

MFS II EVALUATIONS

Joint evaluations of the Dutch Co-Financing System 2011 - 2015

*Civil Society contribution
towards achieving
the Millennium Development Goals*

Synthesis report

8 country studies

July 2015

CONFIDENTIAL

Please be aware that this report is still confidential, pending an assessment of whether its contents could put at risk the safety of any partners mentioned. Once this assessment is completed the report, can be made public.

SGE Stichting Gezamenlijke Evaluaties



Netherlands Organisation for Scientific Research
WOTRO Science for Global Development



MFS II Joint Evaluations

Final Synthesis Report

Jacques van der Gaag, Jan Willem Gunning and Gerton Rongen

16 June 2015

Acknowledgements

The authors wish to thank Lene Böhnke, Stavros Malamas, Kimberley Wallaart and Alejandro Zerain for their assistance in completing this report.

PREFACE

This report is one of a series of evaluation reports, consisting of ten reports in total, reflecting the results of the jointly-organised MFS II evaluation:

- eight country reports (India, Bangladesh, Ethiopia, Uganda, Indonesia, DR Congo, Liberia, Pakistan);
- a synthesis report (covering the eight country studies); and
- a report with the results of the international lobbying and advocacy programmes.

This series of reports assessed the 2011-2015 contribution of the Dutch Co-Financing System (MFS II) towards achieving the Millennium Development Goals, strengthening international civil society, setting the international agenda and changing decision-makers' policy and practice, with the ultimate goal of reducing structural poverty. On July 2nd, 2015, the reports were approved by the independent steering committee (see below), which concluded that they meet the quality standards of validity, reliability and usefulness set by the Ministry of Foreign Affairs.

MFS II has been the 2011-2015 grant framework for Co-Financing Agencies (CFAs). A total of 20 alliances of Dutch CFAs were awarded € 1.9 billion in MFS II grants by the Ministry of Foreign Affairs. CFAs receiving MFS II funding work through partnerships with Southern partner organisations supporting a wide range of development activities in over 70 countries and at the global policy level.

The MFS II framework required each alliance to carry out independent external evaluations of the effective use of the available funding. These evaluations had to meet quality standards in terms of validity, reliability and usefulness. The evaluations had to focus on four categories of priority result areas, as defined by the Ministry of Foreign Affairs, and comprise baseline assessments serving as a basis for measuring subsequent progress.

Out of the 20 alliances receiving MFS II funding, 19 decided to have their MFS II-funded activities evaluated jointly. These 19 alliances formed the *Stichting Gezamenlijke Evaluaties (SGE)*¹, which acted on their behalf in relation to the joint MFS II evaluation. The SGE was assisted by an 'Internal Reference Group', consisting of seven evaluation experts of the participating CFAs.

The Netherlands Organisation for Scientific Research (NWO/WOTRO) managed the evaluation and selected ten research teams to carry out the joint MFS II evaluation: eight teams responsible for carrying out studies at country level, one team responsible for the synthesis of these country studies, and one team responsible for the study of international lobbying and advocacy. Each study comprises a baseline assessment (2012) and a final assessment (2014). Research teams were required to analyse the effectiveness, efficiency and relevance of development interventions funded by MFS II. An independent steering committee was appointed to verify whether the studies met with the required quality standards. In its appraisal, the steering committee drew on assessments by two separate advisory committees.

¹ *Stichting Gezamenlijke Evaluaties can be translated as Joint Evaluation Trust.*

The evaluation has been implemented independently. The influence of the CFAs was limited to giving feedback on the first draft reports, in particular to correct inaccuracies. The contents and presentation of information in this report, including annexes and attachments, are therefore entirely the responsibility of the research team and/or NWO/WOTRO.

However, as SGE we are responsible for adding this preface, the list with parties involved and a table of contents, in the cases that the report is a compilation of several reports.

In addition we would like to note that when reference is made to individual case studies, this should be seen as illustrative examples, rather than as representative of a CFA's entire partner portfolio.

The Dutch CFAs participating in this unique joint evaluation are pleased that the evaluation process has been successfully completed, and thank all the parties involved for their contribution (see the next pages for all the parties involved). We hope that the enormous richness of the report will serve not only accountability but also learning.

Bart Romijn

Chair of the 'Stichting Gezamenlijke Evaluaties'

c/o Partos
Ellermanstraat 18B
1114 AK Amsterdam
www.partos.nl
info@partos.nl

Participants MFSII Evaluation

The SGE¹ alliances (and their lead organisations)

Freedom from fear (PAX)
United Entrepreneurship Coalition (Spark)
Impact Alliance (Oxfam Novib)
Communities of Change (Cordaid)
WASH Alliance (Simavi)
People Unlimited 4.1 (Hivos)
Dutch Consortium for Rehabilitation (ZOA)
SRHR Alliance (Rutgers)
ICCO Alliance (ICCO)
Connect4Change (IICD)
Connect Now (War Child)
Woord en Daad & Red een Kind Alliance (Woord en Daad)
Together4Change (International Child Support)
Child Rights Alliance (PLAN Netherlands)
Ecosystem Alliance (IUCN)
Partners for Resilience (Nederlandse Rode Kruis)
Press Freedom 2.0 (Free Press Unlimited)
Fair Green and Global (Both ENDS)
Kind en Ontwikkeling (Terre des Hommes)

The SGE Board

Bart Romijn (Partos/chair)
Ben Witjes (Hivos)
Harry Derksen (New World Campus)
Jan Lock (Woord en Daad)
Dianda Veldman (Rutgers)
Lucia Helsloot (Partos)
Marouschka Booy (independent consultant)
Mirjam Locadia (Partos, until 09/2014)
Alexander Kohnstamm (Partos/chair, until 10/2013)

The SGE Internal Reference Group

Yvonne Es (Oxfam Novib)
Rens Rutten (Cordaid)
Karel Chambille (Hivos)
Peter Das (ZOA)
Ruth van Zorge (Rutgers)
Dieneke de Groot (ICCO)
Wouter Rijnveld (independent)

¹ Stichting Gezamenlijke Evaluaties

NWO/WOTRO, the Netherlands Organisation for Scientific Research

Dr. Martijn Wienia
Dr. Gerrie Tuitert
Sabine Zinsmeister
Dr. Henk Molenaar (until 11/2014)
Dr. Barbara Plavcak (until 09/2013)

Steering Committee

Prof. Wiebe Bijker (chair), Maastricht University, the Netherlands
Dr. Gavin Andersson, Seriti Institute, South Africa
Prof. Anita Hardon, University of Amsterdam, The Netherlands
Prof. John Rand, University of Copenhagen, Denmark
Ms. Mallika R. Samaranyake, independent consultant, Sri Lanka
Dr. Zenda Ofir, independent consultant, South Africa (until 10/2012)

Advisory Committee for the Country Level and Synthesis studies

Prof. James Copestake (chair), University of Bath, Department of Social and Policy Sciences, UK
Prof. Arnab Acharya, OP Jindal Global University, School of Government and Public Policy, India & London School of Hygiene and Tropical Medicine, UK
Prof. Bagele Chilisa, University of Botswana, Botswana
Prof. Christopher Cramer, University of London, Department of Development Studies, UK
Dr. Nathalie Holvoet, University of Antwerp, Institute of Development Policy and Management, Belgium
Dr. Janice Jiggins, Independent consultant & guest researcher Wageningen University & Research Centre, Communication Science, the Netherlands
Dr. Stefan Leiderer, Deutsches Institut für Entwicklungspolitik (DIE), Department of Bilateral and Multilateral Development Cooperation, Germany
Prof. Desmond McNeill, University of Oslo, Centre for Development and the Environment, Norway
Dr. Howard White, International Initiative for Impact Evaluation 3ie, USA
Mr. Burt Perrin, Independent consultant, France – participated in first two calls
Dr. Jim Rugh, UK – participated in first two calls
Prof. Tom Schwandt, University of Illinois at Urbana-Champaign, USA – participated in first two calls

Synthesis team

Prof. Dr. J.W. Gunning
Prof. Dr. J. van der Gaag
Gerton Rongen

Principal Investigators/Evaluators (per country)

India: Prof. dr. B.W. Lensink, University of Groningen, The Netherlands
Liberia: Prof. dr. ir. E.H. Bulte, University of Wageningen & Research Centre, The Netherlands
Indonesia: Prof. M.P. Pradhan, University of Amsterdam, The Netherlands
Pakistan: J.P. Tranchant, Institute of Development Studies, UK
DRC/Uganda: Prof. dr. ir. E.H. Bulte, University of Wageningen & Research Centre, The Netherlands
Bangladesh: Dr. J. Devine, University of Bath, UK
Ethiopia: Prof. dr. B.W. Lensink, University of Groningen, The Netherlands

Table of Contents

Executive Summary.....	vii
1 Introduction	12
1.1 Structure of the evaluation.....	12
1.2 Structure of the report.....	14
2 Results per priority area.....	16
2.1 MDGs and themes.....	16
2.1.1 MDG 1 Eradicate extreme poverty and hunger	17
2.1.2 MDG 2 Achieve universal primary education	23
2.1.3 MDG 3 Promote gender equality and empower women	27
2.1.4 MDG 4, 5, 6 Reduce child mortality rates, improve maternal health & combat HIV/AIDS, malaria, and other diseases	28
2.1.5 MDG 7a, b Integrate the principles of sustainable development into country policies and programs; reverse loss of environmental resources and reduce biodiversity loss	31
2.1.6 MDG 7c Water and sanitation	33
2.1.7 Good governance, civil society and fragile states.....	35
2.1.8 Conclusion.....	37
2.2 Capacity development	39
2.3 Civil society strengthening.....	50
2.3.1 Changes in CSI scores.....	50
2.3.2 Attribution of outcomes	58
2.4 Summary.....	61
3 Synthesis and analysis of results.....	62
3.1 Countries.....	62
3.2 Priority areas.....	73
3.2.1 MDGs and themes.....	73
3.2.2 Capacity development	75
3.2.3 Civil society strengthening	77
3.3 Efficiency.....	77
3.4 Project size (budget and duration)	81
3.5 Synergies among the three components.....	83
4 Reflection on evaluation process, lessons for future programme evaluation.....	85

5	Conclusion.....	91
5.1	MDGs and themes.....	91
5.2	Capacity development	91
5.3	Civil society strengthening.....	92
5.4	Synergy, efficiency and project size.....	93
5.5	The evaluation process	94
6	References	96
7	Annex	98
A.	The setup of the evaluation	98
1.	Evaluation framework.....	98
2.	The synthesis.....	99
3.	Sampling.....	101
4.	Timeframe.....	103
5.	Agreed methodologies and choices country teams	104
5.1	<i>Overview of employed methodologies.....</i>	104
5.2	<i>Evaluating efficiency</i>	106
5.3	<i>Attribution of capacity development and civil society strengthening outcomes.....</i>	106
B.	Capacity development & civil society strengthening in Bangladesh	109
C.	Civil society strengthening: outcomes (process tracing and contribution analysis).....	112

List of Tables

Table 2.1 - MDG 1 projects: basic information	18
Table 2.2 - MDG 1 Evaluation scores per project	22
Table 2.3 - MDG 2 Projects: basic information	23
Table 2.4 - MDG 2 Evaluation scores per project	24
Table 2.5 - MDG 3 projects: basic information	27
Table 2.6 - MDG 3 Evaluation scores per project	28
Table 2.7 - MDG 4, 5 & 6 projects: basic information.....	29
Table 2.8 - MDG 4, 5 & 6 Evaluation scores per project	30
Table 2.9 - MDG 7a, b projects: basic information	32
Table 2.10 - MDG 7a, b Evaluation scores per project.....	33
Table 2.11 - MDG 7c projects: basic information	34
Table 2.12 - MDG 7c Evaluation scores per project.....	34
Table 2.13 - Fragile states evaluation scores per project	36

Table 2.14 - Good governance evaluation scores per project	37
Table 2.15 - DRC 5C score changes	40
Table 2.16 - Ethiopia 5C score changes.....	41
Table 2.17 - India 5C score changes.....	44
Table 2.18 - Indonesia 5C score changes	45
Table 2.19 - Liberia 5C score changes.....	47
Table 2.20 - Pakistan 5C score changes	48
Table 2.21 - Uganda 5C Changes.....	49
Table 2.22 - Average change in CSI score per country.....	51
Table 2.23 - CSI Score Changes in DRC.....	52
Table 2.24 - CSI Changes in Ethiopia	53
Table 2.25 - CSI Score Changes in India	54
Table 2.26 - CSI Changes in Indonesia.....	56
Table 2.27 - CSI Changes in Pakistan.....	57
Table 2.28 - CSI Changes in Uganda.....	57
Table 3.1 - Average evaluation scores per theme	73
Table 3.2- Overview MDG efficiency scores	77
Table 3.3 - Average scores based on project duration	82
Table 3.4 - Average scores based on budget size	82
Table A.1 Overview of evaluation components and subdivisions	98
Table A.2 Research questions	99
Table A.3 - Sample size by country and component at baseline	101
Table A.4 - Sample size by country and component at endline.....	103
Table A.5 - Overview MDG methods per country.....	104
Table A.6 - Overview of CD methods per country	105
Table A.7 - Overview CS methods per country	105
Table B.1 - 5C Score Changes in Bangladesh	110
Table B.2 - CSI score changes in Bangladesh	111
Table C.1 - Civil Society Strengthening: Outcomes and Attribution (Process Tracing).....	112
Table C.2 - Civil society strengthening: outcomes and attribution (contribution analysis).....	119

List of Figures

Figure 1.1 Organogram MFS II evaluations.....	14
Figure 4.1 - MFS II evaluation & contracting periods (Indonesia)	86
Figure A.1 Example theory of change	108

Executive Summary

This evaluation of the MFS II programme of the government of the Netherlands was unprecedented in its ambition and scope. It aimed at using the most rigorous evaluation methods and covered a very large sample of MFS II-supported activities: almost 200 projects in eight developing countries. Changes over time were measured for each project: baseline information was collected in 2012 and endline data in 2014. The central questions in this evaluation were whether the changes observed over this two-year period were positive, how large they were, and to what extent they could be credibly attributed to MFS II. The analysis covered three areas: contributions to the Millennium Development Goals (MDGs), to capacity development (CD), and to civil society strengthening (CS). Some of the organisations evaluated were involved in only one of these activities, but in many cases they carried out a combination of activities in all three areas.

Methodology

Modern evaluation methods are characterised by their efforts to construct a counterfactual: what would have happened in the absence of the intervention? It is often argued that randomised controlled experiments (RCTs) are the gold standard for such an approach. The synthesis team does not take that position. RCTs, by their nature, are restricted to evaluating clearly defined (often one-dimensional) activities, which are implemented in a single environment. By contrast, the MFS II programme covered projects with multiple, and sometimes hard to quantify, objectives. In addition, external validity of RCT results is a major problem: there is no guarantee that interventions that work in one environment will also be effective elsewhere.

The synthesis team together with the country teams addressed the issue of appropriate evaluation techniques in a number of workshops prior to the roll out of the evaluation. This has led to a common approach for each of the three types of activities (MDGs, CD and CS).

For the MDG component of the evaluations the teams had to accept that, as a rule, Southern Partner Organisations (SPOs) and their beneficiaries had not been chosen randomly. Given that restriction, they used the best available methods (typically double differencing and propensity score matching methods) to construct an appropriate and convincing counterfactual. For this component the key issue is not rigour (except in the case of Bangladesh) but statistical power: sometimes no hard conclusions could be drawn due to small sample size.

For the CD and CS components of the evaluation, it was clear from the outset that there was little scope for statistical analysis. Therefore, the danger of attributing observed changes largely or wholly to the organisation's efforts (and the MFS II support it received), even if alternative explanations are more plausible, could not be ignored. Some country teams avoided this by very carefully setting out alternative theories of change and comparing their implications systematically with the outcomes observed over the evaluation period. In these cases, the answer to the difficult attribution question was convincing even though there was no scope for formal hypothesis-testing. These teams showed they were able to exceed initial expectations, and thus set the standard for process tracing in future evaluations.

Process tracing was used for only some aspects of the evaluation, and much better by some teams than by others.

The baseline synthesis report stressed that scores cannot be compared since some teams may give higher scores than others in similar situations. This limits the scope for comparisons of scores (levels) across countries or projects. However, the current report focuses on *changes* in scores and, provided the teams have not changed, these are less likely to be subject to evaluator bias.

