

A can of worms? Implications of rigorous impact evaluations for development agencies

Eric Roetman

March 2011



About 3ie

The International Initiative for Impact Evaluation (3ie) works to improve the lives of people in the developing world by supporting the production and use of evidence on what works, when, why and for how much. 3ie is a new initiative that responds to demands for better evidence, and will enhance development effectiveness by promoting better informed policies. 3ie finances high-quality impact evaluations and campaigns to inform better programme and policy design in developing countries.

3ie Working Paper series covers both conceptual issues related to impact evaluation and findings from specific studies or systematic reviews. The views in the paper are those of the authors, and cannot be taken to represent the views of 3ie, its members or any of its funders.

This Working Paper was written by Eric Roetman, International Child Support.

Photograph: Eric Miller / World Bank
3ie Working Paper Series Production Team: Radhika Menon and Mukul Soni

© 3ie, 2011

Contacts

International Initiative for Impact Evaluation
c/o Global Development Network
Post Box No. 7510
Vasant Kunj P.O.
New Delhi – 110070, India
Tel: +91-11-2613-9494/6885
www.3ieimpact.org

A CAN OF WORMS? IMPLICATIONS OF RIGOROUS IMPACT EVALUATIONS FOR DEVELOPMENT AGENCIES

Eric Roetman

International Child Support

Email: eric.roetman@ics.nl

Abstract

Development agencies are under great pressure to show results and evaluate the impact of projects and programmes. This paper highlights the practical and ethical dilemmas of conducting impact evaluations for NGOs (Non Governmental Organizations). Specifically the paper presents the case of the development organization, International Child Support (ICS). For almost a decade, all of ICS' projects in West Kenya were evaluated through rigorous, statistically sound, impact evaluations. However, as a result of logistical and ethical dilemmas ICS decided to put less emphasis on these evaluations. This particular case shows that rigorous impact evaluations are more than an additional step in the project cycle; impact evaluations influence every step of the programme and project design. These programmatic changes, which are needed to make rigorous impact evaluations possible, may go against the strategy and principles of many development agencies. Therefore, impact evaluations not only require additional resources but also present organizations with a dilemma if they are willing to change their approach and programmes.

1. Introduction

Nowadays, there is increased attention to result measurements in development co-operation. In the Netherlands, development co-operation is under great scrutiny by journalists and politicians who openly doubt its effectiveness. Some even advocate for limiting the sector to only emergency aid. Development agencies are expected to show that all their efforts lead to impact. As a consequence, there is increased attention to rigorous, statistically sound, impact evaluations. This paper is being written to show that rigorous impact evaluations¹ are more than an additional step in the project cycle; impact evaluations influence every step of programme and project design. These programmatic changes, which are needed to make rigorous impact evaluations possible, may go against the strategy and principles of many development agencies.

In this paper, ICS Africa serves as a case study. ICS² is a development organization that is founded in the Netherlands and has branches in East Africa and South East Asia. For many years, researchers from renowned universities such as Harvard University and Oxford University conducted impact evaluations within the ICS branch in West Kenya. The costs of these evaluations were mostly covered by institutions such as the World Bank. This approach has been praised by William Easterly in several of his articles (Easterly 2005, 2006). However, about ten years after it started, the practice of conducting in-house impact evaluations ended in 2005. The research and evaluation component was transferred from ICS to Innovations for Poverty Action (IPA), an organization that aims to both solve and evaluate social problems.

This paper is based on literature, interviews with ICS management and experiences of the author, who has himself been organizing various evaluations of development programmes since 2003. Although the author is currently employed by ICS, the management has given him permission to do research, draw independent conclusions and publish his findings. Therefore, the views expressed in this article are those of the author and do not necessarily represent the views of ICS, and should not be attributed to ICS. This paper is being written to feed into the current discussion on impact evaluations.

The paper consists of three parts. The first part describes how the impact evaluations within ICS Africa were conducted and the kind of conclusions that were drawn. The second part explains why this tradition ended and what the underlying issues were. The third part highlights some more universal issues, which are valid for NGOs in general.

¹ By rigorous what is meant is that evaluated projects are subject to the rigours of randomized control trials. This means that groups of project participants and control group are randomly chosen from a larger population. The results of a project are measured by making comparisons between the groups of participants and the control groups.

² The abbreviation ICS stood for Internationaal Christelijk Steunfonds. This Dutch name means International Christian Relief Fund. A few years ago ICS changed its name into International Child Support.

