at systematically preparing the in-depth assessment of selected programmes in the case study period. The results of the portfolio-screening provide indications of correlations between certain driving factors and successfully implemented scaling-up-processes as well as broad impacts that were achieved. In the 7 upcoming case studies, on the one hand scaling-up and broad impacts shall be assessed in detail, and on the other hand the driving factors identified shall be validated and analysed in their mode of action.

The design includes the following elements:

- 'Scaling-up archetypes' reflect textbook combinations of different scaling-up mechanisms within an actual project or programme. These combinations of mechanisms *might* be found in this form in any actual intervention, but are more likely to be found in an intervention-specific form. In the Inception Report I, three hypothetical scaling-up archetypes (pilot project, systematic multilevel interventions, programme-building measures) were presented, which were then used in the portfolio screening.
- 'Drivers' are forces, actors or incentives systems that can exert a (positive or negative) influence on the scaling-up process, and can be actively shaped by the project or programme. 'Spaces' are the (financial), political, institutional, etc. conditions that exert a (positive or negative) influence on the scaling-up process, and can in turn be influenced by the project or programme. 'Pathways' are

operational processes that can scale up an intervention. Pathways include mechanisms that allow the project or programme to use drivers and spaces.

- In terms of the 'influencing factors' to be investigated, a distinction must be made between implementation features, conceptual features and structural features.

What expectations underpinned the choices for the communication of IE findings? Which of these expectations more robust (valid) and which were problematic (less valid)?

The evaluation is still ongoing so communication on findings has not yet started.

Typical challenges for so-called 'learning organisations' are on the one hand choosing adequate tools for generating reliable information, and on the other the systematic organisation of lessons learned and how to translate them into concrete improvements of operational activities.

Opportunities for learning through evaluations are highly attached to the quality of information that evaluations produce, and their validity on an aggregate level. Thus the first challenge to effective organisational changes is to make sure that the quality of findings is high and information reliable. The second challenge is to feedback relevant information – lessons learned – into operational activities in order to improve the quality of interventions.

To what extent has the IE been utilized and why/why not? What influenced this?

The evaluation is still ongoing so the findings have not been utilized yet.

Learning cafés which are attached to a management response system can be a useful strategy to influence the utilization of evaluation findings. . A learning café at GIZ can be understood as a dialogue-oriented format in which evaluators present their recommendations based on lessons learned as well as findings on what has worked, what not, and why. Afterwards the presented recommendations are discussed between mixed groups of management and staff who have a stake in the evaluations. Emphasis is placed on dialogue between evaluators and (operative) personnel in order to identify specific conclusions upon the findings and recommendations. Only selected persons (learning addressees) are invited to join the learning café in order to promote learning in a "secure environment" in which also critical experiences can be discussed without putting personnel under pressure to justify negative outcomes. However, learning addressees are not only those responsible/involved in the evaluated objects, but also those who may benefit from the evaluation results, such as expert colleagues and executive managers from other areas within the company. The purpose is to stream the learning effect up to the sector and enterprise level. At the end of the learning café, the conclusions which had been discussed between the evaluators and participants are agreed on and adhered to in a management response system.

The management response is followed by the Corporate Unit Monitoring and Evaluation and presented to the highest executive management line. This way, the transfer of individual learning (on project level) to changes on the institutional level is assured.

Appendix 2

Case 3: Balas "Are the children better off" Balasahyoga Concurrent Evaluation (India)

PROFILE CASE OWNER

Name of case owner:

Pam Baatsen

Organisation:

Royal Tropical Institute, Amsterdam

Country:

The Netherlands

Your current function:

Senior Advisor HIV and SRHR

Your role in the Impact Evaluation case:

Team leader

Other relevant experiences/background:

Pam Baatsen is a Senior Advisor at KIT in Amsterdam, with a background in cultural anthropology. Over the last 25 years she has gained expertise in evaluation, mixed methods research, HIV prevention and SRHR. Pam has extensive experience in South and East Asia, the Middle East and Africa. She has also led a range of large evaluations, assessments and reviews. Currently, she is – amongst others - working on a Sexual Reproductive Health and Rights Intervention with matching operational research for young male in Bangladesh and Kenya through an innovative Motivational Intervention Approach targeting service providers and young MSM/men.

CASE DESCRIPTION

Title case:

“Are the children better off”, a large scale concurrent evaluation in India.

Who commissioned the impact evaluation (IE):

An independent not for profit organization based in the UK

What was the main purpose of the IE:

The evaluation had three central evaluation questions; whereby the first question related to the impact of the project; the second question to what could have led to that impact and the third question to the model of the intervention itself.

The questions were:

1. Has the Quality of Life (QoL) of children infected and affected by HIV, and their households, been improved
(Changes in QoL scores; BMI, MUAC; trait hope scale; morbidity (outcome HIV tests, CD-4 count levels) and mortality (of children and or their parents).
2. How has the quality of care of children infected with or affected by HIV, and their households, been improved as a result of the strengthening of service delivery capacity by the programme?
(Changes in access to health services; improvements in service delivery; appreciation of the programme and services received; increases in access to other services)
3. Is the intervention model implementable, effective, sustainable and replicable?
(Changes in context and model; added value of the programme; sustainability and replicability of the programme)

Describe the impact that is being evaluated, the intervention that is expected to have led to this impact and the context of the intervention.

Impact evaluated

As mentioned above, the IE looked at the Quality of Life (composed of scores around health, nutrition, education, and psychosocial wellbeing; Body Mass Index (BMI), Mid Upper Arm Circumference (MUAC), Trait Hope Scale for children and adults composed of a Hope Path seeing whether a person is equipped to find a way to get something done and a Hope Score seeing whether a person has the ability to begin and sustain goal-directed behaviour (“can I do this”); HIV test results, CD-4 count test results, and mortality (of children and or their parents).

For question 1, the evaluation combined 1) programme monitoring and evaluation data of over 73,000 children belonging to 40,000 households enrolled in the programme (including on HIV test results, mortality, CD-4 count, ART treatment; BMI or Mid Upper Arm Circumference); 2) data (QoL, Trait Hope Scale, BMI/MUAC) from household surveys in one of the districts involved in the programme (baseline 1,800 children – end line 900 children) and 3) data from the State level AIDS authority.

Impact findings

9 – 16 year old children: Overall improvements were seen over time in this group. The QoL and BMI of **children affected by HIV** had increased, and the QoL scores (especially the health sub-scores), BMI, and the Hope agency scale of **children infected with HIV** had also increased.

0 – 8 year old children: The opposite was true for this group. The QoL score of **children affected** fell significantly, except for their Hope Path scale, while their MUAC/BMI remained more or less the same. The QoL of **children infected with HIV** also decreased significantly, but their BMI and Hope scores remained more or less the same.

The mortality rate in the group of children living with HIV was higher than the rate for all children in the programme (4.6% versus 1%) but the programme had also been able to prolong the life of children living with HIV. Of the 271 Children living with HIV who died, only one third was on Antiretroviral treatment. Children without parents (double orphans) were more likely to have died than other children (2% versus 1%). The actual mortality might have been higher as 9% of the children ever registered had been lost to follow up.

The most vulnerable children identified during the first round were children living with HIV and those living with their grandparents. The latter group was also identified as being the most vulnerable in terms of adherence to regular CD4 testing and treatment.

The findings combined suggest that the younger group of children living with HIV (0-8 years) was particularly vulnerable and that the programme has not been able to improve their Quality of Life. However, trend analysis showed substantial increases in scores of the younger children when they were 42 months or longer in the programme (overall Quality of Life increased from 86.3 to 89.3 for the 0-8 year olds), especially in relation to health and nutritional intake. In this light, it might have been too early to measure the full potential of the programme.

The intervention

The intervention itself consisted of a comprehensive, family-centred approach covering a wide range of services across the health, education, psychosocial support, nutrition, safety net and food security). The intervention combined a community-based approach with a health system strengthening approach.

The context of the intervention

- The project was implemented in 11 districts spread over three distinct geographical zones with different features and cultures.
- The 13 remaining districts were covered by a different organization with a different approach, different project components, and funded by a different donor.
- In addition to these two larger interventions covering the entire state, there were also numerous other smaller interventions in different communities.
- The multi-sectoral approach of the intervention required involvement of different ministries/departments that were all used to a vertical way of working.
- The external environment provided challenges, with numerous changes in leadership within the State, an evolving government-run HIV programme, closure of some of the Community Care Centres, which provided an important link between communities and services, new antiretroviral therapy treatment guidelines over the course of the project and regular unrest and insecurity.
- The donor was eager to only fund the intervention for five years maximum after which the government would have to take the programme over.
- Three large INGOs implemented the programme through a multitude of local NGOs. The three NGOs initially used a vertical approach which shifted in a more collaborative approach over time resulting in a more coherent way of working with families and increased capacity for scaling up.
- HIV is a highly stigmatized issue in the state and the majority of people living with HIV are not willing to reveal their status. Confidentiality is therefore a must.
- A size estimate on the number of children affected by HIV, and infected by HIV is not available.

What expectations underpinned the choices for the design of the IE? Which of these expectations were more robust (valid) and which were problematic (less valid)?

Design

The idea for question 1 was to try to use different data sources whereby “perceived” measures (QoL, Trait Hope Scale) were complemented by harder data (e.g. BMI/MUAC, HIV test results, CD-4 count).

The MIS of the intervention was used as one of the data sources for the evaluation. This because the evaluators believed that it would be a good and very rich source of information to capture the overall population dynamics. Initially there was quite some protest from the implementing agencies that such a MIS would not be of good enough quality. To address this concern the evaluation team made an extra effort to ensure that the data was good enough, through a data audit, but also through providing technical assistance to the intervention on how it could best be computerized.