Findings

Millennium Development Goals

Many of the MDG findings are quite positive: the 53 projects scored quite well in terms of project design, implementation, relevance and efficiency. Clearly, Dutch NGOs and their SPOs know what they are doing and, by and large, they are doing it well. Where low scores were reported there are some obvious explanations. The low-scoring projects are concentrated in two countries. The first is the DRC which still suffers from civil strife and violence. The second is Ethiopia, where pockets of violence exist, including terrorist attacks on Kenya and subsequent retributions by the Kenyan and Ethiopian armies. Clearly, it is difficult to get things done in countries that suffer from civil strife and violence. In fact, it is quite remarkable that some of the projects in those two countries achieved positive results at all. The same is true for Liberia, which was hit by an Ebola epidemic during the evaluation period. Leaving out those three countries, the mean for the important score for objectives achieved was 6.4, a very satisfactory result. In view of the methods used, the synthesis team finds the results produced by the country teams highly credible.

For projects where both CD and MDG objectives were evaluated, there is little evidence that projects that aim to improve the capacities of SPOs succeed in making those SPOs more successful in addressing the MDGs that they are supposed to focus on. Furthermore, MDG projects that claim to also address good governance or fragile states, have little to show for it in these respects. In the former case, this may just be a case of overreach: capacity development takes time, and to translate increased capacities in better MDG outcomes takes even more time. Obviously, it may be hard to detect results during a two-year evaluation period. Still, the lack of results does raise the question why focus on capacity building in the first place.

The lack of impact on good governance or fragile states by MDG-focused projects is less surprising. The relatively small-scale activities of many of the projects in, for instance, the DRC or Ethiopia are no match for the magnitude of the governance and security problems that plague those states. It is probably better to simply not pretend that these, most often very relevant and successful projects, also contribute to these larger goals.

Capacity development

The findings for CD activities are rather mixed. The results in India are generally positive, and in some cases substantial. Furthermore, there is clear indication that these positive results are directly linked to MFS II-supported efforts. The results in the DRC are limited, or even negative. In the absence of a formal

counterfactual this finding may be too negative: it is possible that outcomes would have deteriorated even further without MFS II-supported activities. In Ethiopia scores were high to begin with and even showed substantial improvements over the two-year evaluation period. New leadership of some of the projects appeared to have played a major role. The results for Liberia are modest at best. For Pakistan, considerable improvements are recorded, but no clear explanation for these results is given in the country report.

These findings may reflect the methodology, which leaves considerable room for subjective scores, especially if the scores basically come from SPO staff members themselves. Fortunately, as noted above, the detailed process tracing which was used for some aspects of the evaluation was in some cases done extremely well, leading to convincing answers to the attribution question.

Civil society

Overall, the findings are quite positive but there are some significant differences between countries. For the DR Congo there are both positive and negative findings, with little overall effect. One particularly negative finding regards a micro-credit project that lacked the financial capabilities to be successful, received insufficient support from the Dutch NGO and thus, not surprisingly, was unable to improve the societal position of the victims of armed conflict in the project area, its civil society objective. This is a clear example of overambitious objectives and lack of focus resulting in a failed project.

In Ethiopia, where a 2009 law makes the functioning of civil society organisations quite difficult, and India, the country team finds a general improvement in the dimensions of Civic Engagement and Perception of Impact. Uganda also shows very favourable results across the board, as does Indonesia.

The Pakistan team finds only small changes in the scores for the SPOs and, with one exception, the report does not provide convincing evidence that the changes observed can be attributed to MFS interventions.

Efficiency

The evaluation fails to shed much light on the efficiency question. There are two significant problems. MFS II-sponsored projects are very diverse, even if only the group of MDG projects is considered. Benchmarks for all these projects are hard to come by, although this problem should become less severe over time. A more serious problem, and one that needs to be addressed urgently, is the lack of sufficiently detailed financial information, not only on the side of the SPOs but also, perhaps surprisingly, on the side of Dutch NGOs. Clearly, the efficiency question is highly relevant, and there appears to be ample room for improvements. The fact that neither the recipient organisations nor their Dutch counterparts routinely collect financial data that allows for efficiency analysis, and that neither use benchmarks against which to measure their own performance is disturbing.

Lessons for future evaluations

The report draws five lessons for future evaluations of this kind.

1. Evaluation periods should be project-specific

It turned out that many of the projects in the sample had started well before the 2012 baseline survey, in some cases years earlier. In addition, some of the projects had already ended before the 2014 survey or no longer received MFS II funding at that time. Clearly, under these conditions the 2012-14 changes may reflect the impact of earlier MFS-supported activities (MFS I) or indeed of activities unrelated to MFS. In principle such problems can be avoided through stricter rules in sample selection. However, there is a more fundamental problem which is discussed at length in our baseline report: the projects differ enormously in terms of the likely delay between the intervention and its ultimate impact. For a simple training programme the impact can be almost instantaneous, but it may take years for projects aiming at changes in individual or social behaviour. The report therefore recommends that the rule of a common evaluation period for a very large sample of heterogeneous projects, in this case almost 200, be scrapped. This eliminates the need for the evaluations of the projects in the sample to run simultaneously. Instead, evaluation should be built into the project or program design, start with a proper baseline and continue through the project's life (possibly beyond if sustainability is an issue). Evaluations will then differ across sample projects, both in starting time and in the length of the evaluation period. While this may well be inconvenient in administrative terms, it will greatly enhance the scope for learning from the evaluation results.

It is likely that in many of the cases where the country teams did not find a statistically significant impact, the sample sizes, which were in fact determined by the available budget, were too small. In future evaluations, it would be better to start with the objectives of the project, assess how long it will take to reach them and then estimate the required sample size. Budget considerations should not determine the sample size for data collection for an individual project but rather how many projects can be evaluated in total.

2. Simplify the governance and ensure the independence of the evaluation

The organisation of this project was unusually and unnecessarily complicated. In future evaluations, the structure should be streamlined, have fewer layers of management, advice and supervision, and clearer lines of responsibility.

Only independent evaluations can be credible. Independence obviously requires avoiding any actions, in particular of the organisations that are to be evaluated, that attempt to influence the outcome of an evaluation; but they should do much more: they should also avoid even any suspicion of such actions. This crucial requirement should be reflected in the governance structure of the evaluations.

3. Ensure that evaluations use the best possible methods

For MDG projects the methodology that should be used is generally clear. The key issue for future evaluations is whether the samples used are large enough. This is primarily a budget issue.

For the CD and CS components there are two issues. First, scores may not be comparable across time or across projects. This need not be serious if the evaluation focuses on *changes* in scores rather than on their levels, provided the same teams are used for the baseline and the endline. Second, since there is little scope for formal counterfactuals in CD and CS activities, it is of paramount importance that teams very carefully consider and compare alternative explanations for the observed changes, rather than simply accept the SPO's theory of change and the views of stakeholders on the extent to which results can be attributed to SPO activities and the support provided by MFS II. It is recommended that the very high standard for process tracing set by some teams be imposed in future evaluations across the board, rather than selectively. Clearly, this will make the evaluations more expensive, but also much more useful. For a given evaluation budget it is better to select a smaller sample and apply high quality process tracing for all evaluation questions than to choose a larger sample and leave doubts about the attribution of the results.

4. Do not attempt to use a single evaluation for two different objectives

Evaluations can be used for accountability or as a basis for learning. It is necessary to choose between these two objectives since they have very different implications for sample selection. The main lesson here is Jan Tinbergen's fundamental point that the number of objectives should not exceed the number of instruments: the attempt to achieve different objectives with the same sample is bound to fail.

A sample stratified only in term of the size of MFS II support would be entirely appropriate if accountability is the key consideration. IOB would then report on a random sample of recently completed aid-supported activities each year. If, however, the key objective is learning, then the sample design should reflect this. Comparability would be enhanced by limiting the sample to a small number of particular types of projects. Comparisons would be made within fairly homogeneous groups. This would be very much in line with what IOB already practices in other areas, notably in its meta-evaluation of drinking water and sanitation projects. In the present evaluation, the heterogeneity within the sample is daunting. This undermines the learning function.

5. Pay more attention to efficiency

Assessing the efficiency of the SPOs turned out to be a serious challenge. One reason is that there are still few relevant benchmarks. This problem will become less acute as new benchmark estimates become available. In addition, many SPOs clearly did not recognise the importance of the issue: they had made no effort to collect data that could help them assess whether what they had achieved could be done at lower cost.

1 Introduction

This synthesis report is part of a large-scale impact assessment of the Dutch Co-Financing System, the *Medefinancieringsstelsel* (MFS). The different studies evaluated the impact of MFS II, the grant framework for 2011-2015. The aim of MFS II is to contribute to building and strengthening civil society in the global South, as a foundation for structural poverty reduction.¹ The Netherlands has a long history of distributing development aid through the channel of Dutch civil society organisations. MFS II was preceded by MFS I and a number of other grant frameworks.

The evaluation took place in the context of a debate on the effectiveness of development aid and amid large cuts in the Dutch budget for development cooperation. Simultaneously, discussion was and is taking place on the role of civil society in development cooperation, and on which stance the Dutch government should take vis-à-vis civil society organisations. In response, the Ministry of Foreign Affairs has developed a new grant framework to support civil society in developing countries, focusing on lobbying and advocacy in a broad sense.

1.1 Structure of the evaluation

The joint MFS II evaluations consist of ten interlinked studies. Eight of these analyse the impact of MFS II at the country level. The sampled countries are Bangladesh, the Democratic Republic of Congo (DRC), Ethiopia, India, Indonesia, Liberia, Pakistan and Uganda. For each of these countries, the evaluation took place in two stages: a baseline assessment in 2012 and a follow-up study two years later. The studies focused on three broad outcome areas: attainment of Millennium Development Goals and themes, capacity development (CD) of Southern Partner Organisations, and civil society strengthening (CS).² The assessments resulted in almost 200 technical papers evaluating individual projects and organisations, and eight country-level narratives. The study before you, the ninth, is a synthesis of these reports. It describes the work carried out by the country teams and gives a judgement on the methodologies they employed. Taking these methodologies into account, we summarise the evaluation results wherever possible. More importantly, we synthesise the results across a number of dimensions, analysing what the evaluations teach us about the impact, relevance and efficiency of these development projects. Beside the country and synthesis studies, a tenth study evaluated the International Lobbying and Advocacy (ILA) efforts supported by MFS II.

The studies outlined above form the evaluation in a strict sense. They were carried out by external and independent evaluation teams led by, and largely consisting of, academic researchers. Nevertheless, a complex structure was put in place to tender and oversee the evaluations. This merits further explanation to provide insight into the context in which the evaluation took place. From the perspective

¹ See page 4 of the policy framework for the MFS II programme, *Subsidiebeleidskader Medefinancieringsstelsel II 2011-2015* (Ministerie van Buitenlandse Zaken, 2009).

² Abbreviated as the MDG, CD and CS components of the evaluation respectively. See Annex A for a more elaborate description of the evaluation framework and research questions.

of the evaluation teams, NWO-WOTRO was responsible for adjudicating the tender for this evaluation. The teams are accountable to NWO-WOTRO and the Steering Committee that NWO-WOTRO established to oversee the evaluations. Two Advisory Committees assist the Steering Committee, one with regard to the country and synthesis evaluations, and another working on the ILA evaluation.

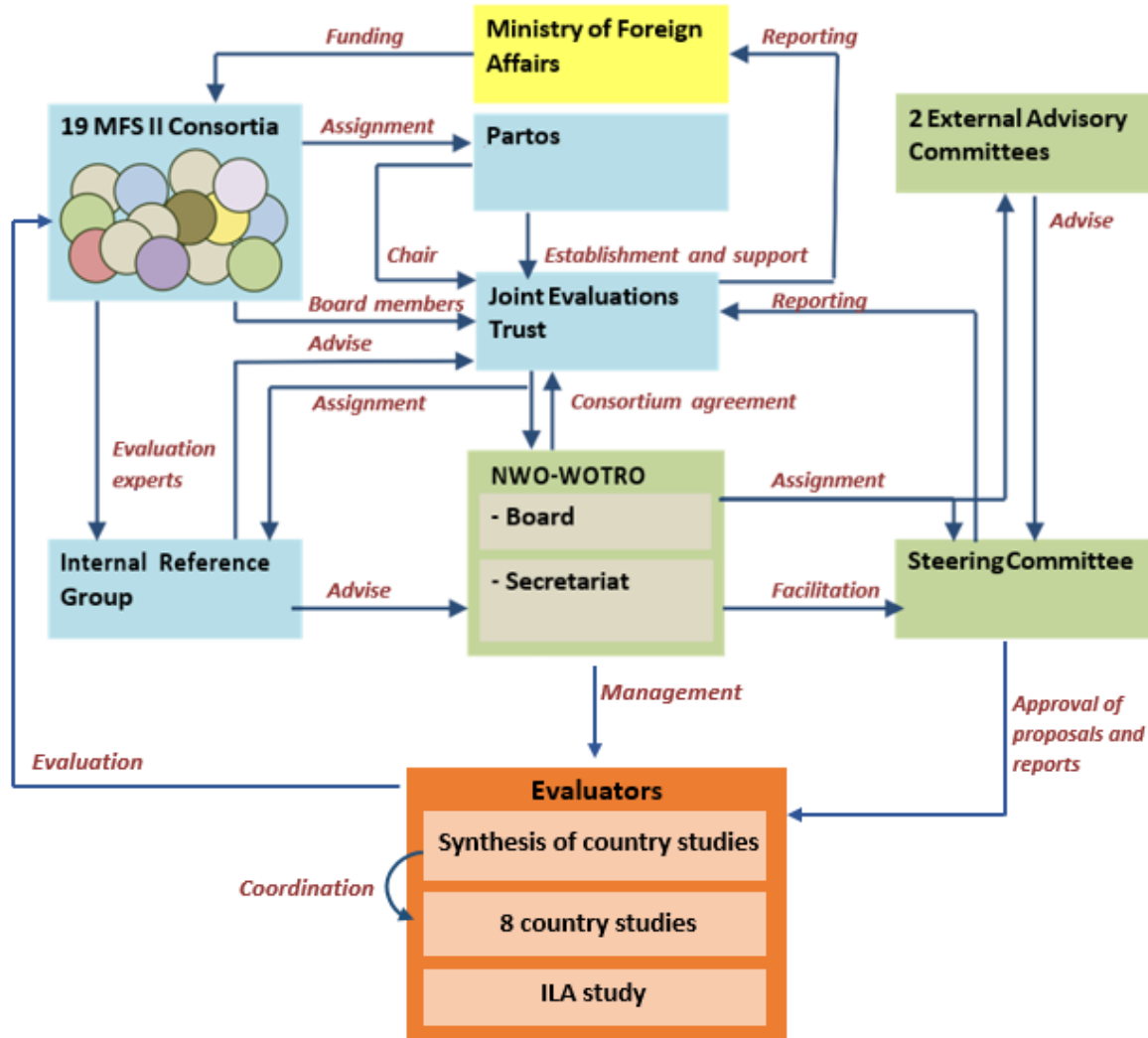
NWO-WOTRO issued calls for proposals on behalf of nineteen out of the twenty MFS consortiums that were granted MFS II subsidies by the Dutch Ministry of Foreign Affairs. In total, the MFS II programme for 2011-2015 covers € 1.9 billion, a substantial amount of the budget of the organisations that form these consortiums. The nineteen MFS consortiums covered by these evaluations established a foundation, *Stichting Gezamenlijke Evaluaties* (SGE), which oversees the evaluation and coordinates with NWO-WOTRO and the Ministry of Foreign Affairs on their behalf. In carrying out these tasks, the SGE is assisted by Partos, the association for Dutch development NGOs, and the *Interne Referentiegroep* (IRG). The IRG groups monitoring and evaluation specialists from the larger NGOs; it was involved in designing the joint evaluation.

The Ministry of Foreign Affairs put forward several conditions to grant MFS II funds to the consortiums. These include requirements concerning monitoring, reporting and evaluation. The nature of these requirements was such that meeting them individually was difficult; therefore, the consortiums decided to set up a joint external evaluation. This joint evaluation, carried out through NWO-WOTRO, which issued the call for proposals discussed below, is intended to satisfy the evaluation requirements of the Ministry of Foreign Affairs.

Figure 1.1 presents a simplified graphic overview of the structure of the MFS II joint evaluations. One of the simplifications is that Figure 1.1 does not show that referees from the Advisory Committee commented on the proposals and the reports. The status of their recommendations to the evaluation teams was not entirely clear.