2. Impact evaluations within ICS

Evaluation Methodology

Duflo and Kremer (2003) led most of the impact evaluations within ICS and described their methodology in the article 'Use of Randomization in the evaluation of development effectiveness'. They refer to a method called 'randomization' or 'randomized evaluation methods', also sometimes referred to as 'field experiment' (Harrison and List 2004, Duflo 2006). Randomization consists of two important elements:

1. For every 'treatment group' there has to be a matching 'control group';
2. Treatment and control groups have to be chosen randomly to avoid selection biases;
3. The situation before (baseline) and after the intervention is assessed to make a comparison over time.

Many evaluations focus only on the participants of a project (in statistics these participants would be defined as the treatment group). By only focusing on this particular group, a crucial question is neglected: What would have happened with the participants if they would not have participated in the project? Therefore the first element of randomization is that a control group (sometimes referred to as comparison group) is needed to make a comparison between the participants and those who are not participating but living in similar circumstances.

A comparison between the treatment group and the control group is needed to make sure that outcomes can be attributed to the project and not to other causes. For example: higher retention rates in schools can be caused by an NGO that introduces a new teaching method. But the government's reduction of school fees also increases retention rates in all schools. By comparing retention rates of schools that participated in the project with retention rates of similar schools, a researcher can conclude if the increase in the retention rates was caused by the project.

The second element of randomization is that the participants of a project should be similar to those who are in the control group. To make a good comparison, the treatment group and the control group have to be identical, as far as possible (Humpreys and Weinstein 2009). Many NGOs are screening potential project areas or participants and base their choice on pre-defined characteristics, such as the willingness to participate. As a result, the treatment group and the control group are harder to compare, as there are already existing differences before the project begins. The best possible solution to avoid such a 'selection bias' is by selecting participants randomly. This is much easier for NGOs than it is for governments because NGOs are not expected to serve everybody.

To do a rigorous impact evaluation, participants have to be randomly grouped in a treatment group or control group before a project begins. Thus the methodology for the evaluation and the overall project design are intertwined and inseparable.

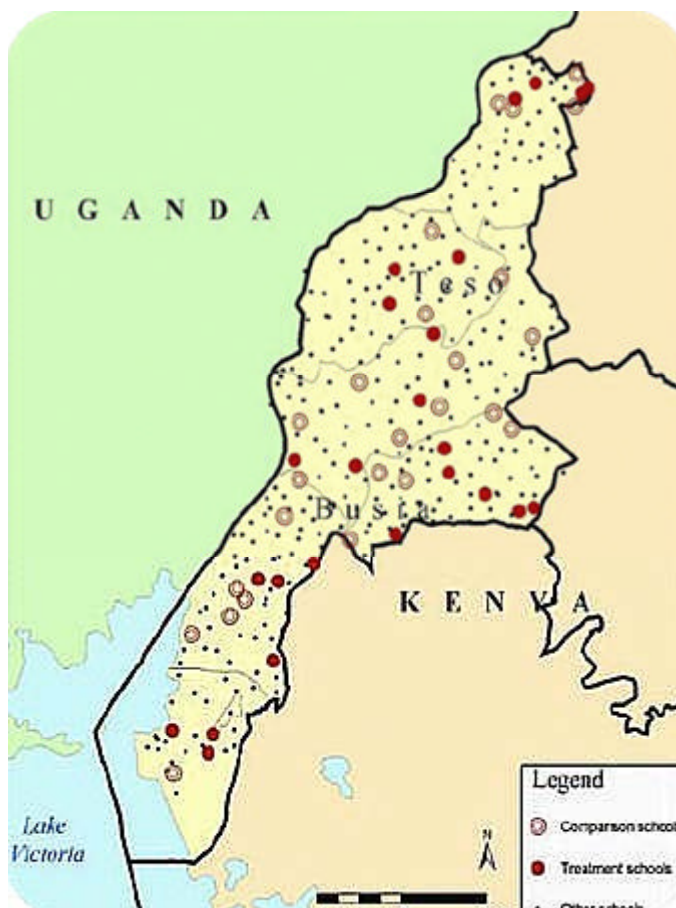
Examples of Impact Evaluations

In this paper two evaluations will be highlighted: the evaluation of a deworming project, which is the most well known evaluation of an ICS project, and an evaluation of several ICS projects that aimed to increase participation and build capacities.