Due to the magnitude of the intervention with children and households scattered over a large number of districts in the state, it was decided to focus on one district for the household surveys. Whereby children who had been highly exposed to the intervention were compared to children with low or hardly any exposure to the intervention either in time or in programme components (in the first round of data collection, but also at end-line as new children and their families would keep enrolling over time in the programme). The basic assumptions for the sample size were that there should be at least a 50% difference in quality of life between the two groups using a 95% confidence interval, a power of 0.8 and an assumed 5% change in the low exposure group. The district selected for the household survey was one of the districts in which the project had started at the time of the first round of data collection. Due to the highly stigmatized nature of HIV, it was not possible to work outside of the intervention to identify households where someone was living with HIV.

The design also included comparison between BMI/MUAC, height by age, weight by height, and weight by age among children in the household survey versus other children in the same district and other districts. Also mortality, morbidity, availability of CD-4 pre-ART and loss to follow up data were compared. In the design phase, it was believed that this would allow for some generalisation of the findings from the household survey.

In addition, also comparison between MIS data on a number of key indicators with state level government data was planned to see whether the intervention districts were having different results (impact) than the other districts where another organization was working (at a much smaller scale). For this, a number of indicators were identified on which the local state AIDS authority reported for all districts. Furthermore, also effort was made to obtain monitoring data from the other larger intervention covering those districts for comparison reasons.

Validity of evaluation design assumptions

Validity of assumptions

The assumption that there would be a 50% difference in the QoL of those receiving support versus those who did not (or had received very little support), was not a valid one. In the first round of data collection, the QoL measured for all those interviewed was a) higher than anticipated (also based on comparative studies in other countries) and b) there was no difference in the QoL between those having received support and those who had not received support. There was however, a significant difference between children being infected with HIV themselves, and those being affected with HIV (where one or both of their parents were living with HIV, or had died because of AIDS). The children living with HIV themselves had significant lower QoL than those who were affected.

The generalization of the survey findings to that of all the districts was challenging, as the intervention dropped collecting certain critical data, such as MUAC and BMI of all children under 5 years of age.

The decision to collect QoL data on the basis of different measurements, also generated some challenges. The different measures often provided different answers (sometimes contrasting answers) but they were able to paint an overall picture (even though with not always significant differences).

The comparison between the 11 and the 13 districts with state level data was eventually only possible for one indicator. This due to the fact that the state level data was not clean, and not all indicators systematically reported on. However, this indicator was an outcome indicator rather than an impact indicator regarding the update of Mother to Child Transmission.

What expectations underpinned the choices for the communication of IE findings? Which of these expectations more robust (valid) and which were problematic (less valid)?

The evaluation protocol mentioned a number of guiding principles, including

- that credit would be given to all those involved in the evaluation either through oversight and funding, programme implementation, being a member of the Evaluation Core Group or being part of the evaluation team; and
- that baseline and end-line evaluation findings would be shared with the project as soon as they would become available so that these findings could help strengthen the programme. This, in line with the evaluation framework which focused on measuring change, learning and action.

The evaluators made great effort to establish an evaluation core group at the onset of the evaluation. This core group was composed of representatives of the implementing agencies, the state level AIDS authority, other organizations working on children in and affected by HIV in the state, including the implementing agency working in the remaining districts, and the network of people living with HIV in the state. This all to facilitate communication, and increase buy-in to the evaluation and its results.

The sharing of the results with the implementing agencies and the funders was done through regular meetings; reflection action learning workshops (where not only state level staff but also staff working in the various districts were present); and evaluation core group meetings (of which implementing agencies were also member).

We worked on and agreed upon a publication protocol right from the start and had various discussions over time on the dissemination of the findings, and who would be responsible for what. The agreed upon protocol stated that representatives of all parties involved would have to be involved. This due to a number of reasons: 1) HIV data is sensitive in the first place, so having agreement on what and how it is shared is important; 2) the data could have a (negative) impact on the project itself (if the outcome was not what was expected); and 3) contractual arrangements.

On the basis of the first round of evaluation findings, the evaluators, besides an elaborate evaluation report, prepared policy briefs on the different components of the project that could be shared with stakeholders by the intervention INGOs. However, these were not shared, as the INGOs were disappointed with the nil-finding.

Also the end line findings, did not show the - by the implementing agencies - desired results. This made sharing of the final results very sensitive, especially as the implementing agencies were aiming to transfer this project to the State Authorities.

The findings were shared with everyone who had been involved in the project, also at district level through a **Reflection, learning and action workshops** at the end of the evaluation. However, it was left to the project implementers to share the findings with other key stakeholders in the state.

To what extent has the IE been utilized and why/why not? What influenced this?

The IE has influenced the programme itself, to for instance better define the role of the outreach workers. However, the final results, have not been shared with a large audience, due to the sensitivity of the findings.

Efforts have been made to reach agreement on the publishing of an article with the main findings, but to date this has not yet been successful.

KIT has used the findings to advocate for the need to give more focus to younger children infected and affected by HIV in different meetings.

Appendix 2

Case 4: Mixed Methods in Impact Assessment of Value Chain Interventions. A case study on the maize sector in Bangladesh

PROFILE CASE OWNER

Name of case owner and organization:

Shocan Chakraborty, Swiss Contact Katalyst, Bangladesh
Marieke de Ruyter de Wildt, LEI, Wageningen UR, The Netherlands
Gideon Kruseman, LEI, Wageningen UR, The Netherlands
Nabanita Sen, Opportunities Unlimited, The Netherlands

Case to be presented by:

Person	Role in this study	Other experiences
Shovan Chakraborty, Manager in Monitoring and Results Measurement Unit, Swisscontact Katalyst, Bangladesh	He played role in supporting the case owners to understand the need and the context of the IE and in validating the information.	<ul style="list-style-type: none">- Built capacity of beginner staff of 'implementation' and 'M&E' teams regarding M&E processes of Katalyst (which are aligned with the DCED standards)- Co-developed an IT based system to aggregate results across 17 sectors of Katalyst
Gideon Kruseman, Senior Research Fellow, LEI, Wageningen UR, The Netherlands	He played role in methodology design, data cleaning, analysis and interpretation.	<ul style="list-style-type: none">- Environmental and energy economics (model building)- Economics of water, air and soil pollution (model building and policy analysis and evaluation)- Institutional analysis and policy design- Rural development policy analysis and evaluation

CASE DESCRIPTION

Title case:

How to evaluate the ultimate poverty impact of value chain interventions? Mixed methods design for attributing indirect interventions to farmers' income.

Who commissioned the impact evaluation (IE):

Swisscontact Katalyst

What was the main purpose of the IE:

Katalyst had been working in the maize sector for 7-8 years through a number of interventions. The first IE on the matured interventions was conducted on 2011 Jun-Jul by Katalyst. However, six months down the line, Katalyst Management felt that the methodology was not sufficiently robust for valid conclusions on impact. As a response, they commissioned this study to experts of 3rd party institutions (LEI and Opportunities Unlimited) in mid-2012. Expectation was that this external study would apply more rigorous scope of methods to evaluate impact and would compare it with the achievements/projections assessed through studies until 2011.

The central methodological challenge of this IE is typical for indirect, value chain intervention. Donors want to see impact at poverty level but interventions are geared to change market systems around poor people. Hence, poverty impacts are indirect and often take 5 to 10 years to become visible in indicators at household level.

New methods need to be developed that can balance between (A) a realistic understanding of how indirect interventions contribute to longer-term impacts pathways and (B) robust conclusions on impact, right after interventions end. This IE developed such new method.

Describe the impact that is being evaluated, the intervention that is expected to have led to this impact and the context of the intervention.

The IE assesses the experience of Katalyst, a multi-donor project in Bangladesh. With private companies, Katalyst started interventions in 2004 to introduce and promote maize cultivation in one of the poorest regions of the country. Maize, a novel cash-crop for that area, was expected to provide additional incomes for particularly marginalized farmers as it grew best on soils cultivated river islands. With high expectations in terms of poverty impact, Katalyst wanted to record and understand its impact properly.

Katalyst designed three types of interventions to boost maize production: it introduced maize contract farming, embedded awareness and knowledge of maize cultivation in retailers's relation with buying farmers and it motivated government extension officers to promote maize-based cropping patterns by .

To assess the impact of these activities, Katalyst maintains an M&E system in compliance with the DCED-standard, which is based on results chains. In the course and after an intervention, indicators for each box of the results chain are assessed by enquiries of service/input providers and beneficiaries. It mixes small n with large n approaches. The IE was expected to identify the people that benefitted from by Katalyst's interventions, how they benefitted, by how much and why.

What expectations underpinned the choices for the design of the IE? Which of these expectations were more robust (valid) and which were problematic (less valid)?

The standard methods for this IE were found inappropriate. RCT, for example, could not be applied because interventions were in a final stage and ex-post 'treated' and 'untreated' farmers could not be singled out post. A mixed methods approach was designed, including surveys (large n) and in-depth interviews with treated and untreated respondents from various levels of the value chains (small n).

Validation of information collected through multiple sources was a primary duty. The design of the IE was based on a rigorous impact logic, linking the various interventions to the expected ultimate impact. For some of the impact pathways attributable impact was difficult to determine while for other pathways, attributable impact could be established.

What expectations underpinned the choices for the communication of IE findings? Which of these expectations more robust (valid) and which were problematic (less valid)?

In preliminary communicating the IE findings emphasis was placed on impact pathways for which attributable impact could be established. A workshop was held to communicate draft results to local stakeholders who could give further feedback to incorporate changes. In final communication of the IE findings, the impact pathways for which contributable impact could not be determined are also treated in order to draw lessons for future interventions. Publication of a paper & presentation in conferences is used to inform a larger crowd and validate the key choices made in the design.