This complicated structure has at times led to confusion about the responsibilities of the various parties. In chapter 4 we will review this issue, leading to a set of recommendations on how such complex evaluations can be better organised. The present arrangements are unnecessarily complicated and costly. At this stage two points are worth stressing. First, the call for proposals did not envisage a coordinating role for the synthesis team. It soon became clear that, in addition to the roles of the Steering Committee and NWO-WOTRO, there was a need for coordination and that both institutions expected the synthesis team to ensure that the same indicators would be used in the various studies. However, once the country proposals had been approved there was hardly any formal basis for such coordination. Informally, the synthesis team has played such a role supported by the Steering Committee, but this somewhat ambiguous arrangement was not ideal. Secondly, once the proposals were approved, the general methodological framework of the evaluation was fixed. In the baseline report we raised some serious concerns about the extent to which the evaluations could be used for an assessment of the effectiveness of MFS II. It is worth stressing that those concerns were largely related to the approved general setup of the evaluation. Nevertheless, the country teams rose to the challenge of improving their originally accepted methodologies. These developments are discussed in Annex A.

Figure 1.1 Organogram MFS II evaluations³



1.2 Structure of the report

Chapter 2 provides an overview of the results of the evaluation: the changes found, their relevance, and the extent to which they can be attributed to MFS II. Chapter 3 discusses what these results teach us. In chapter 4 we reflect on the evaluation process and draw lessons for future programme evaluations. Finally, our conclusions are presented in chapter 5. Annex A outlines the setup of the evaluation, and discusses sampling, timeframe and methodologies. Note that references to the country reports refer to the final versions, unless otherwise stated.

³ Adapted from the figure made by Mirjam Locadia, Partos, as shown in Van der Meer and Kort (2014).

Please note that the synthesis team has serious doubts about the Bangladesh results, notably on the MDG component. There are errors in the indicated statistical significance of results. Also, the magnitude of many of the reported effects is difficult to believe and the country team has not provided a satisfactory explanation for these results. As a consequence, the MDG results have not been included in the synthesis study. The Civil Society strengthening and the Capacity Development components of the Bangladesh report are discussed in Annex B.

2 Results per priority area

2.1 MDGs and themes

This section discusses project results at the beneficiary level. Outcomes for individuals, households or communities were evaluated depending on the type of intervention. The section is subdivided into six MDG areas and two MDG themes, which are discussed in turn. The evaluation scores given by the country teams are an important element of this summary. The teams were asked to summarise their findings by providing scores from 1 (lowest) to 10 (highest) for the following characteristics of the projects: project design, implementation according to design, achievement of objectives, attribution of success to the project, relevance, and cost-effectiveness. The scores are summarised in the tables below.

In most cases the teams used double-differencing or propensity score matching techniques to construct a counterfactual. This is appropriate in the context of the MFS II evaluation; since decisions on where projects would be active and who would be the beneficiaries were made earlier, there was no scope for randomisation (which for this type of evaluation may in any case not provide the gold standard).⁴ Nevertheless, in some cases the results cannot be confidently attributed to the project or organisation and the MFS support it received. In the tables this is indicated by a low score for “attribution”. This reflects low statistical power: the sample is too small to exclude the possibility that the measured impact in fact only reflects noise.

⁴ The use of randomised controlled trials (RCTs) for the evaluation of development interventions has led to sharp controversies. Extreme advocates of the approach (who came to be called the *randomistas*) argued (1) that RCTs are the gold standard in evaluations and that no other method allows causal inference, (2) that policy questions that are not suitable for RCTs are *ipso facto* not worthy of consideration and (3) that if a policy passes the RCT test successfully anywhere it can, and should be, implemented everywhere. All three claims have come under attack (notably Rodrik, 2008; Deaton, 2010; Ravallion, 2012), but the debate mainly focuses on the last claim, that of external validity.

External validity is usually understood as the validity of a result found for a particular group of people or geographical area for another group or area: if an RCT shows that conditional cash transfers are an effective way of getting poor households in Mexico to send their children to school, can this be assumed to be true as well in, say, Nepal? It would seem obvious that external validity of an RCT result in this sense cannot be assumed.

However, there is reason to doubt whether the RCT methodology provides a gold standard even for the same group and area. The intervention may in practice be applied in a way which cannot be reproduced in the RCT so that an RCT result is misleading. This case arises if local staff have discretion in implementing policies, such as selecting who will participate in the intervention and in exercising that discretion use their own knowledge of how the intervention’s effectiveness is likely to differ across individuals or locations. This case is particularly relevant for development NGOs. Rather than using RCTs the evaluation should be based on observational data since these incorporate information about actual implementation (Elbers and Gunning 2014, 2014a; Gunning, 2014).

Therefore, RCTs cannot be said to provide a gold standard. In addition, the MFS II evaluation left virtually no scope for RCTs: where projects would operate and who might benefit from them had already been decided (non-randomly) prior to the evaluation.

2.1.1 MDG 1 Eradicate extreme poverty and hunger

The country teams evaluated the performance of 20 projects that address the objectives of MDG 1. These projects tried to improve peoples' livelihoods (income) or food security in a variety of ways, including by improving the business climate, the working of, and access to local, regional and international markets, agricultural practices, and access to financial markets for low income households often active in agriculture or small businesses. These projects may also address some of the other MDGs or themes. For instance, the Salvation Army project in the Democratic Republic of Congo builds schools (MDG 2), but also tries to improve the income of the targeted beneficiaries to make those schools more affordable. Finally, some of these projects also have components to build SPO capacity (see Sections 2.2 and 3.2.2), or contribute to strengthen civil society (see Sections 2.3 and 3.2.3)

As for almost all MDGs and themes, the sample of MDG 1 projects is very heterogeneous. Table 2.1 shows that the number of potential beneficiaries ranges from 125 to 26,000 and annual budgets range from just over € 100,000 to € 2.1 million. Per beneficiary, costs range from € 6 to € 863.

Table 2.1 - MDG 1 projects: basic information

Country	SPO	Project	ID	# of beneficiaries	Total budget	Budget/ Beneficiaries	% MFS2	Start Date	End Date	Dur. (months)
DRC	Salvation Army	Construction and support of primary schools	B2	.	€ 647,927	.	86%	2011	2014	36
DRC	SOFIBEF	Programme d'appui...	B3	.	€ 122,000	.	100%	Jan-10	Dec-12	35
DRC	VECO	Development of value chains	B4	.	€ 384,693 ²	.	100%	Jan-11	Dec-13	35
DRC	CEPROF	DCR Pamoja	B5	6,127	€ 260,000	€ 42	100%	Aug-11	Feb-13	18
ETH	OSRA	Zero Grazing Project	C10	.	€ 104,265 ²	.	100%	Oct-10	Sep-14	47
ETH	Hundee	CAVC / C4C / OGRI	C5	8,440 ¹	€ 525,049	€ 62	100%	Sep-11	Dec-13	27
ETH	Facilitators for Change	FMO Consortium	C7	26,688 ¹	€ 1,297,188	€ 49	100%	Jun-11	May-14	35
IND	RVGN	Graduating NGOs into ...	D1	21,161 ¹	€ 117,848	€ 6	31.5%	Jul-09	Jun-12	35
IND	Samarthak Samiti	Mobilisation of Community	D2	3856-3492	€ 441,108	.	14.26%	Apr-09	Mar-14	59
IND	FFID	COFA Institution Building	D3	15,279	€ 1,670,106	€ 109	48.75%	Oct-08	Sep-14	71
IND	Pradan	VBN Poultry Coalition	D5	1,030	€ 230,629	€ 224	48.33%	Dec-09	Mar-12	27
IND	Jana Vikas	India People's Participation ...	D6	1000 ³	€ 164,065	€ 164	20%	May-10	Apr-13	35
IDN	SwissContact SE Asia	Local Economic Development ...	E10	3,382 ¹	€ 897,794	€ 265	34%	Jan-10	Dec-14	59
IDN	FIELD	Local Economic Development ...	E4	161	€ 111,981 ²	€ 696	100%	Feb-11	Feb-14	36
IDN	SwissContact	Implementation of GREEN	E5	125	€ 107,929	€ 863	43%	Jul-11	Oct-13	27
IDN	KSP	Seed capital programme	E6	12,362 ³	€ 201,370	€ 16	95%	Dec-09	Dec-12	36
LBR	DEN-L, LSGCE and FOHRD	PAMOJA	F1, F2, F3	5,000 households	€ 2,111,631	.	100%	2011	2015	48
LBR	BSC Monrovia	Business Start-up Centre	F4	1,489 ¹	€ 351,053	€ 236	100%	2011	2016	60
PAK	Lok Sanjh	Food Security in the Changing Climate	G1	871	€ 360,000	€ 413	100%	Jun-11	Jan-15	43
UGA	Kampabits	Establishment of E-learning Centres	H1	140	€ 88,536	€ 632.40	100%	2010	2014	48
UGA	St Elizabeth Girls Home	Rehab of street & orphan girls	H2	200 per year	€ 326,365	€ 543.94	72%	2010	2013	36
TOTAL				.	€ 531,928	€ 242	75%	.	.	40.5

¹Aggregate beneficiaries. ²Incomplete budget. ³Average beneficiaries per year

In most cases, MFSII is the only donor, or by far the largest. In a few cases MFSII only contributes a small fraction of the budget which, as can be seen throughout this report, makes it difficult to attribute any impact to MFS funding exclusively.

For almost all projects it was possible to conduct a baseline (in 2012) and an endline survey (in 2014). But this, too, points to a common thread in this report. Many baseline surveys were conducted years after the projects started, and some projects were already closed at the time of the endline survey. These issues are discussed more extensively in chapter 0. For now it suffices to point out that some of the MDG 1 projects started in 2008 or 2009, well before the baseline survey, and some ended in 2012 or 2013, before the endline survey.

Four projects in the DRC address MDG 1. The Salvation Army project deals primarily with education, but parents of potential students receive training to increase their agricultural output and thus their income. This income component is necessary because the schools built as part of the project are private, and charge fees. Staff training is part of the project (also for existing schools) and children receive a meal in school to reduce the prevalence of hunger, which is a major problem in this region located close to Bukavu. While the project is deemed highly relevant in this post-conflict region, the results are generally poor. Schools have been built, but enrolment remains low, perhaps because the improved agricultural practices have not yet resulted in a sufficient increase in income that allows for parents or students to pay the school fees.

SOFIBEF is a women empowerment project that aims at improving the social and economic situation of women participating in Women Solidarity Groups. Although these women reported having benefitted from the project during focus group discussions, the quantitative analysis did not find any impact. Armed groups are still active in the region, which could help explain this negative result.

The remaining two projects focus on agriculture and try to rehabilitate the local agricultural economy. Focus group discussion show positive impact of the project with more access to markets through the VICO project, and increased income thanks to CEPROF. However, the quantitative analysis does not confirm these impressions. The continuing acts of violence in both regions may influence these findings, but the country team also mentions the small budgets and fragmented activities of the projects as possible explanations for the disappointing results.

The three projects in Ethiopia work with a consortium of local NGOs and, despite their focus on different activities, the projects are quite similar. There is a major capacity building component with the ultimate goal of helping small farmers to improve their livelihood in various ways. Unfortunately, the impact on this MDG 1 outcome is small. Food security improved in some cases, in others cow ownership grew, but no direct link to income improvements can be found in either project. This could be due to spillover effects on small farmers who do not participate in the project. The country team also observed that these projects operate in an environment where government-supported multipurpose cooperatives are also active. That does raise the questions of how project activities differ from those of government-supported cooperatives, and whether additional agricultural extension services are indeed needed in the area.

Five MFS II-funded projects with a focus on MDG 1 were evaluated in India. They cover a large range of topics, from micro-financing, to forest produce collectors and small farmer cotton cultivation, and from women poultry farming to savings and loan cooperatives. Most of the projects have a large training and skill development component. Despite the heterogeneity, the projects all appear to be quite successful, with significant impact and high levels of attribution.

In Indonesia, four projects focused on MDG 1. The FIELD project aims at increasing the means of livelihood of small farmers, by focusing on a sustainable living environment, forest and biodiversity. It has a savings and credit component that seeks to improve access to financial services and tries to improve market access. This is done with a view to increase the prices for produce sold by the farmers. The project started in 2003, long before the baseline was set up. The country team found clear evidence that the project had reached its objectives, although the cost per beneficiary was relatively high.

The GREEN project provides training for female workers in the garment sector, and tried to improve the access of small enterprises to the general market, with the aim to increase earnings. Improved financial services were also part of the project. Using a combination of qualitative and quantitative methods, the country team found that skills were improved as a result of the training, as was access to financial services. But the impact on linkages to the general market was limited and no overall effect on earnings was found.

The KSP project gives seed capital to participants in savings and loans cooperatives in rural communities. Unfortunately the project ended shortly after establishing the baseline. The evaluation team found substantial improvements to access for financial services, but no effect on income.

The LEET-NTT project helps cashew growers in Flores, one of the poorest provinces of Indonesia. In particular, the project tries to further involve farmers in cashew processing, rather than them just selling the crop to traders. This is done through training and quality certification. The project appears to be quite successful: increased selling in (inter)national markets went up by 50 percent, the share of processed cashews increased by 24 percent and the selling price rose 22 percent. As a result, the income of participating farmers increased. However, the income of farmers in the control group also increased, as a result of more crop diversification by that group. The project was also considered a bit expensive.

The BSC project addresses the difficult context in which young entrepreneurs have to operate in Liberia. They lack adequate skills and have little access to credit. The project tries to reduce these bottlenecks but shows little impact as of yet. This may be explained in part by the small number of observations, but another problem is the lack of start-up loans for those who complete the training programme. The PAMOJA project is a rather large anti-poverty project with a large variety of activities. The evaluation finds positive impact on earnings, access to credit and wage labour, but not on other outcomes such as food expenditures or school enrolment rates. The country team argues that the project can be improved if it focuses on core activities.

One project in Pakistan, the Lok Sanjh project, addresses MDG 1. It focuses on food security in a changing climate. The project has been successful in increasing dietary diversity, but the impact on income and yields is not significant.

The Kampabit project in Uganda addresses youth unemployment in Kampala. Each year 40 applicants are chosen to receive basic training in entrepreneurial skills, ICT, life skills, reproductive health, and HIV/AIDS. After a year, the best 20 students are selected for more advanced training. The project has positive impact on employment, which can most likely be attributed to the high quality of the training, especially the ICT component, which is highly appreciated by the participants. Confidence levels, maturity and resourcefulness also increase. The project is seen as very successful, although a bit pricey.

The St. Elizabeth Girls Home project addresses the needs of vulnerable girls, particularly street children and orphans, by providing vocational training, counselling and psycho-social support, and a first approach to income-generating projects. The project has successfully increased the employment level of these disadvantaged girls. There is also a strong positive effect on their psycho-social condition, namely happiness. This effect fades after they leave the project, but it remains higher than before they joined.

Table 2.2 shows the evaluation scores given by the country teams. There are very few projects with low scores across the board. Furthermore, most projects are well designed and implemented as designed. Nevertheless, achieving all goals appears difficult, especially in Congo, Ethiopia and Pakistan.

It is uplifting to see that the country teams concluded that the positive results in the vast majority of cases where the intended objectives were reached can be attributed to the project. In addition, where a cost/efficiency assessment of the projects was possible, most projects scored relatively high.

One more general comment is in order: there is an almost perfect correlation between the country where the projects are located and the country teams who have carried out the analyses and provided the scores. Inevitably, some country teams are, on average, more generous while others are more critical. Fortunately, the country narratives, as well as the underlying technical reports do, for the most part, clarify the basis of the assessments and scores. This will be a thread running through this report and will be reflected upon in the conclusion.