Deworming

The evaluation of the deworming project was part of a bigger evaluation that looked at which interventions were most effective in increasing school participation.

All the ICS projects aimed at more children completing their education were evaluated. Examples are providing free school uniforms (Evans et al. 2009), textbooks (Glewwe et al. 2007) and school meals (Kremer et al 2004). However, the project that proved to be most effective was a deworming project.



Randomization (Kremer et al 2004)

It is very cheap to treat children with a deworming drug, but in spite of these low costs, many children are still chronically infected by intestinal worms. The symptoms are listlessness, diarrhea, abdominal pain and anemia. In fact, about one quarter of the world's children is infected with intestinal worms. In West Kenya, ICS began to treat 30,000 children in 75 primary schools with a deworming drug in collaboration with the Ministry of Health. The costs per child were only US\$0.50. These schools were randomly phased in, one by one. The effects were: 25 percent less absenteeism, especially of the youngest children, and 7 percent more children participating in school (Poverty Action Lab 2007). None of the other ICS projects had such an enormous effect for such a small investment (Duflo and Kremer 2003).

Hambaree – lets pull together

One of the evaluations was remarkable in that it compared three projects, each having a different degree and form of participation and capacity development. While two of the projects had the explicit aim of building social capital through capacity development, one did not. Social capital was understood along three lines: group solidarity within groups, labour and time commitment by group members, and stronger ties of groups to their wider communities. In hindsight, this was a remarkable evaluation because a few years later participation became a key principle of ICS.

The researchers first evaluated and compared two different projects:

- 1 The distribution of textbooks in schools (non-participatory);
- 2 Providing block grants to school management committees of schools (participatory).

In this comparison, the block grants did not turn out to be more effective in increasing social capital than the distribution of the textbook. The distribution of only school books had an overall positive effect on the indicators for school social capital and its organization (Gugerty and Kremer 2000).

The researchers also evaluated a third project aimed at building the capacity of women's groups. The group leaders received training in agriculture and management and the groups received agricultural inputs. It was expected that the social capital within the participating groups would be higher than the control groups.

It, however, turned out that the social capital did not increase, there were even signs it decreased. Although the participating groups received more visits from outsiders, social cohesion in the participating groups became less than in the control groups. The composition of the participating groups had also changed; there was a higher turnover rate; people were quicker in leaving and entering the groups. An increasing number of these groups began to charge entrance fees for these new members. The people who had entered and were willing to pay were more often those who had a salaried job and were higher educated. Moreover, an increasing number of men became members of the group. This also had its effects on leadership: the newly chosen leaders were more likely to be higher educated and men. Thus, the project had completely different effects than ICS had anticipated (Gugerty and Kremer 2002).

All three projects were critically appraised by the researchers. The last project even received a comment like: *"The substantial share of project funds that were spent bringing women's groups to seminars and meetings might have been better spent elsewhere"* (Gugerty and Kremer 2002). At the same time the evaluation posed broader questions for ICS: are projects that have the intention of increasing participation or capacity development (more) effective in increasing social capital? Thus the impact evaluations led to a constant critical reflection on ICS' projects as well as its overall approach.

Seekers

ICS' programmatic approach was described by Easterly as 'piecemeal solutions to poor people's problems'. When comparing this approach with the approaches of institutions, such as the World Bank and the International Monetary Fund, which mainly initiate large-scaled programmes, he concluded that large-scale development initiatives were ineffective and large institutions unaccountable. He referred to large infrastructural projects and national reforms, such as the Structural Adjustment Plans. According to him, all these plans were too fixed and lacked accountability and feedback mechanisms.

Instead of coming with a grand plan, organizations should work on a smaller scale and become more accountable; become so-called piece meal fixers, also referred to as seekers (Easterly, 2005 and 2006). He mentioned ICS as a positive example. A piecemeal fixer is an organization that takes one step at a time and decides how a programme will continue on the basis of monitoring and evaluations. By sharing the findings of these evaluations, a piecemeal fixer can be held accountable. To use the words of Easterly, *"Piecemeal fixers do not promise the miracles that planners do; they just quietly deliver results. Accountability for aid would transfer power from planners to fixers, both African and foreign. Maybe then we can hold some aid agency accountable to get the full value of those three additional dollars to each new mother"* (Easterly 2005).