To what extent has the IE been utilized and why/why not? What influenced this?

The IE is being utilized ex-ante and ex-post. Ex-post for the project and donors, to validate impact claims. Ex-ante to improve on future intervention design (improve aid effectiveness), the M&E system in place (get better utilization of information generated) and on impact expectations by donors (enhance more realistic expectations on what impact can be expected).

Appendix 2

Case 5: A beneficiary picture of development and change (PADev: holistic and participatory approach to development assessment)

PROFILE CASE OWNER

Name of case owner:

Ton Dietz, Dienneke de Groot

Organisation:

African Studies Centre (ASC, Leiden, the Netherlands); ICCO (Utrecht, the Netherlands)

Country:

See above

Your current function:

Director ASC and professor of the Study of African Development at Leiden University; coordinator PMEL unit ICCO Alliance

Your role in the Impact Evaluation case:

Both were involved in the development of the PAdDev instrument. PAdDev is an innovative toolbox of techniques and methods that aims to achieve a bottom-up assessment of development and change in a particular area over a period of time based on the value systems of the population. PAdDev is a holistic and participatory approach to development assessment (see for the toolkit: www.paddev.nl)

Other relevant experiences/background:

Ton Dietz did quite some work on participatory impact research preceding the development of this instrument (e.g., in Kenya Dietz, T. (2008) Participatory evaluation of development initiatives in Western Pokot).

CASE DESCRIPTION

Title case:

A beneficiary picture of development and change

Who commissioned the impact evaluation (IE):

ICCO, Woord en Daad (W&D) and Prisma, in cooperation with the 'Amsterdam Institute for Metropolitan and International Development Studies (AMIDSt, University of Amsterdam), University of Development Studies (UDS, Tamale, Ghana) and the 'Environnement et Développement au Sahel' (EDS, Ouagadougou, Burkina Faso),

What was the main purpose of the IE:

The main purpose to engage in this project was the felt need for developing a robust instrument for participative impact assessment from a beneficiary perspective. Many of the until then known impact evaluations were either done over a rather short period of time, carried out in a fully quantitative way, or from a donors or international institutional perspective. A topped perspective – from those participating in programmes and projects focussing on their development in various ways- was deemed necessary to be able to learn on the why questions and to assess the long term impact of inducing change in a situation of multiple stakeholders.

Describe the impact that is being evaluated, the intervention that is expected to have led to this impact and the context of the intervention.

PADev can be differentiated from conventional assessment methodologies which often focus on a single agency, project/programme or sector with quite a "technical" expert-driven approach to evaluation of output, effect and sometimes impact. In PADev, participants assess a wide range of changes, projects and agencies based on their perceptions. Further, PADev assessments typically look back at development and change over the past twenty to thirty years. This yields extremely valuable information for NGOs and other agencies, e.g. the government and foreign donor agencies' in the area: they learn about their own impact vis-à-vis other actors, and in addition, they find out which types of projects have been regarded as most effective and relevant and more importantly: for whom and why. This can be an important lesson for future interventions. PADev should not be thought of as a replacement for conventional approaches, but rather as an alternative approach that enables a "big picture" of development and change to be constructed for an area

What expectations underpinned the choices for the design of the IE? Which of these expectations were more robust (valid) and which were problematic (less valid)?

Four years of experimenting with the PADev approach have revealed its strengths, but also some limitations or challenges. The strengths lie in the progress made towards addressing some key weaknesses in many other evaluation methods:

- They focus on too short a period;
- Projects are evaluated in isolation of wider developments in the region;
- The opinions of the supposed beneficiaries are largely neglected.

One of the main strengths of the approach – its ability to compare the contribution made by particular projects with other agents of change – also holds a limitation. Its broad scope inevitably reduces the focus on specific projects or interventions. More often than not, evaluations of development projects serve the short-term bureaucratic purpose of meeting donor requirements. For such ends, the PADev approach involves gathering too much 'irrelevant' information. By contrast, if development agencies want to know

their real contribution to societal change and desired outcomes such as poverty reduction, livelihood sustainability, food security and better health, and if they want to learn how to improve their performance, the PAdDev approach is a very suitable tool. It can help agencies to discover their strengths and weaknesses vis-à-vis other organizations operating in the area and it can help them to discover niches and needs.

What remains a problem, also in PAdDev, is reaching the ultra-poor participants and beneficiaries of this method. However, attempts have been made to focus PAdDev specifically on these ultra-poor, and experiments continue to develop that further.

What expectations underpinned the choices for the communication of IE findings? Which of these expectations more robust (valid) and which were problematic (less valid)?

PAdDev chose for a completely transparent dissemination strategy, making use of a dedicated website www.paddev.nl. All basic data, all workshop reports and all further analyses can be found there, as well as a guidebook (now on the website in English, Spanish and Chinese; French will follow soon). Locally attempts have been made to go back to the participants in the PAdDev workshops and to some of the more influential development agencies in the study areas to discuss the findings. In 2012 specific dissemination sessions have been organised in two of the most important study areas. That was also partly based on a student study (Zjos Vlaminck) about the 'uptake' of the 'PAdDev spirit' in one of these study areas between 2008 and 2011. Most problematic was the fact that it often took too long between workshops and the report, and that it was cumbersome to organise follow-up activities with and among the participants.

To what extent has the IE been utilized and why/why not? What influenced this?

The PAdDev methodology has been used by the University of Tamale, and by EDS in Ouagadougou for many additional evaluative and teaching/training activities. At the University of Amsterdam and the ASC many students were informed about the method and use it in their work. At the Royal Tropical Institute in impact assessments for e.g. how investments in SMEs may lead to social change at producer's level in several cases in Eastern Africa. It is a good instrument for use in a quasi-experimental design.

Appendix 2

Case 6: Methods for impact assessment of certification schemes

PROFILE CASE OWNER

Name of case owner:

Verina Ingram, Giel Ton, Yuca Waarts

Organisation:

LEI, Wageningen University & Research Centre (WUR)

Country:

Netherlands

Your current function:

Senior researchers

Your role in the Impact Evaluation case:

Researchers and project leaders

Other relevant experiences/background:

- Lived and worked among commodity producers in Central Africa and South America for over 5 years;
- Backgrounds in multidisciplinary impact assessment (focus on environmental and social).

CASE DESCRIPTION

Title case:

Methods for impact assessment of certification schemes

Who commissioned the impact evaluation (IE):

The case is based on learning from a number of studies with and for the following organisations: UTZ, ISEAL Alliance, Ministry of Foreign Affairs, Ministry of Economic Affairs, Agriculture and Innovation, commodity processing companies, Solidaridad/WAFF and the Dutch Sustainable Trade Initiative.

What was the main purpose of the IE:

Demonstrate impacts of corporate sustainability schemes, particularly the certification of tropical commodities (tea, coffee, cocoa, cotton).

Describe the impact that is being evaluated, the intervention that is expected to have led to this impact and the context of the intervention.

As part of the Dutch Sustainable Trade Action Plan organisations (industry, government and civil society) in cocoa, coffee and tea chains have committed themselves to targets of certified sustainable consumption of these products in the Netherlands by 2025. Certification is seen as a key route to promote, demonstrate and “control” responsibly grown and produced commodities with a triple bottom line focus on economic, social and environmental impacts at farmer, group, company and chain/sector level. The policy context strongly focuses on the use of multi-party partnerships to achieve this goal, whereby different partners provide for example, financial and/or technical support, training and support to build knowledge and skills for certification, implementation of certification schemes, organisation of producers and other organisations in the chain, the implementation of supporting social and environmental programs and of services e.g. credit and farm inputs.

The interventions are expected to lead to better and more sustainable practices which are expected to lead to higher and long term productivity, improved product quality (i.e. better market access and prices), increased efficiency (i.e. lower costs per unit of produce), increased income for farmer-producers and companies involved (i.e. improved profitability) and improved social and environmental conditions, particularly at producer level. Training of producer groups focusses on organizational management and control systems to the standards required by certification schemes, and is expected to lead to more effective trader and farmer organizations with more effective input purchasing, marketing and better service delivery to farmers .

What expectations underpinned the choices for the design of the IE? Which of these expectations were more robust (valid) and which were problematic (less valid)?

Robust

- Combining and using mixed methods of impact assessment
- Using data from multiple sources
- Difference in Difference approach, comparing the target group before and after situation and comparison groups
- Using an impact logic intervention approach
- Methods to collect information on impact indicators (don't be afraid to be innovative and pool resources!)

- Using counterfactual thinking: Ensure plausible rival explanations – external causal factors – accounted for (collecting information for comparison ('control groups'), triangulate with similar case studies to identify plausible alternative explanations for observed outcomes), exploring heterogeneity in outcome patterns (i.e. purposefully sampled groups or areas, which pose high risks to realizing theory of change)
- For Intermediate outcomes, use appropriate proxies for key outcome areas

Problematic

- Ability of ex-ante baselines to allow robust impact assessment
- Lack of availability of (other) baseline data
- Intervening organisations difficulty to look back and reconstruct intervention logic and impact pathways
- Underestimation of timescales needed for data collection
- Need for control groups in the absence of a baseline situation
- Documentation of activities implemented over time expected to lead to impact
- Developing pragmatic indicators for the current IAs and future M&E - focus on impact whereas immediate and intermediate outcome areas more pragmatic
- Selecting only key impact indicators (cost and time implications)
- Recognising where in the intervention logic interdependencies on other factors/actors become dominant
- Monitoring how you manage these interdependencies
- Identify when and where it becomes impossible to derive attribution claims
- Refrain from 'impossible' quantitative-attribution research
- striking a balance between ownership (by multiple stakeholders) and independence of results

What we learnt:

- monitor immediate outcomes
 - Learn for increased performance with immediate outcome indicators that help to build and adjust the specific intervention theory
- measure intermediate outcomes
 - Evaluate changes in intermediate outcome indicators that are informative for benchmarking performance of the intervention
- build a plausible story line for the contribution to the ultimate outcomes
 - Make use of existing quantitative information to reflect on the impact logic
 - Collect qualitative information that can both support and challenge your story-line

What expectations underpinned the choices for the communication of IE findings? Which of these expectations more robust (valid) and which were problematic (less valid)?