Table 2.2 - MDG 1 Evaluation scores per project

Country	SPO	Project	ID	Design	Implementation	Objectives	Attribution	Relevance	Efficiency
DRC	Salvation Army	Construction and support of primary schools	B2	5	7	5	3	6	.
DRC	SOFIBEF	Programme d'appui...	B3	5	4	5.5	3	6	.
DRC	VECO	Development of value chains	B4	8	7	4	3	5.5	.
DRC	CEPROF	DCR Pamoja	B5	5	7	5.5	3	5	.
ETH	OSRA	Zero Grazing Project	C10	5	7	5	6	8	8
ETH	Hundee	CAVC / C4C / OGRI	C5	8	7	5	7	6	6
ETH	Facilitators for Change	FMO Consortium	C7	8	8	5	7	6	6
IND	RVGN	Graduating NGOs into ...	D1	8	10	7	7	8	9
IND	Samarthak Samiti	Mobilisation of Community	D2	8	10	6	6	8	6
IND	FFID	COFA Institution Building	D3	10	10	7	7	7	10
IND	Pradan	VBN Poultry Coalition	D5	10	10	7	5	7	10
IND	Jana Vikas	India People's Participation ...	D6	8	6	6	6	8	6
IDN	SwissContact SE Asia	Local Economic Development ...	E10	8	9	7	6	6	5
IDN	FIELD	Local Economic Development ...	E4	8	8	7	7	7	6
IDN	SwissContact	Implementation of GREEN	E5	8	6	6	7	8	.
IDN	KSP	Seed capital programme	E6	8	9	8	7	7	.
LBR	DEN-L, LSGCE and FOHRD	PAMOJA	F1, F2, F3	6	9	6	7	8	6
LBR	BSC Monrovia	Business Start-up Centre	F4	8	7	5	5	8	7
PAK	Lok Sanjh	Food Security in the Changing Climate	G1	5	6	4	3	4	5
UGA	Kampabits	Establishment of E-learning Centres	H1	8	7	8	8	10	5
UGA	St Elizabeth Girls Home	Rehab of street & orphan girls	H2	7	7	7	7	7	7
	Number of projects	21	TOTAL	7.3	7.7	6.0	5.7	6.9	6.8

2.1.2 MDG 2 Achieve universal primary education

There are ten education projects related to MDG 2. As for most other groups of MDG projects, the list for MDG 2 (education) is very varied (See Table 2.3). There are large differences in intended beneficiaries and budgets range from € 48,000 to more than € 2 million. All projects began before the baseline was established and most are still ongoing.

Table 2.3 - MDG 2 Projects: basic information

Country	SPO	Project	ID	# of benef.	Total budget	Budget/ Ben.	% MFS	Start Date	End Date	Durat. (month)
COD	Salvation Army	Construction and support of primary schools	B2	.	€ 647,927	.	86%	2011	2014	36
ETH	NVEA	Non formal alternative basic education	C1	903	€ 182,929	€ 203	96.65%	Jan-11	Dec-15	59
ETH	TTCA	Improving the Teaching-Learning Processes...	C4	867	€ 61,518	€ 71	83%	Jun-11	Jun-15	48
ETH	SIL Ethiopia	Multilingual education	C8	24,817	€ 357,636	€ 14	57.7%	Oct-09	Sep-15	71
IND	VTTC	VTTC Edu Gate	D10	2,000	€ 285,900	€ 143	100%	Jun-10	Jun-15	60
IND	LGSS	Dahar	D4	1,280	€ 48,447	€ 38	100%	May-09	Oct-14	65
IND	COUNT	Education Programme	D9	935	€ 280,000	€ 299	100%	1993	2020	324
LBR	DEN-L, LSGCE and FOHRD	PAMOJA	F1, F2, F3	5000 hh	€ 2,111,631	.	100%	2011	2015	48
UGA	War Child Holland Uganda Office	Conn@ct.Now	H3	103,798	€ 3,268,704	€ 31	100%	2011	2015	48
UGA	FOKAPAWA	N/A	H6		€ 218,638		54%	2011	2013	24
Number of projects	10	TOTAL	.		€ 746,333	€ 114	90%	.	.	78

Table 2.4 shows the evaluation scores for each project. Most education projects received very good scores, with some notable exceptions. The previously mentioned Salvation Army project in the DRC was both poorly designed and poorly implemented. The teacher training project in Ethiopia (TTCA) was well designed but did badly on all other areas of evaluation. At the other extreme the multilingual education project of SIL Ethiopia received an outstanding evaluation.

The Pamoja project in Liberia had multiple objectives, many unrelated to MDG 2. The econometric analysis faced fairly serious attrition (22 percent) and also statistical power problems: means would have

to change by 20 percent to be detectable.⁵ Four indicators were used to measure the project's contribution to MDG 2: adult and child literacy and gross and net enrolment. None of these measures presented a statistically significant difference between treatment and control groups.⁶ However, the evaluation also considered changes in the quality of education using school tests for reading comprehension and mathematics. The results for these additional measures are also disappointing: for mathematics there are no significant differences between the two groups, for reading they behave opposite than expected: schools in the control group did significantly *better* than those in the Pamoja treatment group.⁷ Thus, the evaluation has found no convincing evidence of a positive effect of the education component of Pamoja. It is possible that this reflects the problem of low statistical power.

Table 2.4 - MDG 2 Evaluation scores per project

Country	SPO	Project	ID	Design	Implementation	Objectives	Attribution	Relevance	Efficiency
DRC	Salvation Army	Construction and support of primary schools	B2	5	7	5	3	6	.
ETH	NVEA	Non formal alternative basic education	C1	10	7	8	7	10	8
ETH	TTCA	Improving the Teaching-Learning Processes...	C4	6.5	5	3	3	4	5
ETH	SIL Ethiopia	Multilingual education	C8	10	9	7	7	8	10
IND	VTRC	VTRC Edu Gate	D10	8	6	6	6	8	6
IND	LGSS	Dahar	D4	8	6	8	7	10	.
IND	COUNT	Education Programme	D9	8	9	8	6	8	9
LBR	DEN-L, LSGCE and FOHRD	PAMOJA	F1, F2, F3	6	9	6	7	8	
UGA	War Child Holland Uganda Office	Conn@ct.Now	H3	5	5	3	5	5	.
UGA	FOKAPAWA	N/A	H6	5	7	2	5	5	5
Number of projects		10	TOTAL	7.2	7	5.6	5.6	7.2	7.2

In Ethiopia the NVEA project aimed at improved access to and better quality of education. The evaluation analysed changes in enrolment but also in learning outcomes. Attrition was a serious problem and double-differencing was applied to data treated as two repeated cross-sections.⁸ Students in the treatment group did better in school performance tests. There is also an effect on grades but this may reflect differences in the quality of teachers rather than in performance.⁹ The project was successful in

⁵ Liberia Technical Papers, p. 9 of Pamoja paper (p. 37 of PDF).

⁶ Liberia Technical Papers, Table 5, p. 12 of Pamoja paper (p. 40 of PDF).

⁷ Liberia Technical Papers, Table 8, p. 16 of Pamoja paper (p. 44 of PDF).

⁸ Ethiopia, MDGs, C1 Endline report, Section 5.2, p. 22.

⁹ Ethiopia, MDGs, C1 Endline report, p. 23.

terms of improving school quality and the impact is probably understated as a result of spillover to the control group.

In the evaluation of the teachers training project (TTCA) the outcomes of students who studied with a teacher who had been trained were compared to those of students whose teachers had not received training. Surprisingly, having a trained teacher did not improve school performance in terms of test scores although it did improve student motivation. That the evaluation found little impact is probably due to very high turnover: only a single teacher remained over the full period. (Clearly, turnover leads to underestimating impact as the benefits of teacher training will largely accrue to students not in the sample.¹⁰

The SIL project in Ethiopia is aimed at multilingual education. This project started in 1996 already, which presents a problem for evaluation. The central question in the evaluation is whether a child receiving education in his or her mother tongue, rather than in Amharic, has a positive impact on school engagement, schooling outcomes, self-esteem and ethnic identification.¹¹ Note that not all of these outcomes can be classified as MDG 2 contributions. Again, double-differencing treated the data as two repeated cross-sections because of high attrition rates. The results are somewhat counterintuitive. For Shinasha students, being taught in that language *negatively* affects their school performance, and has no effect on ethnic identification. The reason appears to be that members of the Shinasha ethnic group no longer use their language at home; therefore teaching children in their “mother tongue” in fact exposes them to a foreign language. Why this was not known when the project was designed remains unclear. The Shinasha case is clearly special since the project is effective for other groups: for Bertha children the effect on school performance was both positive and significant. The effect is quite strong: performance improved by over 16 percent as a result of these children being taught in Bertha rather than in Amharic.¹²

The Salvation Army project in Liberia centres on providing primary education for vulnerable groups. The key result is that none of the MDG 2 impact indicators (enrolment and cash spent on schooling) improved significantly.¹³ School dropout rates are very high, and the Salvation Army project has not succeeded in reducing them. The report suggests that high school fees represent a binding constraint for households. In this situation, an intervention of this type is doomed to fail, and in that sense, project design is to blame.

Of the three projects in India, VTRC had an unexceptional score (6) for realising its objectives but did quite well otherwise.¹⁴

The LGSS project promoted education for marginalised children, in particular school dropouts. Implementation suffered under teachers’ strikes and local violence but also from the project’s narrow

¹⁰ The evaluators also make this point: Ethiopia, MDGs, C4 Endline report, p. 27.

¹¹ Ethiopia, MDGs, C8 Endline report, p. 20.

¹² Ethiopia, MDGs, C8 Endline report, p. 26. This is an important and convincing demonstration of the effectiveness of the approach. The evaluation (like quite a few other MDG evaluations) is of very high quality and obviously publishable.

¹³ DRC Technical Papers, Table 7, p. 27.

¹⁴ Table 19.

focus on education which proved to be difficult when the target group faced a host of other problems.¹⁵ The report suggests that in these circumstances the design could have been different but also, and possibly more importantly, there should have been greater flexibility to adjust the design in response to circumstances.

The econometric analysis is one of the best in the entire MFS evaluation. It focuses on two outcome variables: school attendance and test performance.¹⁶ The effect of the intervention on school attendance was strong and significant. Somewhat surprisingly, there is no effect on test scores.

The COUNT project addresses the quality of education for tribal communities. It offers children shelter (in so called Agape homes) and encourages their school attendance. The evaluation is based on comparing the outcomes for children who are in Agape homes, and those who are not. A major problem for the evaluation was a very high attrition rate, partly as a result of children in Agape homes being placed elsewhere after a period of time.¹⁷

For MDG 2 the key results are reported in Table 2a of the technical paper.¹⁸ A naïve reading of the results would suggest that the project was a failure: scores are overall *lower* for children in Agape homes. The report points out that this would be a mistake. Disadvantaged children in the Agape homes indeed score significantly lower than their counterparts as a result of selection effect, but this difference vanishes in higher school grades. This is the true measure of success: over time the project eliminates the effect on learning of a poor starting position.¹⁹

The VTRC project in India is also concerned with the education of marginalised children. It works mainly through tuition centres, a type of private education institution. The evaluation finds little impact in terms of parental involvement but a positive impact in terms of learning outcomes. Just as in the case of COUNT, the effect is negative at first glance. In effect the project succeeds in offsetting an initial disadvantage so that the difference vanishes in higher school grades. This is an impressive outcome.

The WarChild project aims at improving the wellbeing of children and youth in the post-conflict setting of North Uganda. The evaluation focuses on the education component of this project. The FOKAPAWA project addresses the needs of women and youth with a special focus on improving livelihoods and peace building. It also includes an education component. Unfortunately, both projects get low scores across the board. This underscores the difficulties of getting solid education results in an area that has recently suffered from civil strife and violence.

¹⁵ India Technical Papers, p. 1536.

¹⁶ This evaluation report sets a high standard in its very careful and detailed discussion of statistical power issues. The report is also exemplary in using and comparing different estimation methods.

¹⁷ The evaluation report carefully stresses such limitations: India Technical Papers, p. 1646.

¹⁸ India Technical Papers, p. 1546.

¹⁹ India Technical Papers, p. 1650.

2.1.3 MDG 3 Promote gender equality and empower women

Of the five projects focusing on MDG 3, two started in 2010 and three in 2011 (see Table 2.5). The first two ended in 2012. All projects have modest budgets and all but one are fully funded by MFS. The cost per beneficiary ranges from € 32 to € 173.

The project in the DRC, which contributes to women support houses, gets low scores on almost all indicators. It is considered relevant because of the dire situation of women in the DRC, but it is poorly designed and implemented, and shows no results. The country team underscores that part of the explanation for this poor performance is exogenous. The project is set in a region that is still reeling from conflict. Moreover, many NGOs are still active in the project area and the control area, so the impact of one single (and relatively small) NGO may be hard to detect.

Table 2.5 - MDG 3 projects: basic information

Country	SPO	Project	ID	# Benefic.	Total budget	Budget/ Ben.	% MFS II	Start Date	End Date	Durat. (month)
DRC	SOFIBEF	Programme d'appui...	B3	.	€ 122,000	.	100%	Jan-10	Dec-12	35
ETH	ECFA	Nazareth Child Help Line: Protection	C2	3,100 ¹	€ 243,644	€ 78.59	71%	Jan-11	Jan-15	48
ETH	FSCE	Girl Power	C3	9,848 ¹	€ 323,000	€ 32.80	100%	Apr-11	Dec-15	56
IDN	YPI	YPI project	E12	247 ²	€ 42,864	€ 173.54	100%	Apr-11	Dec-15	56
IDN	LRC-KJHAM	Strengthening Marginalized Women ...	E7	1,768 ¹	€ 95,000 ³	€ 53.73	100%	Oct-10	Dec-12	26
Number of projects		5	TOTAL	.	€ 165,302	€ 85	94%	.	.	44.2

¹Aggregate beneficiaries. ²Average beneficiaries per year. ³Budget allocation unclear

The two projects in Ethiopia are well designed, although the Nazareth project has implementation problems. The projects are part of a larger, nationwide “Girl Power” programme to support equal rights and opportunities for girls.

Table 2.6 - MDG 3 Evaluation scores per project

Country	SPO	Project	ID	Design	Implementation	Objectives	Attribution	Relevance	Efficiency
DRC	SOFIBEF	Programme d'appui...	B3	5	4	5.5	3	6	.
ETH	ECFA	Nazreth Child Help Line: Protection	C2	7	5	5	7	9	6
ETH	FSCE	Girl Power	C3	8	9	6	5	9	6.5
IDN	YPI	YPI project	E12	8	9	7	8	8	8
IDN	LRC-KJHAM	Strengthening Marginalized Women ...	E7	9	9	9	4	10	.
TOTAL				7.4	7.2	6.5	5.4	8.4	6.8

The Yayasan project in Indonesia focuses on improving knowledge of and attitudes toward reproductive rights among youth still in school. The project started before the baseline survey, and few additional participants were added between the baseline and endline surveys. This hampers the evaluation. Cross-section regressions did help find that the project significantly increased knowledge about HIV/AIDS and contraceptive methods. Attitudes toward HIV-positive individuals and homosexuals also improved. The project raised hopes that additional benefits could be achieved when participating students shared their knowledge with non-participants. However, such spillover effects were not found.

The LRC-KJHAM project addresses gender based violence (GBV), particularly home-based violence, which is the largest subcategory in Indonesia. The project tries to empower women by increasing awareness, and improving advocacy, through counselling and paralegal training. The project uses violence survivors as coaches, a very positive element. Participants found the project beneficial, but the support of family and friends, usually the first point of contact for women suffering from domestic violence, were also considered important. Therefore, it is difficult to answer the question on attribution.

Given the small number of projects focusing on MDG 3, it is difficult to make further generalisations.

2.1.4 MDG 4, 5, 6 Reduce child mortality rates, improve maternal health & combat HIV/AIDS, malaria, and other diseases

Seven projects in the sample drawn for the evaluation deal with health issues in Ethiopia (AMREF), India (BVHA), Indonesia (Rifka Annisa), Pakistan (Awaz), and Uganda. The budget per beneficiary ranges from just € 3 to € 1,347, but the MFS contribution is very small in the latter scenario. All projects started before the baseline was set and four even ended in the year when the baseline was conducted (2012), with obvious consequences for the impact evaluation.

Table 2.7 - MDG 4, 5 & 6 projects: basic information

Country	SPO	Project	ID	# benef.	Total budget	Budget / Benf.	% MFS	Start Date	End Date	Dur. (month)
ETH	AMREF	AE Project / Unite for Body Rights	C11	507,523	€1,515,000	€ 3	100%	Jan-11	Dec-15	59
IND	BVHA	Community Awareness building	D7	15,187	€ 90,971	€ 6	47%	Mar-09	Mar-12	36
IDN	Rifka Annisa	RA project	E11	617	€ 830,839	€1,347	3.03%	Jan-11	Dec-12	23
PAK	Awaz CDS	Parwan	G3	.	€ 57,667	.	100%	2011	2015	48
UGA	Health Child	HC Project 1	H4	.	€ 525,562	.	55.2%	Oct-11	Feb-14	28
UGA	Diocese of Jinja	Integrated Health Care	H9	.	€ 231,582	.	44%	Jul-09	Dec-12	41
UGA	SHU	Reducing delays to maternal & child health care	H8	.	€ 250,000	.	57%	Jul-09	Dec-12	41
Number of projects: 7			TOTAL	.	€ 500,232	€ 452	58%	.	.	39.4

The evaluation scores show a very mixed picture.²⁰ Awaz received high scores (8 or 9) on all counts. BVHA did quite well (score 8 or 10) in several dimensions, but received poor scores (5) for design and reaching its objectives. AMREF received a score of 7 for design and for reaching its objectives, 9 for implementation and relevance, 8 for efficiency and an unimpressive 6 for attribution. The focus of the evaluation findings is on the extent to which the projects reached their objectives and on the confidence that can be placed in the attribution.