3. Break with the Past

After about a decade, ICS ended its tradition of impact evaluations in 2005. Two years before this, a new international board, with fresh ideas, was installed. The new board decided, in consultation with the researchers, that the partnership for rigorous impact evaluations should come to an end. The part of ICS that made these evaluations possible would be phased out of ICS and phased into Innovations for Poverty Action (IPA), an organization specifically focused on research and rigorous impact evaluations. This chapter describes this history in greater detail and mentions the main reasons given why ICS ended its tradition of impact evaluations.

Evaluation as driver of programmatic change

In 1995, ICS began its collaboration with international researchers by evaluating a selection of projects. These projects were chosen out of ICS' programme in West Kenya. At that time the idea was simple: by evaluating projects, better informed decisions could be made. The external researchers organized funding for the evaluations, including financial support for ICS to cover the extra costs of these evaluations. The international board of ICS was very positive and considered this collaboration to be a great opportunity, a win-win situation. There was a sound but informal agreement between ICS' international board and the researchers: they could define their own research questions and design their own evaluations as long as they attract their own sources of funding.

In the following years, an increasing number of evaluations were organized within ICS. As a result, the evaluations became harder to manage. Bigger samples had to be taken; more schools and communities had to be involved. ICS had to expand its programme. And, on top of that, information had to be collected in the schools and communities that acted as control groups. The funds for this expansion were drawn by external researchers. This made it more difficult for the management to strictly separate the means and management of programmes and evaluations. The line between programmes and evaluations began to blur.

At its peak in 2004, ICS' branch in Kenya had a budget of about 1.2 million euros to spend on programmes and ICS had more than 80 people on its payroll, of which 70 were involved full time with evaluations. Except for a handful of people, everybody in the Kenya office was, one way or another, engaged with evaluations. On workdays field workers went to villages or schools to implement the programmes with a set of questionnaires to interview programme participants (and non participants). Another group processed these questionnaires by entering the gathered information into a database. Statisticians were checking the databases and doing statistical calculations. At night the researchers were in communication with the United States to discuss methodological issues and findings.

It had not been a problem to gain funds for new impact evaluations. But what increasingly became a problem was the identification of new subjects for the evaluations, that is new projects to evaluate. As a result, there was a lot of pressure on the 1.2 million budget to keep initiating new types of projects that were interesting to research. The choice of research questions had implications for programmatic choices and vice versa. As more evaluations were done, it became harder to fit these projects neatly into ICS' program agenda.

The evaluations were generating an overload of information but this information was not always serving the organization. After an evaluation was initiated, it took a few years before

conclusions could be shared. In the meantime, ICS had often changed its strategy. Not all the evaluations were useful for the organization anymore. ICS had become a development lab, allowing others to upscale projects. This approach no longer matched ICS' simple objective to do evaluations for informed decision-making.

New board, New ideals

Most board members of ICS, including the chairman and the executive director, were replaced within the year 2003. This change marked a discontinuity in the history of ICS. While the previous board viewed impact evaluations mainly as an opportunity, the new board was more critical of how the evaluations were organized. They emphasized that the management of ICS should have more control over its own programmes. The line between programmes and evaluations had become too thin. The evaluations had to be organized separately from the programmes.

The new board thought that rigorous impact evaluations hindered genuine participation: people in schools and communities had too little influence on the design of the projects. Participants were only chosen when the project was already designed, planned and indicators were defined. Instead they wanted to navigate through participants' viewpoints and contributions. This proved to be difficult when rigid evaluations were planned, because the planning was hard to adjust after the baseline. Participation was considered to be crucial; it was not viewed as just an effective way of working. It was a driving principle. It was said that participatory projects can fail (as the example in the previous chapter shows) but such incidents were related to the design of the project, not to the principle of participation. The researchers made participation the subject of evaluations and openly discussed whether participation was effective at all. This clearly shows the different ideas that the management and the researchers had of the future direction of ICS.

Phase out

There was no long term agreement between the universities and ICS when the future of impact evaluation within ICS was being discussed. The management and the researchers talked for many hours and jointly decided to phase out the evaluation support from the organization. This part of ICS was phased into the US non-profit organization 'Innovations for Poverty Action', which made it possible to continue the impact evaluations in Kenya. ICS itself continued as a development organization, which no longer had an exceptional emphasis on impact evaluations and result measurements.

'Civic driven change' became more central in ICS' strategy in 2006. It was defined in ICS' policy that participation could only be effective if a project was driven by people. Instead of asking people to participate in ICS' projects, it would be the other way around: ICS wanted to strengthen already existing local initiatives. From that time onward, ICS went into a new direction. No attempts were made to do a randomized evaluation of ICS' new programme.