Robust

- Requirement for a suite of communication tools designed to meet the demands of different audiences: scientific reports, scientific articles, public orientated brochures, webpages
- Developing guidelines for sharing data e.g. using Chatham House rules
- Methods for sampling, data gathering and data analysis

Problematic

- Confidentiality issues i.e. how to make results anonymous but maintain details that the results make sense
- Sharing (pre-competitive) data on interventions - particularly costs and target groups and their results
- Interpreting results in the light of different partner/stakeholder perspectives and intended uses of data
- Maintaining independence and legitimacy, but taking different parties interpretations of results of impact studies (and their possible consequences) into consideration

Also: the pros and cons and alternatives to “ golden standard” methods of IE.

To what extent has the IE been utilized and why/why not? What influenced this?

The results of the first studies are just being finalised, so this section is details stated intents of different participating organisations, rather than how IA have actually been used:

1. For (internal) organisational monitoring, evaluation and learning;
2. To assess the effectiveness of activities with farmers, groups and implementing partners and demonstrate their contribution to impact;
3. To demonstrate impacts of sustainability schemes/certification interventions to customers and end consumers, Board, shareholders, financiers;
4. To evaluate the efficiency and effectiveness of policy support and interventions concerning private sector development, aid for trade and foreign affairs;
5. To understand direct (also indirect and unintended) achievements of sustainability programs and certification schemes;
6. To assess whether the activities/interventions lead to the desired outcomes/and impacts
7. To (indirectly) evaluate effectiveness, results and impacts of partnerships;
8. To improve the design of credible certification schemes;
9. To aid the design of guidelines and standards for credible certification schemes;
10. Managing expectations - of all stakeholders- but especially the end consumer about what certification and schemes can be expected- and do – achieve;
11. Planning for communication (e.g. commissioning parties find it difficult to decide beforehand on publishing results) and ownership of data.

Appendix 2

Case 7: Impact Evaluation of Rwanda's National Domestic Biogas Programme

PROFILE CASE OWNER

Name of case owner:

Willem Cornelissen

Organisation:

Erasmus University Centre for Contract Research and Business Support - ERBS

Country:

The Netherlands

Your current function:

Senior researcher

Your role in the Impact Evaluation case:

Quality controller on behalf of one of the contracting authorities, the Ministry of Foreign Affairs of The Netherlands.

The second contracting authority is SNV. Both organisations are 'owner' of the study.

The field studies were implemented by the International Institute of Social Studies, Erasmus University Rotterdam

Other relevant experiences/background:

Policy evaluator of well over 40 evaluation assignments worldwide, in particular of subsidy schemes in the area of development cooperation and international affairs.

Expert in public finance management.

Main clients: Regional Development Banks, European Commission and Netherlands' ministry of Foreign Affairs

CASE DESCRIPTION

Title case:

Impact Evaluation of Rwanda's National Domestic Biogas Programme

Who commissioned the impact evaluation (IE):

The impact evaluation was commissioned by the Policy and Operations Evaluation Department of the Netherlands' ministry of Foreign Affairs in co-funding with the Netherlands Development Organisation SNV.

What was the main purpose of the IE:

The study forms part of a series of impact studies that jointly provide the field level evidence of the effectiveness of the implementation of the Netherlands' policy on renewable energy in developing countries.

The main purpose of the biogas study was to identify the impact of access to biogas provided through a Dutch funded Rwandan government programme about biogas.

The choice for rigorous methods was motivated by the fact that scientific evaluations on biogas are extremely scarce.

Describe the impact that is being evaluated, the intervention that is expected to have led to this impact and the context of the intervention.

The impacts have been determined based on a results-chain.

The input: The Netherlands' ministry of Foreign Affairs intervenes with financial resources and SNV with technical assistance in order to enable the Rwandese government programme to disseminate biogas digesters at household level.

The Output: cooperatives promote digesters; subsidy scheme is operational, masons are trained and qualified

The Outcome: bio-digesters are installed and users sensitized

Acknowledging an attribution gap: intermediate impact: biogas provides fuel for lighting and cooking, while the rest product, slurry, is used for agricultural purposes.

The impact: household convenience and feelings of security; slow down of deforestation; savings to household and income generation; gender equity; improved health; less eutrophication; improved agricultural yields.

What expectations underpinned the choices for the design of the IE? Which of these expectations were more robust (valid) and which were problematic (less valid)?

The evaluation design was based on the results chain, as well as the conceptual requirements for applying rigorous methods, using mixed methods (that means both quantitative and qualitative elements).

Robustness is a requirement ex ante that determines –to a large extent- the sample size. The other determinant was pragmatic: the limitations in financial resources and in time. Financial restrictions had an impact on i) the sample size, ii) the choice of a cross-sectional method (and not a double survey) and iii) the use of surveys only (and not additional objectively verifiable methods).

In the biogas study, the survey was sufficiently large (a sample of 600) to produce robust data.

What expectations underpinned the choices for the communication of IE findings? Which of these expectations more robust (valid) and which were problematic (less valid)?

Draft results are communicated to the most important stakeholders involved for consultation and correction. Each study is presented to an independent forum of professionals, the Reference Group.

The communication concerning IOB reports are determined by public regulation derived from the Budget Law, the Regeling Periodiek Evaluatieonderzoek (RPE). The first and most important recipient is the Dutch Parliament and through parliament the Dutch population.

Information to Dutch professionals as well as a broader audience is organised in advance and consists of i) a Newsletter bulletin, b) publication on the IOB-website, c) the printing and active dissemination of approximately 500 hard-copies d) the presentation of results during seminars e) in some cases active dissemination through the national press. Towards the partner country: official presentation of the results; organisation of a workshop.

To what extent has the IE been utilized and why/why not? What influenced this?

Not yet applicable, since report has not been officially released.

Appendix 2

Case 8: The Most Significant Change method for assessing impact with community and children (India)

PROFILE CASE OWNER

Name of case owner:

Inka Pibilova

Organisation:

Czech Evaluation Society

Country:

Czech Republic

Your current function:

Organisational Development Consultant / Evaluator

Your role in the Impact Evaluation case:

External Evaluator

Other relevant experiences/background:

- Fourteen years of experience in project and programme cycle management including planning, monitoring and evaluation. Well experienced with the EC procedures, rules and regulation as a leading agency, trainer and evaluator.
- Five years of work experience in evaluation of community-based as well as national and international projects and programs; focusing on education, health and social causes. Clients include beside NGOs also UNDP / MFA SK and MFA CR.
- Strong in institutional development and capacity building of CSOs as well as government representatives and private sector. E.g. launched innovative tools such as peer reviews, facilitated a number of strategies development.
- Member of the Czech Evaluation Society - ČES, o.s. <http://www.czecheval.cz/> - the International Development Evaluation Association - IDEAS, <http://www.ideas-int.org> and the European Evaluation Society – EES, <http://www.europeanevaluation.org/>.

For more details, see cz.linkedin.com/in/inkapibilova/.

CASE DESCRIPTION

Title case:

The Most Significant Change method for assessing impact with community and children

Who commissioned the impact evaluation (IE):

The IE was commissioned by ADRA Czech Republic, which had funded a tsunami relief project implemented by ADRA India.

What was the main purpose of the IE:

The purpose of the evaluation was to:

- Show evidence of the project impact on access to education, increased education quality and child-friendly learning environment, thus achieving learning outcomes and higher literacy.
- Show evidence of attitudinal changes of community, school staff, children and government representatives in education.
- Assess sustainability of the “Pro Learn” project after ADRA withdrew support.
- Draw lessons learnt and recommendations for improvements in future projects with respect to planning, implementation, monitoring and evaluation.

The primary users of the end evaluation are ADRA Czech Republic and ADRA India. The secondary users are local schools management, teachers and volunteers as well as government Education Officers, who were expected to be provided with lessons learnt and recommendations to increase the impact and sustainability of the project after the withdrawal of ADRA. Another secondary purpose referred to accountability to individual donors, who contributed to the project.

This evaluation did not primarily aim to assess relevance, effectiveness and efficiency of project interventions; however general observations and findings were also noted with the purpose to learn from this experience for future projects. A separate assessment of constructed classrooms was conducted by a civil engineering expert, whereby results were not made available to the evaluators and therefore the results were not included in the evaluation report.

Describe the impact that is being evaluated, the intervention that is expected to have led to this impact and the context of the intervention.

In 2004, tsunami devastated the costal line on South India and a lot of families lost their lives, households and livelihoods. After the tsunami, communities started rebuilding their homes and livelihoods as well as sending more children to schools. However, the government schools were not able to accommodate the increasing number of students. Classes for different standards had to be conducted in one room or at a terrace outside of the schools. This affected the teaching quality and learning outcomes.