The AMREF project aims at strengthening the health system and on improving education on sexuality, raising the quality of sexual and reproductive health services, and reducing gender-based violence. The balancing tests in the evaluation found large differences between the treatment and control groups.²¹ The econometric analysis, using double-differencing, finds negative or insignificant effects for access to health services. However, there are positive treatment effects for both the use of and satisfaction with health services. The effects of education on sexuality are generally positive but not significant.

²⁰Table 2.8. For Rifka Annisa the scores are not complete.

²¹ Ethiopia C11 Endline report, Table 4, p. 23.

Table 2.8 - MDG 4, 5 & 6 Evaluation scores per project

Country	SPO	Project	ID	Design	Implementation	Objectives	Attribution	Relevance	Efficiency
ETH	AMREF Ethiopia	AE Project / Unite for Body Rights	C11	7	9	7	6	9	8
IND	BVHA	Community Awareness building	D7	5	10	5	8	10	10
IDN	Rifka Annisa	RA project	E11	8	8	8	4	10	.
PAK	Awaz CDS	Parwan	G3	8	9	8	8	8	
UGA	Health Child	HC Project 1	H4	9	7	7	8	10	.
UGA	Diocese of Jinja	Integrated Health Care	H9	7	7	9	9	10	.
UGA	SHU	Reducing delays to maternal & child health care	H8	6	5	6	7	6	.
Number of projects		7	TOTAL	7.1	7.9	7.1	7.1	9.0	9.0

This results in quite mixed evaluation results. A likely reason is that the evaluation period did not at all match the project period: the baseline took place long after the start of the project and the endline took place a year before its end. For both reasons the evaluation is likely to understate its impact. Nevertheless, it is encouraging to see that much of the measured effect is in fact positive and significant.

In the BVHA project in India the objective was to improve access to and the quality of services for sexual and reproductive health. Its main objective is to change the strong preference of Indian couples for male children through dissemination of information and campaigns to change behaviour. Change was measured by comparing households in communities that were exposed or not to such campaigns. As previously stated noted, this project was not well designed and did not score well in terms of reaching its objectives. The key problem was that the project spread its efforts too thinly, covering an unrealistically large number of villages.

In Pakistan, propensity score estimates suggest that the Parwan programme by Awaz had substantial and significant effects on school children, but for only on a few of the outcome indicators considered, notably knowledge of HIV/AIDS.²² The report makes much stronger claims but, as it seems, on the basis of unmatched rather than propensity score matched comparisons. Some of the conclusions in the report therefore probably reflect pre-existing differences rather than treatment effects. A worrying finding is that the knowledge of HIV/AIDS of teachers themselves is quite limited.²³

The Rifka Annisa (RA) project in Indonesia sought to achieve women empowerment. Since no control group could be found, the evaluation relies on a before-after comparison. The sample of women in the evaluation is rather small. A number of positive changes were found: women left their husbands or

²² Pakistan Technical Papers, Table 6, p. 190 of PDF.

²³ Pakistan Technical Papers, Table 7, p. 192 of PDF.

divorced them, thus gaining independence and escaping from abuse. However, in view of the methodology used such results cannot confidently be attributed to the programme.²⁴

The focus of the Health Child project is on improving the health of young children and women of reproductive age. It targets, communities along the shores of Lake Victoria.

The Jinja project aims to improve the overall performance of six mid-sized health centres run by the Diocesan Health Office of the Care Third Health project in Uganda. Both projects get high or very high scores across the board. The main objective of the Third Health project (SHU) is to improve mother and child health by providing better access to maternal health. While access indicators show improvements, no major changes on health outcomes are apparent. The country teams notes that the time frame of the evaluation may have been too short to expect such changes.

2.1.5 MDG 7a, b Integrate the principles of sustainable development into country policies and programs; reverse loss of environmental resources and reduce biodiversity loss

The seven projects grouped under MDG 7a and 7b address various aspects of environmental sustainability (Table 2.9). Three of the projects are of modest scale, but the annual budgets for another three range from € 600,000 to € 3.2 million (of which only 15 percent comes from MFS).

The Réseau CREF project in DRC tries to introduce sustainable land use practices through participatory management techniques. However the project has suffered from the lack of support by local politicians. In addition the implementing agency is not trusted by the local villagers because it bought land close to the villages to demonstrate the project. As a result, local villagers now need to travel farther to work on their own land. The project is considered a failure (Table 2.10). A similar problem of mistrust from local government officials has plagued the NTFP-EP project on securing forest livelihoods in Indonesia, but it appears that the implementing agency has been able to overcome it.

²⁴ Indonesia Endline Report on the Achievement of MDGS and Themes, Evaluation of the Rifka Annisa Project, especially p. 43.

Table 2.9 - MDG 7a, b projects: basic information

Country	SPO	Project	ID	# Benef.	Total budget	Budget/ Benef.	% MFS	Start Date	End Date	Durat. (month)
DRC	Réseau CREF / AGIR	Appui au plan de gestion...	B1	.	€ 95,000	.	92%	Dec-11	May-15	41
ETH	CARE Ethiopia	Climate Proof DRR	C6	1,500 HH ²	€ 607,241	.	100%	Jul-11	Jun-16	59
ETH	HOAREC /N	Sustainable energy...	C9	60,000 ¹	€ 3,258,877	€ 54.31	15%	2011	Dec-14	47
IDN	WIIP	Climate-Proof Disaster Risk Reduction	E1	194 HH ²	€ 1,001,000	.	100%	Oct-11	Jun-15	44
IDN	Pt.PPM A	Empowerment of Papua Customary People ...	E2	.	€ 130,328	.	100%	May-12	Apr-14	23
IDN	NTFP-EP	Up-scaling [...] forest livelihoods ...	E3	.	€ 135,217	.	100%	Sep-11	Dec-14	39
UGA	Uganda RCS	CPDRR Uganda	H5	2012	2015	36
Number of projects: 7		TOTAL		.	€ 871,277	.	85%	.	.	41

¹10,000 Households ²HH= households

The two MDG 7 projects in Ethiopia are well designed, but struggle to achieve their objectives (Table 2.10). The HOAREC project, which aims at introducing sustainable energy sources, like efficient stoves and solar home systems, to the rural poor has low scores and has not achieved any of its objectives. It appears that the project was simply too ambitious to be successful in such a poor environment

The WIIP project aims at climate-proof risk reduction in 7 coastal villages in Flores, Eastern Indonesia. It provides grants to communities so they plant and maintain mangrove forests. The evaluation shows clear evidence of increased awareness and mobilisation of the community. One hundred and thirty six hectares of additional forest were realised. However, the impact of the project on the livelihood of the villagers was minimal.

The project Empowerment of Papua Customary People for Sustainable Natural Resource management suffered from delays in implementation. Although the project was well-designed it ended being a failure and was ultimately terminated during the evaluation period.

In contrast, the NTFP project was well executed. The aim was to scale up community-based forest livelihoods using a non-timber forest product exchange program. It has contributed to community awareness regarding the importance of regulating forest use and it has increased more sustainable use of the forest by the villagers. Unfortunately, while the forest mapping exercise was completed successfully, these maps had not yet obtained legal status at the time of evaluation.

The Uganda project addressing MDG 7a and 7b showed some impact on knowledge of measures that can prevent drought-related problems, but the country team does not find any evidence of impact on the prevalence of flood and drought related problems facing crop or livestock production.

Table 2.10 - MDG 7a, b Evaluation scores per project

Country	SPO	Project	ID	Design	Implementation	Objectives	Attribution	Relevance	Efficiency
DRC	Réseau CREF / AGIR	Appui au plan de gestion...	B1	5	2	1	2	5	.
ETH	CARE Ethiopia	Climate Proof DRR	C6	7	9	5	6	8	4
ETH	HOAREC/N	Sustainable energy...	C9	7.5	8	1	8	8	5
IDN	WIIP	Climate-Proof Disaster Risk Reduction	E1	9	9	9	7	9	8
IDN	Pt.PPMA	Empowerment of Papua Customary People ...	E2	8	1	3	1	2	1
IDN	NTEP-EP	Up-scaling [...] forest livelihoods ...	E3	8	7	6	7	8	7
UGA	Uganda RCS	CPDRR Uganda	H5	5	5	3	10	5	.
Number of projects		7	TOTAL	7.1	5.9	4.0	5.9	6.4	5.0

The performance scores for the MDG 7a, b projects are among the lowest for all MDGs. Four projects score very low on objectives achieved, although two of the projects in Indonesia show that well-designed projects can be implemented successfully and efficiently in the area.

2.1.6 MDG 7c Water and sanitation

There are only two projects in this category: HOAREC and AMREF, both in Ethiopia and both fully funded through MFS II (Table 2.11). The number of beneficiaries was quite large: some 20,000 in the case of HOAREC and 38,000 for AMREF. The scores in Table 2.12 show that both projects did quite well in terms of efficiency and extremely well in terms of design, implementation and relevance.

Table 2.11 - MDG 7c projects: basic information

Country	SPO	Project	ID	# Benef.	Budget	€/Benef	% MFS	Start Date	End Date	Dur
ETH	HOAREC /N	Innovative WASH	C12	19,805	€ 251,100	€ 12.68	100%	Aug-12	Sep-15	37
ETH	AMREF Ethiopia	Pastoralist WASH	C13	38,147 ¹	€1,245,000	€ 32.64	100%	Jan-11	Dec-15	59
Number of projects: 2		TOTAL	.	€ 748,050	€ 23	100%	.	.	.	48

¹Target population

HOAREC, however, did very badly (score 1) in terms of reaching its objectives. The econometric analysis shows that there was no impact whatsoever. Most importantly, the project failed to achieve its key objectives: increasing the use of safe water and sanitation services and improved hygiene practices. The intervention aimed to improve hand washing but that practice increased more in the comparison villages and that perverse difference was statistically significant. There was a similar perverse effect for the use of improved latrines.

Table 2.12 - MDG 7c Evaluation scores per project

Country	SPO	Project	ID	Design	Implementation	Objectives	Attribution	Relevance	Efficiency
ETH	HOAREC/N	Innovative WASH	C12	9	8	1	8	8	8
ETH	AMREF Ethiopia	Pastoralist WASH	C13	9	9	6	6	9	7
Number of projects		2	TOTAL	9	8.5	3.5	7	8.5	7.5

The evaluation considered an entire series of outcome indicators, but found no significant differences irrespective of the methods used (various combinations of double-differencing and propensity score matching).

These negative findings suggest total failure. This is especially disconcerting since the analysis is quite convincing. The distribution²⁵ of propensity scores, for example, is virtually identical in treatment and control villages.²⁶ The report suggests that the project still had one year to run at the time of the endline as a possible explanation for it not yet having realised its potential, but this cannot be an adequate explanation since no effect can be seen after two years.

²⁵ While the HOAREC case is striking, the point is more general. For future evaluations there is a need to ensure that capabilities are measured correctly, which HOAREC staff apparently did not do, and that they are relevant to the SPO's objectives.

²⁶ Ethiopia C12 Endline report, Figure 1, p. 10.

Comparing the MDG contribution of the project with the five capability scores suggests a remarkable disconnect.²⁷ HOAREC capabilities were quite good at the time of the baseline intake and somewhat improved in the following two years. Its score for ability to deliver on development objectives rose from 3.2 to 3.5. It is difficult to reconcile this with the evidence that HOAREC failed to achieve any results.

The Amref project also aimed at improved access to safe water and sanitation for rural populations. In this evaluation, treatment and control groups were not well balanced. This makes the double-differencing method, which corrects for initial differences between the two groups, particularly appropriate. The analysis shows that the project was successful in terms of reducing the distance to safe water. This effect is highly significant in any way it is measured.²⁸ The project was less successful in making people switch to a safe source as their main source for drinking water. While an effect could be observed, it was not significant. In the endline survey 40 percent of the respondents reported that their main source for drinking water was unsafe. The problem the project unsuccessfully sought to address is therefore an important one.

In terms of sanitation the results are quite encouraging.²⁹ Access to public latrines improved, as did their use and the practice of handwashing after defecation. All three effects are statistically significant. It is important to note that the baseline took place after the start of the project. To that extent there was a bias towards finding *no* effect: whatever the project's impact, much of it could have been realised prior to the baseline and the evaluation would not have picked up the change.

The results of the Amref evaluation are therefore mixed. Outcomes were disappointing in terms of safe water use, encouraging in terms of sanitation practices (although only in terms of public latrines: there was no change in latrines at the household level). In terms of econometric methodology the situation was far from ideal because of the long delay between the beginning of the project and the establishment of the baseline. This reinforces the findings for sanitation, which were positive in spite of the delay. Nevertheless, in the case of safe water it could have caused a false negative: there could well have been a positive impact that went undetected because it occurred prior to the baseline. This underlines the importance of aligning baselines and endlines with a project's lifetime.³⁰

2.1.7 *Good governance, civil society and fragile states*

Six of the projects that had one or more MDGs as their main focus also aim to contribute to good governance, or to strengthen civil society. Five also address fragile states. Some of the projects overlap in these two categories. Not surprisingly, three of the projects related to fragile states are in the DRC and two are in Liberia. They range from women empowerment projects and improvements in the agricultural value chain, to community development and improvements in the business environment. Unfortunately,

²⁷ Endline Report MFS II Evaluation for Ethiopia, Table 11, p.76.

²⁸ Ethiopia C13 Endline report, Table 7, p. 27. Distance is measured in two ways and for both the wet and the dry season.

²⁹ Ethiopia C13 Endline report, Table 8, p. 28.

³⁰ In chapter 0 we suggest that this will be much easier if the attempt to evaluate projects all at the same time rather than on a continual basis is abandoned.

the evaluation results of these fragile state components by and large fail to show positive results. For instance, it could be reasonably argued that improving the value chain for agricultural products may over time improve livelihoods, and thus contribute to a more stable environment. But the potential impact of such an approach is no match for the magnitude of the problems caused by violence and civil war. The projects may, and often do reach their MDG related goals, as seen in previous sections, but the claim that they also contribute to strengthen fragile states seems overly ambitious.

The five projects listed under fragile states and shown in Table 2.13 were previously discussed in section 2.1.1 since they also address MDG 1. These projects essentially had no MDG impact except for the PAMOJA project which lacked focus and succeeded in some dimensions while failing in others. For the three DRC projects the attribution score is extremely low (3) so very little, if anything, can be concluded. Note that the attribution score is also poor for the BSC project in Liberia.

Table 2.13 - Fragile states evaluation scores per project

Country	SPO	Project	ID	Design	Implementation	Objectives	Attribution	Relevance	Efficiency
DRC	SOFIBEF	Programme d'appui...	B3	5	4	5.5	3	6	.
DRC	VECO	Development of value chains	B4	8	7	4	3	5.5	.
DRC	CEPROF	DCR Pamoja	B5	5	7	5.5	3	5	.
LBR	DEN-L, LSGCE and FOHRD	PAMOJA	F1, F2, F3	6	9	6	7	8	6
LBR	BSC Monrovia	Business Start-up Centre	F4	9	7	5	5	8	7
Number of projects		5	TOTAL	6.4	6.8	5.2	4.2	6.5	6.5

A similar point holds for the six components related to good government or strengthening civil society in the MDG projects listed in table 2.14. These projects focus on improvements in agricultural practices, community rights, and women empowerment, among others. One of these projects is also listed in Table 2.13. The other five include the DRC project Réseau CREF/AGIR. Here the attribution score was extremely low (2). There is the (remote) possibility that such a project, if successful, would contribute to better forms of more participatory government practices. But the impact is likely to be small and, indeed, the score for objectives reached was only 1.