IPA Kenya has since developed itself to be a true development lab. In five years, the organization has employed nearly 200 staff members and is doing many research projects, mostly in West Kenya. IPA is considered to be a leader in randomized control trials and has used this methodology to investigate many complex research topics. It functions as a development lab, evaluating topics, and allowing others to upscale. It also supports other organizations and governments to upscale projects that were rigorously evaluated.

ICS and IPA went in separate directions. A staff member of ICS Africa said: *"Just after the separation many of us felt pessimistic. But now, five years later, I can clearly see that IPA and ICS both have grown and developed themselves strongly. This could never have happened if this was still one organization"*. Thus in hindsight, the phase out and the foundation of IPA Kenya was inevitable. ICS could never have retained a research institute that was growing faster and bigger than the organization itself.

The researchers, staff and management who carried out the impact evaluations and facilitated the research achieved many results: many brilliant articles, the foundation of IPA? a renowned research institute - and fame for ICS Africa. Above all, internal tensions between programmes and evaluation laid the groundwork that allowed ICS and IPA Kenya to blossom in the following years.

4. Discussion

The case of ICS is unique because it was one of the first NGOs that conducted impact evaluations on a regular basis. Some of the underlying issues are more universal and relevant for the debate about the use of rigorous impact evaluations. In this chapter, the lessons learnt are explored for further discussion. The main argument is simple and straightforward: rigorous impact evaluation is not the gold standard for monitoring and evaluation practices of NGOs. The timelines, demands and requirements of research are not responsive to the immediate and medium term needs of NGOs. Thus there are many reasons for NGOs to rely on other methods. However, impact evaluation is a very valuable methodology that contributes to organizational learning and improves accountability of many programmes.

Resources

If the evaluations within ICS had not been done on the basis of 'randomization' but on the basis of a methodology that only evaluated a project in retrospect, the conclusions would have been weaker. As the projects were rigorously evaluated, the conclusions have shown a value beyond ICS and served NGOs and governments worldwide in making programmatic choices.

What this paper highlights is that rigorous impact evaluations deliver valuable findings but also require many resources and are difficult to organize for a relatively small NGO. It is questionable if rigorous impact evaluations should be organized by an NGO on its own. This is because impact evaluations are costly and it is difficult to develop an organization in a way that there is a good balance between programmes and evaluations. Therefore organizations are recommended to join forces with likeminded organizations before starting collaborations with universities and knowledge centers. Instead of evaluating the results of an individual programme, more fundamental issues can be jointly evaluated. For example: frequently made assumptions in 'theories of change' or log frames. By acting together, NGOs can identify these broader issues, take into account lessons learnt from organizing rigorous evaluations, and share the costs and benefits.

Role of NGOs

Before thinking of the opportunities and challenges of rigorous impact evaluations, attention should be given to why NGOs should start doing rigorous evaluations. This is a question that seems to be easy to answer: rigorous evaluations deliver valid conclusions and recommendations to organizations. Based on this information, NGOs can organize projects

that generate valuable results. But this answer neglects an important fundamental question: what is the role and added value of an NGO? Why does an organization want to have its results rigorously evaluated?

There are very few NGOs that think their role is to be the deliverer of services. In most cases this is thought of as a task of the government. For example: deworming has had a great effect on child healthcare and school attendance. But this effect could only be sustained if children received deworming pills on a regular basis. It was the Kenyan government that was in the best position to take up this responsibility. *"Development NGOs... are increasingly recognising that the denial of people's rights often lies at the root of many of the problems that the NGOs aim to address"* (Harris-Curtis, Marley and Bakewell 2005).

Another way to think of the added value of an NGO is by viewing these organizations as development labs: NGOs do short term interventions and evaluate the results, leaving it to others to upscale successful interventions. This is essentially what ICS Africa did and what IPA is currently doing. This is definitely a legitimate and valuable role but many NGOs view their role differently. Different NGOs perform different roles.

It is important to think of the roles of NGOs because rigorous impact evaluations deliver very specific information: to what degree was a project successful in meeting its objectives? This is mainly information about the outcome of projects, not its process. This information also has a price. The methodology of a programme is hard to change after a baseline. On top of that, it takes a few years before the conclusions can be shared. There is also the issue of scale. To use a simple metaphor: a local bakery does not need a research and development department. And a small scale NGO may not need to have its results rigorously evaluated. If an NGO does think of itself as a development lab, the costs and constraints of a rigorous evaluation will often be much higher than the benefits.