In June 2007, ADRA India conducted a needs assessment of 20 schools in two blocks of Nellore District, which were damaged in tsunami in 2004 the most. The findings supported the need of infrastructural and capacity building in the schools. ADRA India along with the District Education Office selected 8 schools and launched project ProLearn – **“Promotion of learning conditions in 8 tsunami affected schools in Nellore District, Andhra Pradesh, India.”** The overall goal was a *“higher literacy rate in Nellore District in Andhra Pradesh in India”*, whereby the specific objective of the project was *“increased and retained number of school children in Allur Block of Nellore.”* Schools, including teachers and children (age 6-12 years) were key target groups, whereby the community as whole was expected to benefit from the intervention. The project activities were concentrated on the following areas, which formed a holistic

approach to education in project schools: establishment of “Child Friendly Schools” and construction of classrooms, teachers training in activity-based learning solutions, supply of teaching learning materials, local government sensitisation, community sensitisation and children participation. The project was implemented from December 2007 to April 2009.

In April 2009, a 3-member evaluation team conducted the end evaluation of ProLearn with active participation of all stakeholders. The evaluation team found that all project activities took place and **benefited around 1500 children**, their teachers, families and the community in large. Although it could be argued that ADRA’s investments in infrastructure to certain extent substituted the government’s responsibility, the high quality classrooms set a standard for other government schools in the District and were used later as role models. The new team managed to incorporate the holistic approach in the project and achieved visible changes in behaviour in the last 6 months of the project. This was achieved also thanks to project volunteers, who acted as education champions in the community and supported well all new institutions.

The **impact** of ProLearn **could not be fully assessed** as classrooms construction was just completed and institutions have been functioning for 2-6 months at the time of the end evaluation. Only enrolment and performance figures in the academic year 2008-09 could reveal a progress towards the project objective and goal. However, several improvements were already apparent in April 2009. The **Child Clubs were** evaluated by the majority of stakeholders as **the “most significant change” brought by ProLearn**. Child Club members developed their self-confidence, their leadership skills and ability to lead a constructive discussion and present opinion to higher authorities. Moreover, they delivered a number of successes ranging from improved school infrastructure to attendance, to higher quality mid-day meals, to community involvement.

As per the evaluation, children started to attend the school more regularly and felt attracted especially to the new teaching - learning and playing materials. Several schools reported that there was an increased unity among children. Some teachers adopted innovative techniques, while others at least started to attend the schools more regularly. The cooperation of teachers and the community improved and community became more aware of children’s progress. Community was proud of the quality of education in their respective villages and therefore preferred to send their children to the project schools rather than to private schools. Community members were also more aware of their role in education and of established institutions. A number of community members were involved in school management and development either through participation in School Development Committees or individually. Donations of land or a playground were among examples. The local government officers were involved in the same and therefore mainly school infrastructure could be further enhanced. Students’ performance reportedly improved; however, this is to be attributed to the Block Education Officer, who strongly focused on bringing low performers back on track.

Sustainability was the weak point of ProLearn project. The **phase-out strategy was not a part of the project plan**. Sustainability plans were not developed even later when it was clear that newly established institutions would need further support. Rotation of some teachers, headmasters and even the Block Education Officer posed a further challenge to keep the project successful path. Former project volunteers, remaining teachers, communities and the new government officers therefore needed to unite again to provide quality education especially to underprivileged children, who were still out of school.

It was **recommended** to the school administration to focus on out of school children and address the root causes of their absence at school e.g. during a door-to-door campaign and home visits. New teachers and headmasters should have received induction in concept brought by ProLearn. Activity-based education needed a systematic support from the government or an expert institution to achieve a desired change. Continuous support and mentoring of Child Clubs and School Development Committees was also needed.

To address attendance and performance, close cooperation of teachers and parents was necessary. Remedial classes were also recommended in all schools. Further, schools should have considered alternative solutions to increase literacy among youth as well as adults, such as evening literacy classes, vocational trainings etc. In order to share project outcomes, discuss the project recommendations and agree action plans, an evaluation workshop was proposed for autumn 2009, whereby representatives of all stakeholders should participate. Government officers and a potential partner from local non-government organisations should have also been involved. The success of ProLearn should have been promoted locally, as well as on-line in receiving as well as donor's country. ADRA could have adopted a holistic approach to education and community development, as partially piloted in Nellore District.

For ADRA's future projects, the key **lesson learnt** was to involve all stakeholders from initiation and planning, to execution, to monitoring and evaluation. Baseline survey would help identify root causes of the problem, set adequate objectives and indicators and also provide quantitative and qualitative data for comparison. Project teams need to be carefully selected and fully supported. Finally, monitoring and evaluation need to be planned systematically and deserve constant attention.

What expectations underpinned the choices for the design of the IE? Which of these expectations were more robust (valid) and which were problematic (less valid)?

1. Logical matrix (theory of change) was incomplete, it had changed several times and did not correspond to the final project strategy. A number of other activities were initiated that could have potentially led to unplanned results / effects.
2. The donor / implementing organisation had already undertaken an audit of the construction and were interested especially in planned and unplanned impacts and sustainability.
3. The donor did not have experience with evaluations and did not have a specific method for IE in mind.
4. There was lack of funding to conduct IE.
5. There was a lack of baseline data.
6. The evaluation had to be conducted in 3 weeks, i.e. before upcoming elections and the end of project as there would be no organisational support for the evaluator/s (e.g. a car) thereafter.

What was the particular (innovative) approach for impact evaluation (IE):

In order to collect unexpected (positive and negative) impacts, the Most Significant Change (MSC) was selected by the lead evaluator as a useful method to assess even "soft" changes, such as changes in attitudes of children, without having a baseline study. The Most Significant Change (MSC) method is explained in the free MSC Guide at <http://www.mande.co.uk/docs/MSCGuide.pdf>. The method was used together with other quantitative and qualitative methods – data review, key informant interviews, focus groups, community group discussions, survey among schools and site visits.

Community volunteers were trained and sent to different villages to collect MSC stories (see the Guide for details). Additionally, children in intervention schools were involved in an MSC drawing exercise - following an introduction, children closed their eyes, imagined themselves at the beginning of the academic year and then at the end. They were supposed to draw a picture (and the older ones could include a story) of the biggest change for them in the academic year. Children were also asked to come up with a headline for the story (e.g. for newspapers). The results were quite astonishing - a lot of children drew the new school building and sports / learning materials, provided during the project. Further, they mentioned Child Clubs (established during the project), where they discussed school problems like lack of potable water or dirty toilets. Others drew their friends, homes, flowers, local animals. Children generally pictured happy moments. No children drew a teacher or parts of their curriculum. At the final MSC panel, criteria were set for selecting the Most Significant Change story, and diverse stories were debated jointly with project staff and community volunteers. The selected top MSC story concerned the Child Clubs, which were not originally planned in the project logical framework and were just a "side activity" of the project manager.

This was rather surprising for the project staff as well as for donors, who spent most of the resources on enhancing the school infrastructure. It needs to be noted that the MSC method was triangulated with document review, semi-structured interviews, community (group) discussions and evaluator's observations. Community volunteers and all other actors were involved without any remuneration – their commitment to the cause was high and thus they participated with quite a passion. Due to established networks, it was possible to undertake the evaluation within the required short time span.

What expectations underpinned the choices for the communication of IE findings? Which of these expectations more robust (valid) and which were problematic (less valid)?

1. The project finished a week after the evaluation was completed and the project staff moved elsewhere – they stopped working for the implementing organisation. Thus it was clear that key communication had to be done at the end of the evaluation.
2. The local community was to a big extent illiterate. Therefore, community volunteers participated in selection of the Most Significant Change story and the evaluation debriefing. They were expected to debrief the respective communities (they continued in their roles and therefore they held subsequent meetings with communities on diverse issues – here the community could be debriefed).
3. The implementing organisation did not plan to further work in the target/intervention area. The debriefing of the headquarters in India was done only by Skype due to lack of availability for a personal meeting.
4. The donor organisation did not plan to further work in the target/intervention area. The debriefing of the donor and debate on subsequent actions was held in person back in Europe.

To what extent has the IE been utilized and why/why not? What influenced this?

The government school staff did not own much the project (teachers' absenteeism was a local issue). Therefore the principals did not attend the final debriefing. It is assumed that evaluation report did not reach the authorities, though the interviews with them during the evaluation revealed a good awareness especially on the quality of school construction and interest to copy the design for other schools.

The implementing organisation had no staff to work subsequently on communication of evaluation findings. We have originally planned a short, 2-page debriefing paper for schools and even press. The papers were expected to contain a specific story about the Child Club. Finally, as these were already communicated previously to the press, the implementing organisation had not communicated the IE results in paper.

The donor organisation decided to delay the communication to individual donors in Europe as accounting of the newly established www.edonation.cz had to be consolidated. Ultimately, the full evaluation report was not published, though parts were mentioned to media. Finally, the evaluation report was used only internally.

What I would do differently:

Next time, I would keep the MSC method, but would clarify better how the targeted villages and schools will be debriefed in a coordinated manner. The community volunteers could have also been systematically trained how to disseminate IE findings. A story in pictures (painted by children on their schools) may have been an interesting solution, for instance. Further, I would explore more the possibility how to bring together principals of the schools and the Educational Bureaus in order to own the findings and recommendations – probably at an official meeting, which had to be budgeted for.

Appendix 2

Case 9: Systematic analysis and synthesis in qualitative evaluation. A case study evaluation of the Oxfam Novib programme in Burundi (2005-2009)

PROFILE CASE OWNER

Name of case owner:

Ferko Bodnar

Organisation:

Part-time: Policy and Operations Evaluation Department (IOB), Ministry of Foreign Affairs; part-time: freelance consultant tropical agriculture and project evaluation.

Country:

The Netherlands

Your current function:

Inspector (Evaluator) for the theme Food Security

Your role in the Impact Evaluation case:

I was leader of a team consisting of 2 Dutch consultants, 2 Local consultants and 7 enumerators in Burundi. (At that time, 2010, I worked as freelance consultant).