The projects in India, Indonesia and Pakistan showed much better results and these are, with the exception of the Smile Foundation project, credible in view of the high scores for attribution. The mostly advocacy project Pakistan PFF (Pak FisherFolk) was well designed and implemented. However, it cannot be considered very successful in terms of good governance. Here the evaluation used two measures: access to fishing licences, and voter registration. There was no improvement in terms of the former and limited in terms of the latter. In Indonesia the YRBI project aimed at empowering indigenous communities.

The evaluation found it effective and well designed, but costly, and change for this project was negative in terms of CD. The HuMA project also sought to empower indigenous communities focusing on forest management and the containment of violence. One of the project’s activities was training to use local courts to prevent natural resource extraction. The survey found no evidence of success in terms of better use of regulations or an improvement in the villagers’ sense of control over the use of their land. Therefore the score for reaching objectives is modest.

Table 2.14 - Good governance evaluation scores per project

Country	SPO	Project	ID	Design	Implementation	Objectives	Attribution	Relevance	Efficiency
DRC	Réseau CREF / AGIR	Appui au plan de gestion...	B1	5	2	1	2	5	.
DRC	CEPROF	DCR Pamoja	B5	5	7	5.5	3	5	.
IND	Smile Foundation	Action for Children	D8	8.5	8.5	8.5	5	.	8
IDN	YRBI	Empowerment of Mukim and Gampong Capacity ...	E8	8	9	7	8	7	.
IDN	HuMA	Security of Strengthening Communities Rights ...	E9	8	10	7	7	7	.
PAK	PFF	Just and sustainable livelihood	G2	9	9	7	8	8	8
Number of projects		6	TOTAL	7.3	7.6	6.0	5.5	6.4	8.0

2.1.8 Conclusion

In sum, the scores given by the country teams on projects addressing the MDGs and themes are overwhelmingly positive (6 or higher, often much higher), especially for Design, Implementation, Relevance and Efficiency. Clearly, Dutch NGOs and their SPOs know what they are doing and, by and large, they are doing it well.

Projects that do not achieve all objectives have a few things in common. First, it appears to be very difficult to work in the DRC of Ethiopia, where external events like violence and a pervasive lack of security, make it hard even for well-designed and well-implemented projects to live up to their ambitions. Second, a few projects are just badly designed, without taking local circumstances sufficiently into account. Finally, some projects show mixed results and get a low score because their ambitions seemed to be unrealistically high. Fixing the last two problems is easy. It is fully understandable that projects in countries that suffer from civil strife and its aftermath will not always be successful.

It is worth noting that the synthesis team believes the results produced by the country teams are highly credible. In general, the country teams used the best available evaluation methods given the

circumstances, and where “second best” methods had to be used, the results are presented with great care and often after triangulation with information from secondary sources. As a result, the performance scores are, for the most part, highly credible.

The scores of the attribution question are addressed in section 3, where learning points from the evaluations are also discussed.

2.2 Capacity development

The capacity development (CD) evaluations assess changes in the capacity of Southern Partner Organisations (SPOs) of Dutch MFS organisations. More importantly, they attempt to determine whether changes occurred as a result of the interventions of the latter.

The 5C methodology defines capacity in terms of five core capabilities (see Section 1 in Annex A). It puts the five capabilities on an equal footing. There is much to be said, however, for viewing the capability to deliver on development objectives as paramount: after all, this capability is what makes an SPO effective. The other four capabilities are clearly important, but indirectly, as a means for achieving the capacity to deliver, rather than as capabilities which are valuable in their own right. Therefore it becomes particularly important to consider changes in the scores for capability to deliver.

Table 2.15 to Table 2.21 show that these changes were modest: evaluators typically found a small improvement of about 0.3 over the two-year period. In some cases no change was found and in a few, one in Congo and two in Indonesia, the score deteriorated.

All five capabilities were considered to some extent in the CD evaluations, but the process tracing outcome was selected for further analysis. In this case more attention was paid to the theory of change, to possible alternative explanations for the findings, and for the views of actors other than stakeholders. Process tracing was meant as an attempt to make attribution of the results as convincing as possible in a situation where there was no scope for formal hypothesis testing. The country teams differed in how far they took process tracing. We will indicate below that in some cases this work was exemplary and much more convincing than was envisaged in the baseline report.

It should be noted that the evaluations of capacity development usually relied heavily on the perceptions of SPO staff even in the more detailed analysis of process tracing. This could have introduced a positive bias. The answer to the attribution question of to what extent the observed changes are the result of MFS support is also largely based on staff judgement. In both respects the notable exception is the India report, and to a lesser extent the Indonesia report, where in most, but by no means all, cases there is a very serious effort to consider alternative explanations for observed changes in addition to the theory of change, and also to other external sources than opinions of staff and other stakeholders.³¹

DRC

There was no clear evidence in any of the five Congo cases that MFS support causing a noticeable improvement. For AFEM-SK, an organisation concerned with sexual violence, there was no improvement in the CD scores and the evaluation team pointed out that considerable room for improvement remains. In any case, Cordaid was only one of multiple international organisations that offered organisational support to this SPO.³² Even without this dilution, large CD impact cannot be expected since Cordaid changed its approach shortly after the baseline was established. The previous programme of intensive

³¹ In most reports, all sections on key organisational capacity changes begin with the same phrase: “During the endline workshop some key organisational capacity changes were brought up by [the SPO] staff”.

³² DRC Technical Papers, p. 184.

training was stopped and no further training took place after late 2012.³³ The in-depth analysis focused on AFEM’s capability in the field of monitoring and evaluation. This confirmed the general picture: after 2012 no more training in this area was initiated by Cordaid, other organisations were also involved and much of the impact that might have been achieved was lost when the staff member responsible for monitoring and evaluation left.

All five capabilities of KMS, an SPO seeking sustainable development that received support from ICCO, deteriorated over the period (see Table 2.15). The team found that KSM staff were focusing on checking boxes for donors rather than on rectifying major problems in a water project.³⁴

Table 2.15 - DRC 5C score changes

SPO	Baseline					Endline					Changes				
	Act & Commit Objectives	Adapt & Renew	Relate external	Coherence	Act & Commit Objectives	Adapt & Renew	Relate external	Coherence	Act & Commit Objectives	Adapt & Renew	Relate external	Coherence			
AFEM-SK	3.5	4	3	4	3	3.5	4	3.5	4	3	0	0	0.5	0	0
KMS	3	3	3	2.5	2.5	2.5	2.5	2	2	2	-0.5	-0.5	-1	-0.5	-0.5
RHA	3	2.5	2.5	3	2	2	2	2	2	3	-1	-0.5	-0.5	-1	1
VECO	4.5	4	3.5	3	4.5	4.5	4	3.5	3.5	5	0	0	0	0.5	0.5
CEPROF	3	3	2	2.5	2.5	3	3.5	2	2.5	2	0	0.5	0	0	-0.5
TOTAL	3.4	3.3	2.8	3	2.9	3.1	3.2	2.6	2.8	3	-0.3	-0.1	-0.2	-0.2	0.1

The same was true for RHA. There was a general deterioration in capabilities: the capacity to achieve coherence improved, but the scores for the other four capacities all diminished down substantially.³⁵ The network is headed by a person who is overburdened with coordinating tasks and whose leadership is challenged by the staff. The organisational problems at RHA have been recognised for at least three years but remain unresolved. A key problem is that of long delays in donor funding payments. Because of this, salaries go unpaid and activities grind to a halt. Predictably this has undermined the reputation of RHA among local contacts. Monitoring data are collected, but staff do not know what use to give to them. RHA did succeed in improving coherence by attracting funding for a portfolio of similar activities. A key issue is that more coordination is required but that donors want to reduce funding for coordination. However, the evaluation concludes that without the MFS involvement of PAX and Cordaid the slow progress of restructuring the organisation would not even have gotten underway.

³³ DRC Technical Papers, p. 190.

³⁴ DRC Technical Papers, p. 223; cf. p. 222.

³⁵ DRC Technical Papers, p. 238.

A general point about the DRC evaluation is that, unlike most other CD evaluations, this one went well beyond staff opinions. The team clearly went out of its way to collect oral and written evidence from a variety of sources.

Ethiopia

For the SPOs in Ethiopia the situation at the time of the baseline measurement was already good in terms of capabilities. CD scores were high, averages across the nine organisations ranging from 3.3 to 3.7 (Table 2.16). Nevertheless scores improved substantially over the two-year period and there were only 2 cases of regress out of 45 scores.³⁶ In one of these cases the change in score was small (-0.1). The scores suggest that, with one notable exception, capacity development in Ethiopia was successful. The exception is the deterioration (-0.5) in the Capacity to Act and Commit of the Forum on Sustainable Child Empowerment.

For CARE, the evaluation raises the important point that policies and strategies are designed by the international organisation, and CARE-Ethiopia only adapts them to local circumstances. To that extent local capabilities are not very important: outcomes for CARE and similar organisations may well be driven by the capabilities of the international organisation.

Table 2.16 - Ethiopia 5C score changes

SPO	Baseline					Endline					Changes				
	A&C	Obj	A&SR	Rel	Coh	A&C	Obj	A&SR	Rel	Coh	A&C	Obj	A&SR	Rel	Coh
AMREF	3.5	3.7	3.3	3.5	3.2	4.5	4.1	3.8	4.2	3.8	1	0.4	0.5	0.7	0.6
CARE Ethiopia	3.8	4	3.8	4.2	4.1	3.9	4.2	4.1	4.4	4.3	0.1	0.2	0.3	0.2	0.2
ECFA	3.1	3.4	2.9	3.9	3.5	3.7	3.7	3.4	4.3	3.7	0.6	0.3	0.5	0.4	0.2
FSCE	3.6	4	3.8	4	4.1	3.1	4	3.8	3.9	4.1	-0.5	0	0	-0.1	0
HOAREC	3.4	3.2	3.4	3.9	3.2	3.7	3.5	3.9	3.9	3.6	0.3	0.3	0.5	0	0.4
NVEA	3.1	3.5	3.4	4	4.1	3.9	3.7	3.6	4.2	4.6	0.8	0.2	0.2	0.2	0.5
HUNDEE	3.7	3.8	3.9	4	3.9	4.2	4	4.2	4.2	4.2	0.5	0.2	0.3	0.2	0.3
OSRA	3.2	3.6	3.1	3.9	3.9	3.4	3.8	3.7	4	4.2	0.2	0.2	0.6	0.1	0.3
TTCA	2.9	2.75	2.4	1.9	3.3	3.2	2.9	2.6	2.4	3.3	0.3	0.15	0.2	0.5	0
TOTAL	3.4	3.6	3.3	3.7	3.7	3.7	3.8	3.7	3.9	4.0	0.4	0.2	0.3	0.2	0.3

In some cases it was quite clear why capabilities improved. For example the case of AMREF. New leadership—a new country director and deputy director—led to major changes: a survey was used to assess staff capabilities, performance targets were set and extensive staff consultation took place. At ECFA, an SPO for child protection, better leadership, improvements in physical infrastructure and more

³⁶ Five capabilities for nine organisations.

staff helped to improve capabilities. It should be noted that for ECFA, the changes in the Capacity to Act and Commit was quite large (0.6, Table 2.16).

For the Forum on Sustainable Child Empowerment the evaluation stresses staff training and improved reporting as the key changes. At the time of the baseline measurement, reporting was done orally and unsystematically+ this has since changed. Nevertheless, the scores for this organisation show little improvement, indeed a negative change on two of the five capabilities. This is because while the conclusion that new leadership was required had already been reached in 2013, the new director only arrived at the time of the endline. The evaluation argues that, to a large extent, improvements can be attributed to MFS-funded training activities.

In the case of HOAREC the improvement in scores is accounted for by a few simple organisational changes: new premises, better software, and more staff. It is useful to consider HOAREC in some detail. During the endline workshop the staff identified a long series of reasons for the observed changes in capabilities, including more and better staff, clearer responsibilities as a result of the development of an organogram, training on logframes and on monitoring and evaluation.³⁷ The evaluation makes clear that some of the improvements, like the ability to attract better staff as a result of affiliation with the university, have nothing to do with MFS while others, such as the MFS-supported logframe training, do.

The example shows that while it is usually clear that MFS has made a positive contribution, in most cases little can be said about the *extent* to which observed changes can be attributed to MFS. The second noteworthy point is that many of the changes identified by HOAREC staff (better offices, better software, and more and better trained staff) are not at all specific to that organisation. Why such changes have greater impact in one organisation than another remains something of a mystery.

For SPO HUNDEE, the change shown in Table 2.16 in the Capacity to Act and Commit is large and positive. One of the reasons identified for this improvement is that the organisation improved its ability to write successful proposals. Interestingly, this was not due, as would be expected, to training in this area but to “the long experience they had in implementing different rural development projects.”³⁸ Since that experience was built up on over an extended period, the evaluation picks up the effect of efforts prior to the baseline measurement. MFS training in financial management provided by ICCO staff was one of the reasons for the improvements seen at HUNDEE. Similarly ICT capacity improved through various ICCO initiatives, including training on using smart phones for data collection.

For OSRA, donor demands stimulated better monitoring and evaluation. Its capabilities also improved as a result of decentralisation. As in the case of HUNDEE, some improvements in this SPO reflect long term processes, notably the trust it gained from the community through its past experiences with project implementation. As the evaluation report makes clear, many of the activities credited with having improved OSRA’s capabilities were funded through MFS.³⁹

For the last Ethiopian SPO, TTCA no large changes were recorded.

³⁷ Ethiopia Endline report, HOA-REC report, p. 34.

³⁸ Ethiopia Endline report, HUNDEE report, p. 35.

³⁹ Ethiopia Endline report, OSRA report, especially p. 38.

India

For India the CD component of the evaluation initially covered 12 organisations. The two cases where there was no endline are not considered. For the remaining ten organisations, changes in the scores for the five capabilities are almost always positive.⁴⁰ These changes are in general substantial, given the short period over which they were measured. However, for RGVN they are so small (0.1 is the maximum change in Table 2.17) that one might as well say that the scores did not change. The workshop did identify two important changes: improved relations with networks and improved capacity to train partners.

Process tracing focused on the capacities to Act and Commit and to Adapt and Self-renew. For BVHA, an organisation for sexual and reproductive health rights (SRHR), there were major improvements in planning and monitoring related to MFS, largely as a result of biannual meetings. The capacity of the organisation to deliver SRHR services was strengthened through better knowledge and this can be partly attributed to MFS, but was also the result of training initiated by other donors. Improved planning was partly due to frequent Simavi consultations but also had causes without MFS links. The report clearly shows that MFS supported activities played a major role in strengthening the organisation's capabilities. Much of the training which led to improved capabilities was MFS funded.

For COUNT, the evaluation documents that the phasing out of support from Woord en Daad (which started in 2010 and is to be completed in 2020) had positive effects on the organisation, stimulating better planning and a diversification of funding sources. This was not only because the prospect of losing MFS funds led to efforts to retain existing donors and find new ones, but also because the SPO received guidance from Woord en Daad on how to collect and analyse data.

FFID, an organisation that promotes organic farming, experienced only small changes in capabilities. The imminent ending of funding appears to have had a positive effect in this case as well. Jane Vikas (JV) registered some of the largest improvements in capabilities. There are multiple reasons for this but no very clear answer to the attribution question emerges. JV staff considered MFS funded training as important, but this is unfortunately not discussed in any detail in the report. For Shivi Development Society (SDS), MFS activities played no role at all in improving its capabilities. For Smile, the main changes in the evaluation period were improved abilities to raise funds and to organise workshops. This is another case where the end of donor funding from Wilde Ganzen played a stimulating role. Yet fundraising also improved through training activities supported by Wilde Ganzen and the greater ability to organise training reflected a requirement in the contract with Wilde Ganzen as well as extensive feedback on trainings from that organisation. For this SPO the MFS contribution to the observed improvements is fairly clear, although the evaluation points out that some training predates the MFS period.

⁴⁰Table 2.17. The one exception is the deterioration in three scores for Samarthak Samiti.