Clashing Principles

"You have a unique personality, just like everybody else"

Although some people and organizations within the Dutch development sector are advocating for result measurement, few Dutch NGOs have organized rigorous impact evaluations. There is resistance to this type of evaluation. The first argument is that there is less flexibility as rigorous evaluations limit the space for organizations to adapt a program to unexpected changes. The second argument relates to the statistical rigour of rigorous impact evaluations: *"Many real-world problems are not easily described with the kind of precision that professional mathematicians insist upon. This is due to the limitations of data, the costs of collecting and analyzing data, and the inherent difficulties of giving mathematical expression to the complexity of human behavior"* (Lawry 2010).

Not all development agencies are willing to organize their projects like a grand field experiment to test their interventions. This is especially true for impact evaluations based on 'randomization'. Although Dutch development agencies follow different approaches and have different ideas about organizing evaluations, there are two strong opposing views that dominate the debate about result measurement. The first view focuses mainly on the outcomes and the second on the process. These two positions can be organized under two broader ethical principles: 'consequentialism' and 'relational ethics'.

The first is consequentialism, which means that a morally right action is one that produces a good outcome. Rigorous impact evaluations neatly fit in that line of thinking. Programmatic

changes in development co-operation to make rigorous impact evaluations possible are thought of as a more professional way of operating. The second way of thinking is relational ethics, which is built on the premise that morally right actions come out of a process of joint learning. Thus, morally right actions come out of an informed consent by all stakeholders. Not surprisingly, rigorous impact evaluations do not fit well in that line of thinking because there is an emphasis on uniqueness, complexity and perceptions. Programmatic changes to make monitoring and evaluations possible are accepted to the extent that it is not hindering the joint learning processes with stakeholders.

Results	Processes
Based on the ethics of Consequentialism.	Based on relational ethics
Emphasis on commonalities, efficiency, effectiveness and impact.	Emphasis on uniqueness, emergence, complexity, and effects on relationships and perceptions.
Participation is a means (and has to be compared with other means).	Participation (or co-creation) is a principle.
You need proof: outcomes should be measurable. Perceptions of stakeholders are no indicators of success.	'Not everything that counts can be counted'. Perception is a crucial indicator of success.
Learning follows accountability (thus a strong emphasis on accountability).	Accountability follows learning (thus a strong emphasis on learning).
Accentuates that accountability is a public affair: organizations should act responsible, be transparent and answerable to society for their use of resources.	Underscores that the highest form of accountability is internal: staff within organizations should 'walk their talk'.

The different viewpoints become most clear in discussions about treatment and control groups. From the perspective of consequentialism, the fairest way of choosing who can participate in a project is by randomly selecting participants for the treatment and control group. It is like a lottery in which personal preferences and relations do not influence the selection process. From the perspective of relational ethics there are objections to this selection process. People who are in the control group are being interviewed and informed about the project but they are not included in the project. This could make people feel excluded, leading to frustration and even discontent.

Development organizations are put under pressure by donor agencies to measure results but are often unwilling to give up their own value sets (which are often more or less in line with relational ethics). These organizations find themselves doing a balancing act; they are trying to adhere, more or less, to the ethical positions of consequentialism and relational ethics. This does not necessarily lead to an improved evaluation practice. It is, above all, important not to attempt to create synergy between these two positions but rather create a healthy tension by acknowledging the added value and limitations of each ethical position.

Seeking the Piecemeal Fixers

Easterly referred to ICS as a Piecemeal Fixer or Seeker because the organization was willing to move step by step and made its decisions based on monitoring and evaluation findings. By being transparent about this process, ICS was accountable to itself and its stakeholders. This could be interpreted as an argument for all NGOs to become development labs. Only organizations that have projects externally evaluated are truly accountable. A broader interpretation is that Easterly referred to ICS as a Piecemeal Fixer because of the quality of the studies or the type of methodology used. He could also have identified ICS as a

Piecemeal Fixer because it dared to be scrutinized by independent researchers who even published critical articles on ICS' performance in academic journals and online magazines. To phrase it differently, perhaps he mentioned ICS as a good example simply because of the attitude to seek, learn continuously and be fully transparent about its activities and the intended and unintended results.