Other relevant experiences/background:

I recently did a systematic review of the impact of food security interventions (2011) and I am currently involved in the evaluation of the Dutch food security programme 2012-2015. Before I joined IOB, I first worked in Malawi and in Mali on agroforestry and soil and water conservation (1994-2002). Later I worked on value chain development of organic and Fairtrade products from Africa (2005 - ongoing). Meanwhile I regularly evaluated development projects and programmes and advised on M&E in project design. I hold an MSC in tropical crop science and a PhD in Production Ecology and Resource Conservation at Wageningen University, the Netherlands (2005).

CASE DESCRIPTION

Title case:

Systematic analysis and synthesis in qualitative evaluation. A case study evaluation of the Oxfam Novib programme in Burundi (2005-2009).

Who commissioned the impact evaluation (IE):

Oxfam-Novib (ON), Netherlands

What was the main purpose of the IE:

For Oxfam Novib, this was one of the 'Core Country Evaluations', in which the whole portfolio of projects in a given country (Burundi) and in a given period (2005-2009) is evaluated. This type of evaluation goes beyond the individual project evaluations. It serves two purposes:

- Accountability: towards the Dutch Ministry, other donors, and the Dutch public;
- Learning: improve strategic decisions in the current and upcoming Oxfam-Novib (ON) programme in Burundi; provide input for ON thematic policies and ON strategic programme 2011-2015, and for Oxfam International; support organisational learning by local partner organisations and other stakeholders.

Describe the impact that is being evaluated, the intervention that is expected to have led to this impact and the context of the intervention.

Oxfam-Novib (ON)'s programme in Burundi worked on five themes: 1) sustainable livelihood: food security, income, employment, markets, 2) social services: health, education, 3) security: emergency aid, conflict prevention, 4) social and political participation, and 5) gender and diversity. In the period 2005-2009 ON supported 31 projects, implemented by 12 partner NGOs, with a total budget of 5.7m Euros. Based on the project documents and discussions with Oxfam Netherlands staff, we identified a series of evaluation questions, from output to impact, for each of these themes.

Burundi was still in the reconstruction and reconciliation phase after the end of the civil war in 2005. Many people have been displaced and the process of returning is still ongoing. My personal impression is that the social cohesion within a '*colliné*', a farming community spread out over a hillside, is very weak compared to the strong social cohesion that you find in West African villages, e.g. in Sierra Leone, which also came out of a civil war (in 2002).

What expectations underpinned the choices for the design of the IE? Which of these expectations were more robust (valid) and which were problematic (less valid)?

Before discussing expectations, I will first describe the design, the methodology. In this programme evaluation, we did not want to go into detailed outputs of 31 individual projects, but searched for more generic programme outcomes for each of the strategic objectives (themes). Because the strategic objectives described by Oxfam Novib often lacked clear indicators or baseline data, we reconstructed for each strategic objective a result chain from intervention to impact, based on programme documents and discussions with Oxfam Novib staff in the Netherlands. In these result chains, we distinguished impact and outcome on final beneficiaries, and outcome on partner and other organisations. This resulted in an evaluation matrix with evaluation questions before going into the field. Information sources included, besides the directly involved project staff and targeted beneficiaries, also independent sources such as other NGOs, government organisations, and not directly targeted households. Besides focus group discussions, a survey was done among 293 randomly selected households in the areas targeted by one or more ON projects, which automatically included targeted and (many) non-targeted households. In

discussions and survey, we first asked for changes per theme at impact level before discussing the causes and the possible contribution by projects.

One of our concerns was how to deal with different opinions in a systematic and transparent way, so that opinion holders recognise their own opinion, understand the opinions of others, and understand how an overall conclusion is drawn from these different opinions. This was even more important because there were not much 'hard data' to prove the results of the programme. The different opinions were presented in disaggregated form in tables, based on the 'evaluation matrix'. One table was made per evaluation question, with a few judgment criteria. The different statements and opinions were inserted, indicating from which group these opinions came. These indications summarized in a way the findings: + for positive – confirming the judgment criterion, - for negative, in 5 narrow columns next to the statements. In my case I distinguished targeted people (rural households), non-targeted people, involved project staff, other staff not involved, and government. This overview showed clearly where opinions converged and where they contradicted. This way, the answer to the evaluation question, at the bottom of the table, becomes a logical, almost unavoidable conclusion. In the restitution meeting, people could still add or reinterpret statements and opinions, but the logic of how to come to conclusions could not be dominated by one particular stakeholder.

The expectations underpinning this design were:

1. Individual project objectives matched the programme objectives even though the latter were described in more detail afterwards.
 - a. I think this worked more or less well because we focused at outcome and impact level to which all projects intended to contribute.
 - b. However, some partner organisations felt that their project specific outputs were underappreciated in this evaluation.
2. The inclusion of non-targeted households and not-directly involved organisations strengthened the assessment of the Oxfam Novib programme contribution to changes at household level.
 - a. I think this worked really well. For example, some spontaneous discussions with not-directly targeted households in the project area added valuable information about the context and unintended effects that were not mentioned by targeted beneficiaries.
 - b. However, one disadvantage of the way we took our survey sample was that many more non-targeted households than targeted households were interviewed. This also contributed to the feeling that some project results were underappreciated. On the other hand, it also showed the importance of scale: small, isolated interventions that were not replicated or scaled up, lost importance in our evaluation.
3. The systematic presentation of different opinions facilitated drawing (tentative) conclusions and reduced evaluator bias. This was done just before the restitution / synthesis workshop.
 - a. It worked well in considering and balancing all different findings, and reduced our own (evaluator) bias.
 - b. It is time consuming for the evaluators, and takes up many pages in the annex... but I would do it again this way.

What expectations underpinned the choices for the communication of IE findings? Which of these expectations more robust (valid) and which were problematic (less valid)?

Before discussing expectations, I will first describe the communication of the IE findings.

We expected a certain resistance of partner organisations where our draft conclusions would be disappointing. Therefore, we organised the restitution and synthesis workshop in two days. The first day, several sub groups discussed one theme each (partner organisations working on more themes had sent

more delegates). They discussed the results table that presented survey results and different opinions. People could add clarifications and add more opinions to the table (not directly into the conclusions). The second day, we discussed the draft conclusions with the whole group, considered the additions made in the results, and revised the conclusion.

The expectations underpinning this communication were:

1. The transparent presentation and synthesis makes the conclusions more acceptable, and thus more useful, for the different opinion holders.
 - a. Although some partner organisations felt that their project results were underappreciated, they had the opportunity to add information and opinions without dominating the variety of opinions. The inclusion of external experts (although they were only few in the restitution meeting) and organisations working on other themes in amending the conclusions also mitigated the tendency to overvalue results.
 - b. Partner organisations, often involved in just one project, appreciated the overview and focus on higher level programme outcome and impact. Some mentioned that they would like to repeat such an exercise at the programme planning stage.

To what extent has the IE been utilized and why/why not? What were critical influences and expectations during design (front of an IE) and communication (end of IE) that limited the use of the findings?

Design:

I estimate that the design (including non-participants, disaggregated presentation of opinions) gave room for more doubt and a more critical view on the programme achievements. This in turn resulted in stronger recommendations, stronger in the sense of adapting the ON programme in Burundi. Besides my own preference to work systematic and transparently, Oxfam Novib also stressed that the evaluation should meet the IOB evaluation assessment criteria. This certainly made the formulation of evaluation questions and judgment criteria more rigorous, and assured sufficient validation and triangulation. The design could have been stronger by including more directly targeted households in the survey. This would have yielded more programme results and stronger conclusions about the effect that Oxfam Novib made at household level. (This was important because of the lack of such survey results from Oxfam Novib or partner organisations). The costs of a larger sample size (local staff) are modest compared to the total evaluation costs.

Communication:

The restitution workshop went well, but could have been more valuable if more external experts were included (other NGOs, gov't) in the discussion of conclusions.

The ON management response (included in the report) has taken over some of the recommendations. Some points from the management response:

- More focus on fewer themes.
- Invest more in context and baseline studies at the start of a programme, jointly with other donors.
- More effort to include vulnerable groups within a community.
- Continue (successful) national-level lobbying.
- Integrate disaster risk reduction in programmes.
- Programme more open for local initiatives with a local contribution.

The follow up, during the planning of a new 5-year plan for Burundi 2011-2015, could have been enhanced if it were more closely linked to this evaluation. ON developed this almost simultaneously but independently from our evaluation. I only briefly participated. A few months later Oxfam Novib asked me to participate in a planning workshop, but unfortunately I was not available. It would have been good to take the evaluation as one of the starting points for the elaboration of a new programme. It would also have been good to invite other donors and external experts in the planning process, especially given some of the successful examples of collaboration between donors in the past. The risk of planning with the existing partner-organisation is a certain bias to continue what they were doing.