Table 2.17 - India 5C score changes⁴¹

SPO	Baseline					Endline					Changes				
	Act & Commit Objectives	Adapt & Renew	Relate external	Coherence		Act & Commit Objectives	Adapt & Renew	Relate external	Coherence		Act & Commit Objectives	Adapt & Renew	Relate external	Coherence	
BVHA	3.5	3	3	3.75	4	4	3.5	3.5	4	4.1	0.5	0.5	0.5	0.25	0.1
COUNT	3	3.5	3	3	3.5	3.5	3.5	3.5	3.25	3.75	0.5	0	0.5	0.25	0.25
FFID	3.7	3.5	3.4	3.7	3.7	3.7	3.8	3.6	4	3.7	0	0.3	0.2	0.3	0
Jana Vikas	3.1	2.8	2.3	3	3.4	3.8	3	2.9	3.9	3.8	0.7	0.2	0.6	0.9	0.4
NEDSSS	3	3.4	2.9	3.5	3.1	3.3	3.5	3.2	3.7	3.2	0.3	0.1	0.3	0.2	0.1
RGVN	3.6	3.7	3.4	3.4	3.5	3.7	3.7	3.4	3.4	3.6	0.1	0	0	0	0.1
SDS	2.8	3.6	3.2	3.9	3.5	3	3.8	3.4	4	3.9	0.2	0.2	0.2	0.1	0.4
Smile	3.5	3.5	3.5	3.5	3.75	4	3.8	4	4	4	0.5	0.3	0.5	0.5	0.25
SS	3.1	3.7	3.5	3.8	3.8	3	3.5	3.7	4.1	3.6	-0.1	-0.2	0.2	0.3	-0.2
VTRC	2.5	3	2.9	3.1	3.3	3	3.25	3.5	3.8	3.5	0.5	0.25	0.6	0.7	0.2
TOTAL	3.1	3.3	3.1	3.4	3.5	3.5	3.5	3.5	3.8	3.7	0.3	0.2	0.4	0.4	0.2

VTRC, an SPO supported by Red een Kind, improved its capabilities through an MFS-supported training workshop on theory of change, and another workshop on participatory integrated child development. This organisation managed large improvements in the scores for the capabilities to Adapt and Self-renew and to Relate to External Stakeholders.⁴²

Samarthak Samiti (SS) was the only organisation for which small decreases in capabilities were registered. It now “receives feedback on its strategies and policies from a more diverse set of stakeholders,”⁴³ and reduced funding made it more cost effective.

A recurring theme in the evaluations in India is the effect of reduced or terminated donor funding which in many cases seems to have been positive. While the advantages of reliable long-term partnerships are well known, these findings suggest some caution:⁴⁴ It is possible that the security of a long-term relationship led to complacency: actions to strengthen the organisation which were feasible were not undertaken until the SPO was forced to do so by lack of funds.

There are substantial differences in the quality of the studies in the CD component of the India evaluation. In some cases extensive use was made of informants other than the SPO staff, unlike in many of the other country studies. Also, in some cases the evaluation is commendably careful in its approach

⁴¹ Number of decimals in 5C score changes is displayed as given by country teams.

⁴² Table 3.18

⁴³ India Endline report , Samarthak Samiti, p. 35.

⁴⁴ This is not a point made in the reports.

to the attribution question. This makes it an impressive model for what can be done with qualitative analysis. In other evaluations, like that of Smile, alternative explanations are not seriously considered.

Indonesia

In Indonesia the CD component included evaluations of 12 SPOs. Table 2.18 shows that on average the changes in their 5C scores were modest and only some were negative.⁴⁵ The average improvement for the capacity to deliver on development objectives was a modest 0.2.

Large improvements (over 0.5) were found for the Capacity to Act and Commit for organisations ASB, RA and WII, the Capacity to Deliver on Development Objectives for ASB and WIIP, and the Capacity to Relate to External Stakeholders of WIIP.

A key development for ASB was a radical change in leadership style, apparently as a result of a HIVOS-sponsored workshop. The organisation also benefited from an extension of the network of ASB. When the TIFA foundation agreed to become the new sponsor, the financial situation of ASB improved resulting in better office infrastructure, an increase in programme activities and more staff. The TIFA funding also led to better employment conditions, including health insurance. Similarly, better financial reporting could be traced to a non-MFS supported training program. ASB became more visible through the public events it sponsored. This improved its reputation, thus facilitating contacts with external parties, including journalists and academics, which in turn led to learning and improved capabilities.⁴⁶ For ASB, the support of Hivos played a role but there were multiple other reasons for their improved scores.

Table 2.18 - Indonesia 5C score changes

SPO	Baseline					Endline					Changes				
	A&C	Obj	A&SR	Rel	Coh	A&C	Obj	A&SR	Rel	Coh	A&C	Obj	A&SR	Rel	Coh
ASB	3	3.3	3.1	3.6	3.5	3.8	3.9	3.6	4.1	3.6	0.8	0.6	0.5	0.5	0.1
ID	3.1	3.6	3.1	4.1	3.9	3.6	3.6	3.1	4.1	3.9	0.5	0	0	0	0
Lembaga Kita	2.5	3.3	2.6	3.4	2.5	2.7	3.4	2.6	3.7	2.9	0.2	0.1	0	0.3	0.4
ECPAT	3.3	3.3	3.2	3.1	3.5	3.7	3.7	4.1	3.6	4	0.4	0.4	0.9	0.5	0.5
GSS	2.8	3.2	2.9	3.4	2.9	2.9	3.2	3.1	3.6	2.9	0.1	0	0.2	0.2	0
PT.PPMA	2.9	3.2	2.6	3.5	3.2	3.2	3.3	3	3.8	3.2	0.3	0.1	0.4	0.3	0
Rifka Annisa	2.9	2.7	2.8	3.5	3.5	3.6	3	3.2	3.8	3.6	0.7	0.3	0.4	0.3	0.1
WIIP	3.4	3.4	3	3.6	3	4.2	4	3.8	4.2	3.6	0.8	0.6	0.8	0.6	0.6
YADUPA	3.4	3.3	3.3	3.7	3.7	3.7	3.3	3.5	3.7	3.9	0.3	0	0.2	0	0.2
YK	3.2	3.4	3.3	3.9	3.2	3.4	3.5	3.3	4	3.2	0.2	0.1	0	0.1	0
YPI	3.4	3.2	3.1	3.7	2.9	3.7	3.7	3.4	4.1	3.1	0.3	0.5	0.3	0.4	0.2
YRBI	2.7	3	2.5	3.2	2.7	2.4	3	2.3	3.4	2.7	-0.3	0	-0.2	0.2	0
TOTAL	3.1	3.2	3.0	3.6	3.2	3.4	3.5	3.3	3.8	3.4	0.4	0.2	0.3	0.3	0.2

⁴⁵ Two out of 60. These negative changes were quite small.

⁴⁶ Indonesia Endline report, Aliansi Sumut Bersatu, p. 27.

For RA the report is unfortunately unclear regarding the extent to which the substantial improvement recorded can be attributed to MFS. This also applies to the YK and PT.PPMA reports which carefully describe changes in capabilities, but remain quite vague on attribution. Table 2.18 shows that the changes to be explained are in any case modest. For YPI there are clear indications that staff quality improved through MFS (training) activities.

The Wetlands International–Indonesia Programme (WIIP) registered substantial improvements in capacities. Nevertheless, MFS supported CD was quite modest in terms of funding: € 4,000 for a workshop.⁴⁷ WIIP staff attributed much of their improved capacity to their participation in an international community related to wetlands unrelated to MFS.⁴⁸ They did credit MFS interventions with bringing about key changes in the capacities of their organisation.⁴⁹ The mechanism for this impact is not entirely clear from the report.

In the case of ECPAT only fairly minor capacity building initiatives were MFS- supported so major impact was unlikely.⁵⁰ At ID improvements could be traced to changes in office space funded by Cordaid but also to a series of MFS supported training events which enabled ID to spread information successfully. Capacity change was minor at LK, Good Shepherd Sisters (GSS)⁵¹ and YRBI. This is somewhat disconcerting in the RBIF case since there was a costly series of activities under the heading “Making Markets Work for the Poor” training which only started in March 2014, not long before the endline⁵²; on the other hand ICCO funding ended in late in 2013.

Liberia

The five organisations covered by the evaluation are BSC, DEN-L, NAWOCOL, REFOUND, and RHRAP. Changes in capabilities were in general positive but modest, with two notable exceptions. Large improvements in scores (0.5 or more) were recorded for NAWOCOL (for 3 out of 5 scores) and BSC (2 scores).

Reduced ICCO funding affected several organisations. While this sometimes led to reduced capabilities because staff left, the effect was also positive in a few cases, inducing SPOs to diversify their sources of funding. In the case of BSC training in networking and monitoring, partly funded through MFS, was identified as a major cause for improved capabilities. During the evaluation period NAWOCOL developed a new strategy, to some extent due to reduction in ICCO funding, and improved its capabilities by doing so.

⁴⁷ Indonesia Endline report, Wetlands International Indonesia, p. 27.

⁴⁸ Indonesia Endline report,, Wetlands International Indonesia, p. 35.

⁴⁹ Indonesia Endline report,, Wetlands International Indonesia, p. 40.

⁵⁰ Indonesia Endline report,, ECPAT, p. 23. The report is not clear on the attribution question.

⁵¹ This report is somewhat confusing, e.g. section 4.1 is not about changes at GSS but at ECPAT.

⁵² Indonesia Endline report, YRBI, section 4.1. [The acronyms YRBI and RBIF are used interchangeably.]

Table 2.19 - Liberia 5C score changes⁵³

SPO	Baseline					Endline					Changes				
	Act & Commit Objectives	Adapt & Renew	Relate external	Coherence		Act & Commit Objectives	Adapt & Renew	Relate external	Coherence		Act & Commit Objectives	Adapt & Renew	Relate external	Coherence	
BSC	2.8	2.8	2.6	4	3.3	3.5	3	3.1	4.4	3.6	0.7	0.2	0.5	0.4	0.3
DEN-L	3.5	3.4	3.1	3.9	3.9	3.8	3.8	3.4	4.1	3.9	0.3	0.4	0.3	0.2	0
NAWOCOL	1.6	1.75	2.4	2.25	1.82	2	2.25	2.9	2.75	2.3	0.4	0.5	0.5	0.5	0.48
REFOUND	2	2.4	2.5	2.1	2.25	2.1	2	2.5	2.3	2.5	0.1	-0.4	0	0.2	0.25
RHRAP	3.2	3	3.1	3.5	3.7	3.3	3	3.4	3.6	3.7	0.1	0	0.3	0.1	0
TOTAL	2.6	2.7	2.7	3.2	3.0	2.9	2.8	3.1	3.4	3.2	0.3	0.1	0.3	0.3	0.2

The Liberia evaluation used quite detailed theories of change, set out in graphic form as so-called causal maps in the report. These maps are often of such daunting complexity that it is a difficult task to identify any of the pathways shown.⁵⁴ Attribution is based entirely on staff opinions; there is no systematic comparison in the process tracing of alternative explanations for the changes observed.

Pakistan

For Pakistan capacity development was evaluated at four SPOs: LSF, Awaz, PFF and Scope.

For LSF capabilities improved considerably, notably for Act and Commit (+0.7) and Relate (+1.0).⁵⁵ The report notes that CD is not an objective of the organisation; therefore there is no theory of change. The evaluation fails because the baseline was supposed to be used to articulate a theory of change. The staff provided plausible reasons for changes observed, but only in the endline workshop.⁵⁶ The reasons identified in the process tracing explain *negative* changes (notably better salaries offered by other organisations), while in fact scores improved. MFS-initiated trainings were considered quite useful by LSF staff.

For Awaz the capability to Act and Commit declined while the scores for the other four capabilities improved.⁵⁷ The decline is due to the inability of the SPO to attract and retain qualified staff. Staff are dissatisfied with the Awaz HR procedures. Improvements in the Capability to Relate appear to be brought about by MFS funding which enabled Awaz to participate in various networks. The evaluation notes that much of what led to the improvements predates MFS funding. In that sense, the MFS contribution was quite limited.

⁵³ Number of decimals in 5C score changes is displayed as given by country teams.

⁵⁴ For an example see Liberia Technical Reports, p. 96.

⁵⁵ Pakistan Technical Papers, LSF, p. 37 (PDF p. 248).

⁵⁶ Pakistan Technical Papers, LSF, p. 39 (PDF p. 251).

⁵⁷ Pakistan Technical Papers, Awaz, p. 33 (PDF p. 198).

Table 2.20 - Pakistan 5C score changes⁵⁸

SPO	Baseline					Endline					Changes				
	Act & Objectives	Adapt & Renew	Relate external	Coherence		Act & Commit	Objectives	Adapt & Renew	Relate external	Coherence	Act & Commit	Objectives	Adapt & Renew	Relate external	Coherence
Awaz	4.2	3.2	3.3	3	3.3	3.5	3.5	3.6	3.5	3.3	-0.7	0.3	0.3	0.5	0
LSF	3	3	3.5	3	3.5	3.7	3.3	4	4	3.8	0.7	0.3	0.5	1	0.3
SCOPE	3.3	3.3	3	3	3.5	3.3	3.5	3.5	3.8	3.5	0	0.2	0.5	0.8	0
PFF	3.8	3	2.5	3.25	3.25	3.7	3.5	3	3.5	3.5	-0.1	0.5	0.5	0.25	0.25
TOTAL	3.6	3.1	3.1	3.1	3.4	3.6	3.5	3.5	3.7	3.5	0.0	0.3	0.5	0.6	0.1

Capabilities improved at PFF. Unfortunately, the CD part of the report is completely descriptive: it addresses the first research question about the changes that have occurred, but it contains no analysis of the reasons for these changes and simply ignores the remaining evaluation questions.⁵⁹ No conclusions can be drawn from this report.

This is also true for Scope: the report describes changes in capabilities, noting some substantial improvements, but does not even begin to analyse the changes.⁶⁰ It states its conclusion that MFS support was important, but MFS funding was not in a few lines. There is no further analysis.

A general point about the CD component of the Pakistan evaluation should be stressed: the evaluation is superficial, even in the more detailed process tracing. Without a prior theory of change, as was done by LSF but not Awaz, theory testing, no matter how informal, is impossible because outcomes and expectations cannot be compared. Alternative explanations for the changes observed were not systematically considered in some cases (at LSF) or remained restricted to staff informants (at Awaz). As a result, much of the evaluation does not go beyond repeating the judgments of staff (in some cases only a few staff members). The basis for causal statements is therefore very weak in parts of the Pakistan case.

Uganda

On average, the scores for the nine projects evaluated in Uganda changed very little (Table 2.21), but are positive across the board. However, these averages hide significant differences between the projects. The Dado project, which focuses on peace and reconciliation activities, shows large positive changes on Act and Commit and Adapt and Self-renew and these improvements can be attributed to MSF-funded CD activities.

⁵⁸ Number of decimals in 5C score changes is displayed as given by country teams.

⁵⁹ Pakistan Technical Papers, PFF, pp. 24-31 (PDF pp. 81- 88).

⁶⁰ Pakistan Technical Papers, Scope, p. 16 (PDF p. 17).

The FOKAPAWA project also shows significant positive improvements. It focuses on women and empowerment, and MFS funding supports education activities. In the beginning, progress on CD activities was slow, due to delays in getting a Board of Directors in place; after this was resolved, things have improved. For instance, the General Assembly has increased from 255 to 410 participating groups.

The Health Child project capability score was already high at baseline measurement, but the capability to Act and Commit has significantly improved as a result of focused MCH training.

The RWECO, and TWaweza, projects show very little change, but that is not surprising: they already showed solid scores on all five capabilities, leaving little room for improvement.

A similar observation can be made for the SHU project, which performed quite well at baseline level already. A stronger focus of MCH delivery significantly improved the capability to Act and Commit, but that has not yet translated into improvements to Achieve Objectives.

The negative results reported of the HESP project can partly be explained by changes in focus over the evaluation period, from CD activities to support for project activities, in particular regarding HIV prevention and treatment. The scaling down of activities at War Child Holland headquarters could partly explain the decline in the scores for War Child Uganda.