In my personal view, the best Piecemeal Fixers or Seekers are those organizations that can create healthy tension, thinking in terms of outcomes and processes. These are organizations that have clearly defined their principles for monitoring and evaluations, taken both processes and outcomes into account, based their approaches on the lessons learnt and are transparent and answerable about this.

Paradox of Impact Evaluations

The current debate about impact evaluations is, above all, about measurability and methodologies. But this paper has drawn attention to the broader picture, making an argument to debate the relevance and attainability of impact evaluations or any other form of result measurement. Patton (2009) addressed this issue by asking if "*evaluation is designed and implemented in a way that really makes a difference to improving programs and improving decisions about programs?*" I would like to add to this question: 'are evaluations designed and implemented in such a way that development agencies really become more accountable to societies, both in the North and the South?'

In the Netherlands, the argument made by some is that rigorous impact evaluations are needed to change the public discourse about development co-operation. It is assumed that such evaluations will contribute to proving the effectiveness of development co-operation and quieten the critical voices. There is, however, no strong evidence for this way of thinking. The paradox is that the impact of rigorous impact evaluations on public discourse has not been rigorously evaluated.

5. Epilogue: ICS' Road Ahead

ICS has adopted Social Return on Investment (SROI) as a method for planning and evaluating projects and programmes. In this method, economic, social and environmental costs and benefits of a project are compared and valued. In 2009, ICS mainly relied on SROI to facilitate communities in defining their main concerns. From 2010 onwards, the analyses will be further disseminated for linking and learning.

Social Return on Investment

Social Return on Investment (SROI) is an innovative method that can be used as a tool for planning, monitoring and evaluating projects. In a SROI analysis, all key stakeholders are included: they jointly attribute values to the economic, social and environmental costs and revenues of a project. The analysis is finalized by comparing the total costs and total revenues, and drawing conclusions from that comparison.

In an SROI analysis, qualitative and quantitative information is gathered and analysed. The quantitative data is used to calculate the SROI ratio. To calculate this ratio, quantitative data has to be collected about the economic, social and environmental costs and benefits of all key stakeholders.

Qualitative data is collected by gaining a more in-depth understanding of why stakeholders have captured certain costs and revenues. Qualitative methods, such as 'most significant change' and story-telling exercises are used to capture these viewpoints. Thus, both numbers and words count in an SROI analysis. An SROI analysis is conducted in the planning phase and after the project is completed.

Process versus Outcome thinking

SROI does not promise to 'solve' the dilemma between process-thinking (relational ethics) and thinking in terms of outcomes (consequentialism). In fact, this has been an ongoing debate amongst researchers and practitioners who are involved with SROI in the last few years. The advantage of SROI, however, is that it is possible to put the emphasis on the process and the outcomes. SROI can be applied as a learning process with stakeholders by focusing on perceived values. On the other hand, it is also possible to focus more on outcomes and calculations based on research material. For example: in the planning phase of a vocational training programme, the estimated revenues can be based on the perception of local stakeholders (teachers, government officials, parents and students) or calculated by looking at what the benefits were of similar vocational training programs elsewhere. The strength of SROI is that it allows organizations to rely on both types of information, as long as the methodological choices are made clear in the final report and widely shared. Thus ICS applies SROI to create and maintain a healthy tension between thinking in terms of process and thinking in terms of outcomes.

More information on SROI can be found on the website of Context, international co-operation (www.developmenttraining.org) or Social Evaluator (www.socialevaluator.eu)