Appendix 3 – Panel members

The panel members were:

- Abubakar Muhammad Moki ('policy maker'), Ministry Of East African Community Affairs, Uganda
- Don Seville ('business'), Sustainable Food Laboratory, USA
- Ebere Joy Nneka Ubah ('implementing agency'), Action Aid Nigeria
- Kaidar Uali ('funder'), Soros Foundation, Kazakhstan
- Nirasha Amaratunga ('implementing agency - NGO'), LEADS, Sri Lanka
- Regina Josephine Akello ('implementing agency - GO'), National Agricultural Advisory Services (NAADS), Uganda
- Ruerd Ruben ('commissioner'), Radboud University Nijmegen & independent Policy and Operations Evaluation Department (IOB) at the Netherlands Ministry of Foreign Affairs.
- Ton Dietz (African Studies Centre; Leiden University; University of Amsterdam) the Netherlands

Appendix 4 – List of participants

Surname	First name	Organization
Keynote speakers		
Guijt	Irene	Learning by Design
Stern	Elliot	Lancaster University
Case owners		
Baatsen	Pam	Royal Tropical Institute
Bodnar	Ferko	Ministry of Foreign Affairs - NL
Chakraborty	Shovan	Swisscontact Katalyst
Cornelissen	Wilhelmus Jan	Erasmus University Rotterdam
De Groot	Dieneke	ICCO - NL
Ingram	Verina	Agricultural Economics Institute (LEI), Wageningen UR
Kornahrens	Claudia	Deutsche Gesellschaft Für Internationale Zusammenarbeit (GIZ)
Pibilova	Inka	Freelancer
Van Hemelrijck	Adinda	Institute for Development Studies (IDS)
Panel members		
Akello	Regina	Akello Regina Josephine
Amaratunga	Nirasha	Leads
Dietz	Ton	African Studies Centre
Moki	Abubakar	Ministry Of East African Community Affairs
Ruben	Ruerd	Ministry of Foreign Affairs - NL
Seville	Donald	Sustainable Food Lab
Uali	Khaidar	Soros Foundation Kazakhstan
Ubah	Ebere	Actionaid - Nigeria
Facilitators		
Baguma	Sylvester Dickson	NARO, Uganda
Brouwers	Jan	Centre for Development Innovation (CDI), Wageningen UR
Desalos	Caroline	Centre for Development Innovation (CDI), Wageningen UR
Kusters	Cecile	Centre for Development Innovation (CDI), Wageningen UR
Roefs	Marlene	Centre for Development Innovation (CDI), Wageningen UR
de Ruyter de Wildt	Marieke	Agricultural Economics Institute (LEI), Wageningen UR
Vugt, van	Simone	Centre for Development Innovation (CDI), Wageningen UR
Wigboldus	Seerp	Centre for Development Innovation (CDI), Wageningen UR
Zerfu	Elias	IFPRI, ESARO Ethiopia

Reporters

Chibvuma	Pauline	Wageningen University student
Dekker	Marco	Private consultant
Dorresteijn	Hilde-Marije	Wageningen University student
Janssen	Inge	Wageningen University student
Osharode	Modupe	Wageningen University student
Peters	Bram	Centre for Development Innovation, Wageningen UR
Rizopulos	Matilda	Centre for Development Innovation, Wageningen UR
Schrijver	Annemiek	Wageningen University student
Visser	Irene	Vis-à-vis Advies
Wolsky	Anja	Centre for Development Innovation (CDI), Wageningen UR

External participants

Arkesteijn	Marlen	Capturing Development
Baatsen	Pam	Royal Tropical Institute (KIT)
Back	Lucien Roger	
Bathily	Moussa	Centre Forestier De Recyclage A Thies
Berghmans	Mieke	Kuleuven
Bijman	Wilhelmus (Jos)	Wageningen UR
Blok	Meija Lucia	Royal Tropical Institute (KIT)
Bodnar	Ferko	Ministry of Foreign Affairs - NL
Boeren	Adrianus Jacobus Joseph Maria	Nuffic
Boshuizen	Mireille Kirsten	Ministry Of Economic Affairs
Makkink		
Cabello Arribas	Raquel	FAO
Chakraborty	Shovan	Swisscontact Katalyst
Chambille	Karel	Hivos - NL
Cornelissen	Wilhelmus Jan	Erasmus University Rotterdam
De Bruin	Irene Hildegard	Solidaridad
De Graaff	Donatien	Avance
De Groot	Dieneke	ICCO - NL
De Ruijter de Wildt	Marieke	Wageningen UR - SSG - LEI
Dietz	Ton	African Studies Centre
Getachew	Tarikua	CTA
Giani	Alberto	Wageningen UR - CDI
Gilhuis	Henk	UTZ Certified
Groverman	Verona	Verona Groverman Consultancy
Haagsma	Ben	Fair & Sustainable Advisory Services
Herens	Marion Christine	Wageningen UR
Herngreen	Anna Quirina	Wetenschappelijke Raad Voor Het Regeringsbeleid
Horsten	Jeroen	FMO
Hubeek	Francisca Birgit	Solidaridad
Ingram	Verina	Wageningen UR - SSG - LEI
Jacobs	Carolien Ida Maria	Wageningen UR
Kamm	Moniek Anne Augusta	Cos Gelderland, Partner In Sustainability And Diversity

Korevaar	Adriaan Laurens	Shape Your World
Kornahrens	Claudia	Deutsche Gesellschaft Für Internationale Zusammenarbeit (GIZ)
Kruseman	Gideon	Wageningen UR - SSG - LEI
Lakeman	John Douglas	Care International
Lewinsky	Thomas	MDF Training and Consultancy
Makleff	Shelly	IPPF/WHR
Marriott	Heidi	IPPF
Meerkerk	Alex	Pum Netherlands Senior Experts
Mensink	Nico	FMO
Monaghan	Joanna	Comic Relief
Morag	Fran Margaret	INASP
Ooms	Sonja Antonnetta	Oikocredit
	Cornelia	
Pibilova	Inka	Freelancer
Plavcak	Barbara	Wotro Science For Global Development
Plomp	Jacoline	Avance
Poel	Margriet	SNV Netherlands Development Organization - NL
Prins	Ester	Solidaridad
Ruben	Ruerd	Ministry of Foreign Affairs - NL
Scheewe	Selma	Royal Tropical Institute
Seville	Donald	Sustainable Food Lab
Tegels	Josephina	SNV Netherlands Development Organization - NL
	Jacoba Helena	
Ten Hoorn	Esther	Centraal Bureau Fondsenwerving
Uali	Khaidar	Soros Foundation Kazakhstan
Van Den Berg	Celine Laura	Leger des Heils
Van Der Graaf	Johannes	Leger des Heils
Van der Laan	Anita	SNV Netherlands Development Organization - NL
Van Gent	Marije	NCDO
Van Hemelrijck	Adinda	Institute For Development Studies
Van Noort	Lissy	Akvo Foundation
Van Trigt	Henk	Ministry of Foreign Affairs - NL
Van Vugt	Simone	Wageningen UR - CDI
Vazquez	Maria Jose	Fundacion Etea Para El Desarrollo Y La Cooperacion
Vlaminck	Zjos	African Studies Centre
Weijs	Bart	Wageningen UR
Woel	Birgitte	Sipu International

PPME-M4I course participants

Abraham	Rita	Eve Organization For Women Development
Abreha	Amleset Haile	Cascade
Abubakar	Yakubu	Bauchi State Agency For The Control Of HIV/AIDS, Tuberculosis/Leprosy & Malaria [BACATMA]
Aiyenigba	Bolatito	Fhi 360/Malaria Action Program For States Project
Akello	Regina	Akello Regina Josephine
Akpabio	Ebong	Centre For Health And Development
Al Qur'an	Randa	Ministry Of Planning
Alarango	Florence	Ministry of East African Community Affairs
Ali	Kimwaga	Zanzibar Aids Commission
	Muhiddin	
Amaratunga	Nirasha	Leads

Antwi	Kwaku	Ministry Of Food And Agriculture
Anusim	Manfred	National Emergency Management Agency
Ayun	Qurrota	Ministry Of National Development Planning
Bareyei	Johnson	National Planning Commission Commission
Belew	Mesfin	Sasakawa Africa Association/Sasakawa Global 2000-Ethiopia
Bisanga	Hillary	Uganda Prisons Service
Blaettler	Dominic	Bern University Of Applied Sciences
Camelia Tri		Partnership for Governance Reform (Kemitraan Bagi Pembaruan
Lestari		Tata Pemerintahan)
Chammag	Carmel	National Economic And Development Authority
Chinoona	Morgen	Leonard Cheshire (Disability) Zimbabwe Trust
Chirwa	Masauso	University Of Zambia
Chongo	Mangiza	Lusaka City Council
Dabon	Maria Gracia	National Economic and Development Authority VII
Dachew	Tesfaye	Amhara National Regional State Bureau of Agriculture
Dokurugu	Theophilus	Northern Ghana Network For Development
Dy	Sitha	World Vision Cambodia
Guneph	Hussein	Ministry Of Health And Social Welfare
Haider	Shamim	Rural Community Development Society
Haji	Hamid	Zanzibar Society For The Prevention Of Cruelty To Animals(Zspca)
Imam	Nawal	Ministry of Agriculture - SUDAN
Irema	Kahema	Ifakara Health Institute
Juma	James	Ministry Of Health And Social Welfare, National Aids Control Programme (NACP)
Juma	Biubwa Suleiman	Zanzibar Aids Commission
Kanem	Veronika	Skp Keuskupan Agung Merauke
Kasitila	John	Regional Secretariat - Tabora
Khamis	Mohammed Rashid	College of Health Sciences Zanzibar
Kiminta	Enock Senteu	Lake Naivasha Water Resource Users' Association
Kitingati	Happy	Ministry Of Agriculture Food Security And Cooperatives
Kyomugisha - Nuwagaba	Charity	Makerere University School Of Public Health
Mahmoud	Faiza	Ministry Of Water Resource
Mansour	George	Qader For Community Development
Maseki	Salome	Prime Minister's Office- Regional Administration And Local Government
Milambo	Carol	Zambia Open Community Schools
Mohammed	Mohammed	Zanzibar National Chamber Of Commerce, Industry And Agriculture (Znccia)
Moki	Abubakar	Ministry Of East African Community Affairs
Momoh	Mohammed	National Youth Service Corps
Monstersino	Enrica	Madera
Muhumuza	Catherine	Ministry Of Water And Environment
Mukasa	Charles	Uganda Human Rights Commission
Mwango	Mutale	World Vision Zambia
Mwidunda	Patrick Evariste	National AIDS Control Program
Mwimuka	Niyonsenga (Jimmy)	City Of Kigali
Ngussa	Peter	Arusha District Council

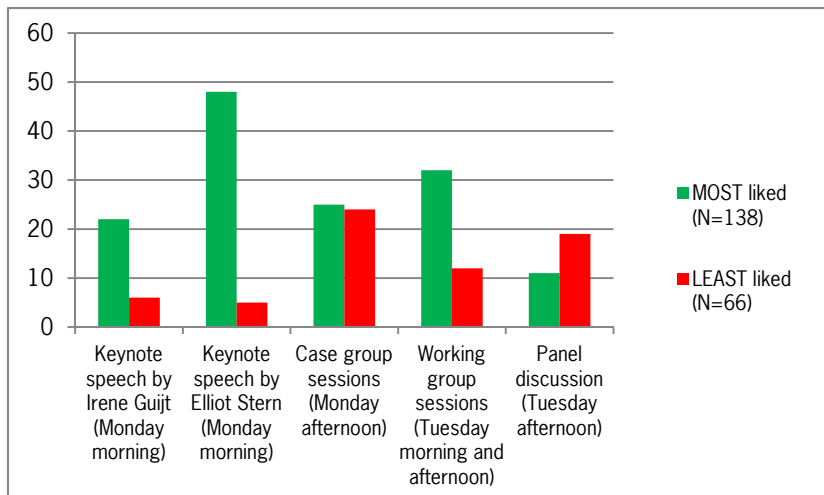
Nguyen	Mai Thi	Navi Communications and Community Development Co. Ltd. (NAVI DECOM)
Nyamukapa	Hilton	Streets Ahead Welfare Organisation
Nyanguwo	Manuel	Department Of Irrigation Development
Obafemi	Hassan	Kogi State Community And Social Development Agency
Odukoya	Atinuke	Centre for Women's Health and Information
Oduori	Praxides	Laikipia Wildlife Forum
Ogbogu	Adeline Chinyelu	Korea International Cooperation Agency
Okwuofu	Justin	Pro-Natura International (Nigeria)
Olalusi	Oluwafunmilayo	MOTT MACDONALD NIGERIA
Olode	Lanrewaju	Office Of Transformation, Strategy And Delivery Otsd, office of the Governor
Patel	Rajesh	J K Trust Gram Vikas Yojana
Permanasari	Yurika	Ministry Of Agriculture
Piya Shrestha	Resha	Alternative Energy Promotion Centre
Rijal	Kopila	Rural Development Services Centre
Sanga	Aneth Freddy	Kinondoni Municipal Council
Sanmugampillai	Balasubramania m	Balasubramaniam
Shahab	Zabihullah	Afghan Public Health Institute
Sharma	Tika	Project For Agriculture Commercialization & Trade (PACT)
Shirzai	Tooryalai	Ministry Of Agriculture, Irrigation And Livestock - Afghanistan
Shrestha	Yamuna	Centre for Peacebuilding and Reconciliation Promotion (CPRP)
Shrestha	Lal	Ministry of Agricultural Development - Kathmandu, Nepal
Simwa	Viviene	Monitoring And Evaluation Directorate
Ssemombwe	Joseph	Entebbe Municipal Council
Suratmiari		Bappeda
Tadlip	Maricel	Department of Environment and Natural Resources
Taqwaningtyas	Safitri	Agriculture Agency Of Semarang City
Teenstra	Edsger D.	Wageningen UR
Tilahun	Meseret	Debre Birhan Agricultural Research Centre
Tiwari	Durga	Development Society
Ubah	Ebere	Actionaid - Nigeria
Un Jan Liau Hing	Emilio	Universidad Norbert Wiener
Wellala Hetti	Dias	Ministry of Economic Development
Wubaye	Fekadu Mekuria	Ministry Of Civil Service-Engineering Capacity Building Program
Yeboah	Franklyn	Ministry of Food and Agriculture - GHANA
Yunah	Hamis	Lindi District Council

Appendix 5 – Evaluation of the conference

Just over half the conference participants (74 of 156) answered a set of evaluation questions, which was handed out on paper at the end of the second conference day. Questions asked were: (1) Which elements of the conference did you like most? (2) Could you explain why you liked the(se) elements most? (3) Which elements of the conference did you like least? (4) Could you explain why you liked the(se) elements least? (5) What highlights or lessons learned do you take home? (6) What suggestions do you have for improving a conference like this? and (7) What suggestions do you have for a topic for the next 'M&E on the cutting edge event' in 2014?

Most respondents mentioned several elements they liked most. Table 2 shows the results when multiple answers per person are taken into consideration. The most liked element was the keynote speech by Professor Elliot Stern, followed by the working group sessions on the role of design, and communication of impact evaluations (Tuesday morning and afternoon). Least liked elements were the case group sessions and the panel discussion. The audience had mixed feelings about the case group sessions, in which real-life examples of impact evaluations were presented and discussed. These sessions scored highest as least liked element, whereas an equal number of people indicated they liked them most. The panel discussion was least mentioned as most liked element. Several participants mentioned a preference for more 'real' stakeholders in the panel, and more time for an open discussion. One participant pointed to the defensive response mode panel members were pushed into by the way questions were asked, whereas these questions did not always resemble the advises formulated during the working group sessions.

Table 2. Which elements of the conference did you like most/least (more than one option possible; N=74)

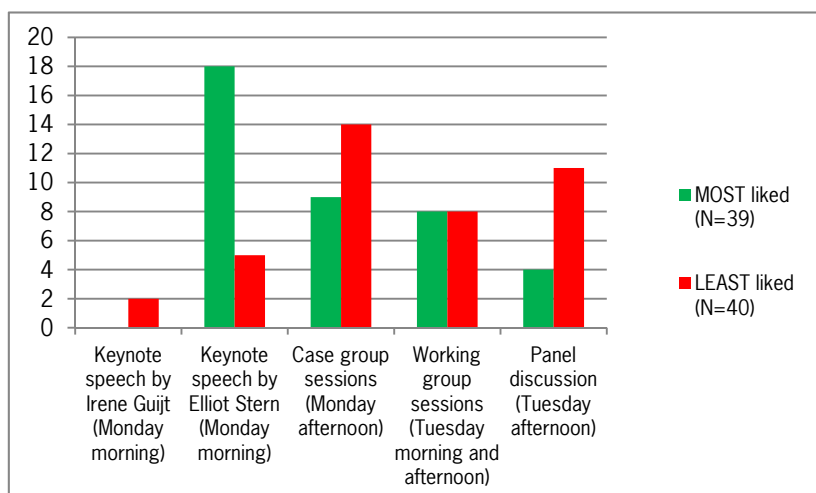


When we focus on those participants that mentioned just one element as most and least liked, the scores change as you can see in Table 3. The high scores for Dr. Irene Guijt and the working group sessions seem to have been primarily caused by participants who rated all elements as most interesting, as you can see when comparing tables 2 and 3. The keynote speech of Professor Elliot Stern is still the highest valued element. Participants especially liked the theoretical framework, overview of techniques and comparison of different IE approaches he presented. The case group sessions and the working group sessions are second highest valued elements. There was a huge appreciation for the group work and dialogue during both case studies and working group sessions. People liked to reflect on their own

situation, share experiences and the better understanding of IE this provided. The mixed feelings about the working group sessions change for a more negative score, whereas the appreciation for the case group sessions is now overshadowed by negative ratings.

Suggestions for improvements showed the same wave between requests for more time on case studies and having less case studies, and more key notes instead. Other suggestions include improving the preparation for the case studies (providing participants with more details prior to the conference, or using poster presentations), a more uniform way to present the different cases and a better focus on good practices in terms of connecting various international standards to IE and on the themes (design, communication, utilisation) instead of content. More time and better focus per case study could have increased in-depth discussion, and reduced the amount of stating the obvious. Several participants saw the working group sessions as a repetition of the case studies.

Table 3. Which elements of the conference did you like most/least (only those people who choose one option)



Insights gained by the participants were mainly related to the ways design, communication, and stakeholder participation can increase utilization of impact evaluations. The design stage of an impact evaluation is of utmost important. The role and way to include or address different key stakeholders, including primary beneficiaries, should be well-thought out. Different stakeholders often have different paradigms, expectations and ideas about the impact of a programme. Taking these into consideration in the design and communication strategy will benefit the utilization in the end. Stakeholder participation will increase the sense of ownership and thus of utilization. A communication strategy is an important element of the design of an impact evaluation, as is positively affects its utilization. IE should not be seen as an end in itself. If (the outcome of) an evaluation is not utilized well, it could as well not have been executed. There is no 'blueprint' for a good impact evaluation. The design of an IE should be driven by the research question, and not by the methodology.

Suggestions for next year's topic ranged from participatory IE to using mixed methods for IE, learning from IE experiences from other sectors (health, education, safety), building capacity for IE, impact monitoring (using e.g. sensemaker), and to using tools for communicating IE findings to better present evidence to policymakers.

Impact evaluation is crucial in current times of change, with increased attention for evaluation and an increasingly complex development sector. The aid architecture is changing, and power relations are shifting, both in the development world as in the professional evaluators world. Despite the increased attention for impact evaluation, its utilisation is yet to be improved.

This report summarises the presentations and discussions of the Conference 'Impact evaluation. Taking stock and looking ahead', which took place in Wageningen on March 25 and 26, 2013. The Conference was organised and funded by the Centre for Development Innovation (CDI), Wageningen UR in collaboration with Learning by Design; the Agricultural Economics Institute (LEI), Wageningen UR; and with funding from the Netherlands Ministry of Foreign Affairs (DGIS) and Hivos. The report describes current issues around impact evaluation (IE) and focuses on the following question that was central during the conference: 'What can be done in design and communication to enhance utilisation of IE findings?'

More information: www.wageningenUR.nl/cdi