6. Bibliography

- Duflo, E & Kremer, M. "Use of Randomization in the Evaluation of Development Effectiveness." Washington D.C.: Worldbank, 2003. 35.
- Duflo, E. "Field Experiments in Development Economics." *Prepared for the World Congress of the Econometric Society*. Massachusetts: Massachusetts Institute of Technology, 2006. 30.
- Easterly, W. *The white man's burden: why the West's efforts to aid the rest have done so much ill and so little good*. Penguin Press, 2006.
- . *Tone Deaf on Africa*. New York: New York Times, 2005.
- Evans, Kremer, Ngatia . *The Impact of Distributing School Uniforms*. www.poveryactionlab.org: Poverty Action Lab, 2009.
- Eyben, R. "The political and ideological context of assessing and reporting on making a difference in development." *For the confrence: Evaluation Revisited: Improving the quality of evaluative practice by embracing complexity" – a Conference on Evaluation for Development*. <http://evaluationrevisited.wordpress.com/>, 2010.
- Glewwe, P & Kremer, M & and Moulin, S. *Many Children Left Behind? Textbooks and Test Scores in Kenya* . www.poveryactionlab.org: Poverty Action Lab, 2007.
- Glewwe, P & Kremer, M. & Moulin, S & Zitzewitz E. "Retrospective vs. prospective analyses of school." *Journal of Development Economics* 74 (2004), 2004: 251– 268.
- Gugerty, M. & Kremer, M. "The Impact of Outside Funding on Community Organizations - The Organizations of the Disadvantaged." www.polisci.ucla.edu/wgape/papers/1_Gugerty.doc, 2002.
- Harris-Curtis, E. Marleyn O. Bakewell, O. "The Implications for Northern NGOs of Adopting Rights-Based Approaches." *Occasional Papers Series No: 41*. INTRAC, 2005. 52.
- Humphreys, M. & Weinstein, J. "Field Experiments and the Political Economy of Development." *Annual Review of Political Science*, 2009: 367-378.
- Kremer, M & Vermeersch, C. *School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation*. www.poveryactionlab.org: Poverty Action Lab, 2004.
- Kremer, M. & Gugerty, K. "The Impact of Development Assistance on Social Capital: Evidence from Kenya." In *The Role of Social Capital in Development- An Empirical Assessment*, by C. Grootaert, 213-233. The World Bank, 2002.
- Lab, Poverty Action. "Mass Deworming: A Best -Buy for Education and Health." *Poverty Action Lab: Policy Briefcase 4*, 2007.
- Lawry, S. "When Too Much Rigor Leads to Rigor Mortis: Valuing Experience, Judgment and Intuition in Nonprofit Management." <http://hausercenter.org/iha/2010/07/12/when-too-much-rigor-leads-to-rigor-mortis-valuing-experience-judgment-and-intuition-in-nonprofit->

management/. Harvard: Harvard University: The Hauser Center For Non Profit Organisations, 2010.

Patton. "In Conversation: Michael Quinn Patton." *Interviewed by IDRC*.
http://www.idrc.ca/en/ev-30442-201-1-DO_TOPIC.html: IDRC website, 2002.

Publications in the 3ie Working Paper series

Behind the scenes: managing and conducting large scale impact evaluations in Colombia by Bertha Briceño, Laura Cuesta and Orazio Attanasio, Working Paper 14, December 2011

Can we obtain the required rigour without randomisation? by Karl Hughes and Claire Hutchings, Working Paper 13, August 2011

Sound expectations: from impact evaluations to policy change by Vanessa Weyrauch and Gala Díaz Langou, 3ie Working Paper 12, April 2011

A can of worms? Implications of rigorous impact evaluations for development agencies by Eric Roetman, 3ie Working Paper 11, March 2011

Conducting influential impact evaluations in China: the experience of the Rural Education Action Project by Mathew Boswell, Scott Rozelle, Linxiu Zhang, Chengfang Liu, Renfu Luo, Yaojiang Shi, 3ie Working Paper 10, February 2011

An introduction to the use of randomized control trials to evaluate development interventions by Howard White, 3ie Working Paper 9, February 2011

Institutionalisation of government evaluation: balancing trade-Offs by Marie Gaarder and Bertha Briceño, 3ie Working Paper 8, July 2010

Impact Evaluation and interventions to address climate change: a scoping study by Martin Prowse and Birte Snilstveit, 3ie Working Paper 7, March 2010

A checklist for the reporting of randomized control trials of social and economic policy interventions in developing countries by Ron Bose, 3ie working paper 6, January 2010

Impact evaluation in the post-disaster setting by Alison Buttenheim, 3ie Working Paper 5, December 2009

Designing impact evaluations: different perspectives, contributions from Robert Chambers, Dean Karlan, Martin Ravallion, and Patricia Rogers, 3ie Working Paper 4, July 2009. Also available in Spanish, French and Chinese

Theory-based impact evaluation by Howard White, 3ie Working Paper 3, June 2009. Also available in French and Chinese.

Better evidence for a better world edited by Mark W. Lipsey University and Eamonn Noonan, 3ie & The Campbell Collaboration, 3ie Working Paper 2, April 2009

Some reflections on current debates in impact evaluation by Howard White, 3ie Working Paper 1, April 2009

For the latest 3ie working papers visit:
http://www.3ieimpact.org/3ie_working_papers.html