Table 2.21 - Uganda 5C Changes

SPO	Baseline					Endline					Changes				
	Act & Commit Objectives	Adapt & Renew	Relate external	Coherence		Act & Commit Objectives	Adapt & Renew	Relate external	Coherence		Act & Commit Objectives	Adapt & Renew	Relate external	Coherence	
DADO	2.9	2.6	2.6	3.5	3.4	3.2	3.2	3.3	3.5	3.5	0.3	0.6	0.7	0	0.1
FOKAPAWA	2.9	2.5	2.9	3.3	2.9	3.1	3.1	2.9	3.6	3.4	0.2	0.6	0	0.3	0.5
Health Child	3.2	3.4	3.3	3.5	3.7	3.6	3.5	3.5	3.9	3.8	0.4	0.1	0.2	0.4	0.1
HEPS Uganda	3.2	2.8	2.6	3.7	3.7	3.3	2.7	2.9	3.7	3.5	0.1	-0.1	0.3	0	-0.2
RWECO	3.5	3.1	3.3	3.7	3.8	3.5	3.2	3.4	3.8	3.7	0	0.1	0.1	0.1	-0.1
SHU	3.4	3.3	3.1	3.5	3.7	3.5	3.1	3.5	3.7	3.8	0.1	-0.2	0.4	0.2	0.1
TWaweza	3.5	3.2	3.5	3.7	3.7	3.6	3.2	3.4	3.8	3.7	0.1	0	-0.1	0.1	0
WarChild	3.1	3.2	3.3	3.7	3.6	3	3.1	2.9	3.5	3.7	-0.1	-0.1	-0.4	-0.2	0.1
TOTAL	3.2	3.0	3.1	3.6	3.6	3.4	3.1	3.2	3.7	3.6	0.1	0.1	0.2	0.1	0.1

2.3 Civil society strengthening

This section discusses changes in the civil society arena around the selected SPOs in two ways. First, it provides an overview of changes in the CSI scores over the evaluation period, as assessed by the country teams. This includes the entire sample of SPOs. Nevertheless, these changes are not necessarily linked to MFS support. Second, the attribution of results to MFS II is discussed based on the SPOs for which teams carried out process tracing or contribution analysis.

Some caution is warranted in interpreting the results. As with the CD component, the CS assessments depend to a large extent on the perceptions of SPO staff. In attributing changes to SPO interventions and MFS II support, teams have attempted to explicitly consider alternative, and sometimes rival, explanations for selected outcomes. Regardless, the outcomes explained by these in-depth methods rarely constitute negative or unexpected changes.⁶¹ Typically, outcomes for which MFS II involvement was expected were selected.

By definition, activities within civil society take place in a context that is influenced by state, private sector and family actions. For example, civil society organisations act because the state fails to provide adequate healthcare, because corporations monopolise seed varieties, or because parents believe their daughter should be married at age fourteen. The evaluation teams have sketched these contexts in detail in the country level narratives and the individual SPO reports. For example, the Bangladesh team notes that the sometimes violent political struggle in the country has complicated NGO work, limiting space for civil society activities and bringing the danger that organisations are co-opted by political parties.⁶² In the DRC, many international organisations work more with the NGO sector than with the state, making the former an interesting source of employment and business opportunities.⁶³ We cannot summarise all the different contexts here. However, a general observation is that the laws and policies in place in Ethiopia, India, Indonesia and Pakistan which regulate the work of NGOs may limit the type of activities and sources of funding of the NGO. Often, this concerns the amount of foreign funding organisations can receive.

2.3.1 Changes in CSI scores

Note that changes for Bangladesh, DR Congo and Pakistan are interpreted as the difference between endline and baseline scores (both on a scale from 0-3). For Ethiopia, India and Indonesia, change is expressed by a number on a scale from -2 to +2, indicating considerable deterioration at the negative end, and considerable improvement at the positive end. Unfortunately, it is possible that differences in average scores between countries (Table 2.22) reflect that an evaluation team has taken a more or less critical stance, rather than actual differences in performance by the SPO.

⁶¹ The India team concedes this point: “There was a tendency to selecting positive outcomes achieved” (India Civil Society Endline report, p. 25).

⁶² Bangladesh Narrative Report, pp. 25-6.

⁶³ Joint MFS II Evaluation DRC, p. 18.

Table 2.22 - Average change in CSI score per country

Country	# of SPOs	Civic Engagement	Organisation	Practice of Values	Perception of Impact	Environment
BGD	16	0.4	-0.1	0.5	0.3	-0.6
DRC	19	-0.3	-0.1	-0.2	0.2	0.1
ETH*	9	0.8	0.4	0.0	0.9	-0.4
IND*	10	0.6	0.6	0.0	1	0.4
IDN*	10	0.6	0.9	0.3	0.9	1.1
UGA**	8	1.0	1.8	1.6	1.8	
LBR	0
PAK	7	0.2	0.0	0.3	0.3	0.1

*These countries have a different scale for relative changes (-2 to +2)

**UGA has a different scale: cumulative score of relative changes (-3 to +3)

DRC

On average, the DRC team did not find a significant improvement in the CS arena; positive and negative changes cancel each other out. For example, Table 2.23 shows that perception of impact somewhat increased (+0.2), but Civic Engagement of SPOs decreased on average (-0.3). The team writes that beneficiaries are generally satisfied, and that there are some concrete examples of policy influencing.⁶⁴ However, this is the only team reporting decreases along multiple dimensions (see Table 2.22).

The team is most positive about changes at IFDP (+0.5 on level of organisation, +1 on environment) and at the Salvation Army (an increase of 0.5 on three dimensions). Nevertheless, for the latter the narrative shows that the SPO had difficulty reaching the already modest CS objectives of establishing parent committees⁶⁵ (see Table 2.23). IFDP is considered a relevant actor in the field of policy influencing, one that actively contributes to Bukavu civil society.⁶⁶ Its role in the adoption of a new environmental law in the attribution section is discussed later. For the SPO VICO, however, the team recorded deteriorations along four of the five CSI dimensions, and concluded that it is unlikely that the intervention— a microcredit project—had led to any durable improvements in the position in society of victims of armed conflict, the intended civil society outcome. Their analysis identifies two main causes: VICO did not have the financial capabilities needed to implement a microcredit project, and Co-Financing Agency (CFA) Cordaid, aware of these shortcomings, did not sufficiently address them. This is an example of a project in which both intervention design and implementation failed.

⁶⁴ DRC Country Report, pp. 67-8. See also the text under section 2.3.2 and the Annex for an overview of analysed outcomes.

⁶⁵ DRC Technical Reports, p. 34.

⁶⁶ DRC Technical Reports, p. 348.

Table 2.23 - CSI Score Changes in DRC

SPO	Baseline					Endline					Change				
	Civic Engagement	Organisation	Practice of Values	Perception of Impact	Environment	Civic Engagement	Organisation	Practice of Values	Perception of Impact	Environment	Civic Engagement	Organisation	Practice of Values	Perception of Impact	Environment
ADI-Kivu	2	1.5	1.5	1.5	1	1.5	1	1.5	1.5	1	-0.5	-0.5	0	0	0
AFEM-SK	2	2	1.5	2	1.5	2	2	2	2	1.5	0	0	0.5	0	0
Caritas	2	2	1.5	1.5	1.5	1.5	2	1.5	2	1.5	-0.5	0	0	0.5	0
CME de Nyankunde	1.5	2	1.5	2	1.5	2	2	1.5	2	1.5	0.5	0	0	0	0
Groupe Jérémie	3	2	2.5	2	2.5	3	2	2	2.5	2	0	0	-0.5	0.5	-0.5
IFDP	2.5	2	2	2.5	2	2.5	2.5	2	2.5	3	0	0.5	0	0	1
KMS	1	1.5	2	1	1	0.5	1	1.5	1.5	1	-0.5	-0.5	-0.5	0.5	0
RECIC	2.5	2.5	2.5	2	2	2	1.5	2	1.5	1.5	-0.5	-1	-0.5	-0.5	-0.5
Réseau CREF	2.5	2.5	3	2	2	2.5	2.5	3	2.5	2	0	0	0	0.5	0
REMACOB	2	1.5	2	1.5	1.5	1.5	1	1.5	1.5	1.5	-0.5	-0.5	-0.5	0	0
RHA	2.5	2	1	2	2	1.5	1.5	1	1.5	2	-1	-0.5	0	-0.5	0
RRILRP	1.5	2	1.5	1.5	2	1	2	1.5	2	1.5	-0.5	0	0	0.5	-0.5
RFDP	2	2	1.5	2	2	2.5	2	1.5	2.5	2	0.5	0	0	0.5	0
Salvation Army	1.5	1	0	1	1	1.5	1.5	0.5	1.5	1	0	0.5	0.5	0.5	0
SOFIBEF	2.5	2	2	2	1.5	2	2	1	2.5	2	-0.5	0	-1	0.5	0.5
UPDI	2	2	2.5	2	2	1.5	2	2	2.5	2.5	-0.5	0	-0.5	0.5	0.5
VECO	2.5	1.5	2	2	2.5	2.5	2	2	2.5	2.5	0	0.5	0	0.5	0
VICO	2	1.5	0.5	1	1	1	1	0	0.5	1	-1	-0.5	-0.5	-0.5	0
CEPROF	1.5	1	1	1	0.5	1.5	1	1	1.5	1	0	0	0	0.5	0.5
TOTAL	2.1	1.8	1.7	1.7	1.6	1.8	1.7	1.5	1.9	1.7	-0.3	-0.1	-0.2	0.2	0.1

Ethiopia

The evaluation team found general improvements in the dimensions of Civic Engagement and Perception of Impact; no changes on average on practice of values, and deterioration in the environment dimension (see Table 2.24). The Environment score declined by 1 for four SPOs: Ethiopian Kale Heywet Church, Catholic Church Gamo Gofa and South Omo (CCGG&SO), MKC-RDA, and Organisation for Social Service for AIDS. This is related to a 2009 law on civil society organisations, which regulates the organisations. It prohibits organisations where more than ten percent of the funding is derive from foreign sources from employing human rights and advocacy initiatives. In addition, a 2011 administrative guideline limits

administrative costs to 30 percent of the total budget of these same organisations.⁶⁷ Clearly, some SPOs have managed to cope with these rules better than others.

From the CSI scores and accompanying narrative it also becomes clear that CCGG&SO has gone through a dramatic episode in the evaluation period: four of its five dimension scores decreased by 1. To account for this, the team points at the dismissal of three staff members who allegedly accused the organisation of fraud and filed a complaint with authorities. The ensuing investigation and audit dismissed all allegations, but the office of the SPO was closed during that time. This proved to be a major setback.⁶⁸ Nevertheless, the team did not obtain evidence that the health microfinance schemes set up by CCGG&SO were having an effect on the food security of targeted households.⁶⁹

A considerable improvement was recorded for the Ethiopian Rural Self Help Association, as Perception of Impact increased by 2. The SPO has constructive relations with the local government, which has a positive influence on the farmer market organisations that the SPO supports.⁷⁰

Table 2.24 - CSI Changes in Ethiopia

SPO	Baseline					Change ¹				
	Civic Engagement	Organisation	Practice of Values	Perception of Impact	Environment	Civic Engagement	Organisation	Practice of Values	Perception of Impact	Environment
African Development Aid Association	2	2	3	2	2	1	1	0	1	0
Education for Development Association	2	2	1	1	2	1	0	0	1	0
Ethiopian Kale Heywet Church	2	1	2	.	.	1	1	0	1	-1
Ethiopian Rural Self Help Association	2	2	.	1	1	1	1	0	2	0
CCGG&SO	2	2	2	2	2	-1	-1	0	-1	-1
RiPPLE	1	2	2	1
JECCDO	2	2	1	2	2	1	1	0	1	1
MKC-RDA	2	1	3	2	2	1	0	0	1	-1
Organisation for Social Service for AIDS	2	2	1	2	1	1	0	0	1	-1
TOTAL	1.9	1.8	1.9	1.6	1.7	0.8	0.4	0.0	0.9	-0.4

¹Change reflected on a scale of -2 to +2

India⁷¹

The civil society picture regarding the SPOs concerned in India is generally positive. Increases are recorded for Perception of Impact (+1), and to a somewhat lesser extent on Civic Engagement and Level of Organisation. On average, there is no change in the practice of values dimension (see Table 2.25). We

⁶⁷ Ethiopia Civil Society Endline report, p.11.

⁶⁸ CCGG&SO Endline report, p. 26.

⁶⁹ CCGG&SO Endline report, p. 22.

⁷⁰ ERSHA Endline report, pp. 7, 17.

⁷¹ The SPO REDS-Tumkur is not yet included, average scores and narrative may change slightly.

should take into account, however, that the team was unable to carry out a follow-up assessment for two SPOs in the sample (Cenderet and Prithvi Theatre) – the results of these organisations could have changed the picture.

Especially for the four SPOs working on MDG 1, the team finds that they have established new organisations capable of defending the rights of the people they represent. Moreover, the SPOs supported the creation of new networks. Six of the seven SPOs contributed to creating new or more intense relations between people from different backgrounds, which leads the evaluation team to claim that social inclusion improved. The four SPOs working on MDG 1 themes generally do not engage with the private sector. Only anecdotal evidence for impact at the household is available.⁷²

The Centre for Sustainable Agriculture (CSA) seems to have exerted very positive change on the civil society arena in which it operates. The team records an increase of 2 on the dimensions Civic Engagement and Perception of Impact, and a smaller increase on Level of Organisation, and Environment. CSA has established three more cooperatives, bringing the total to ten. In general, it has increased participation of women and strengthened internal organisation at all cooperatives.⁷³

Table 2.25 - CSI Score Changes in India

SPO	Baseline					Change				
	Civic Engagement	Organisation	Practice of Values	Perception of Impact	Environment	Civic Engagement	Organisation	Practice of Values	Perception of Impact	Environment
CECOEDECON	3	3	3	2	2	1	1	0	1	1
Cenderet	3	2	2	2	1
Centre for Sustainable Agriculture (CSA)	2	2	3	2	3	2	1	0	2	1
Centre for Workers' Management (CWM)	2	2	1	1	1	0	1	1	0	-1
Gene Campaign	3	3	1	2	3	0	0	-1	1	0
Gram Vikas	3	2	3	3	2	0	1	0	1	1
NNET/Legal Cell for Human Rights	2	1	1	2	3	1	0	0	1	1
NINASAM	3	3	2	2	3	1	1	0	1	0
Prithvi Theatre	1	2	2	1	2
REDS-Tumkur	3	3	3	2	2	0	0	0	1	0
TOTAL	2.6	2.4	2.2	1.9	2.1	0.6	0.6	0.0	1	0.4

⁷² India Endline report, MFS II Evaluation pp. 49-50. There is no overlap between the MDG and CS samples for India; if there had been, the two evaluation components could have reinforced each other.

⁷³ Centre for Sustainable Agriculture Endline report, pp. 22-30.

A lower level is only found for Centre for Workers' Management (CWM) (-1 on Environment) and Gene Campaign (-1 on Practice of Values). CWM did not develop the coping strategies necessary for an environment that is progressively less favourable due to rising informal employment, economic decline, and legislation that prohibits the formation labour unions with foreign funds.⁷⁴ The Jharkand office of Gene Campaign, and the farmer groups that it supports, do not know about its vision, mission and financial resources, which are determined and managed at the national level.⁷⁵ In its summary, the team also comments that there is a slightly negative change on Perception of Impact: respondents had expected Gene Campaign to be capable of expanding to other areas, but its collaboration with other NGOs in the Jharkand area declined. However, this assessment is not reflected in the dimension score (+1).⁷⁶

*Indonesia*⁷⁷

Scores reported by the Indonesia team are positive across the board: no decreases were recorded for any of the SPOs (see Table 2.26). This may indicate a bias to downplay negative change and overstate improvements. What also stands out is a significant improvement in the dimension Environment, the largest increase of all countries. SPOs in the Indonesia sample generally seem to cope well with the challenges and opportunities presented by the environment in which they operate. For example, in its lobby for incorporating community-based forest management in forest policy, WARSI built on the opportunities provided by both a Presidential Instruction and a Corruption Eradication Committee statement on prioritising forest conflict resolution.⁷⁸ In another field, Yayasan RUANGRUPA uses the internet to showcase artwork and promote events, making good use of the strategic opportunities offered by increasing internet access in Indonesia.⁷⁹

WARSI and LPPSLH show positive change along all dimensions, with considerable improvement on Perception of Impact (+2). WARSI has successfully promoted its community-based forest management model and has influenced provincial forest policy in West Sumatra.⁸⁰ According to the team, LPPSLH established well-functioning crystal sugar cooperatives, and has a strong market position and a good quality control system in place. Also, LPPSLH and the cooperatives improved collaboration with district officials.⁸¹

⁷⁴ Centre for Workers' Management Endline report, p. 22.

⁷⁵ Gene Campaign Endline report, p. 19.

⁷⁶ Gene Campaign Endline report, pp. 7, 20.

⁷⁷ Two reports do not report average dimension scores.

⁷⁸ KKI-WARSI Endline report, p.33.

⁷⁹ Yayasan RUANGRUPA Endline report, p. 21.

⁸⁰ KKI-WARSI Endline report, p.46.

⁸¹ LPPSLH Endline report, pp. 25-6.

Table 2.26 - CSI Changes in Indonesia

SPO	Baseline					Change ¹				
	Civic Engagement Organisation	Practice of Values	Perception of Impact	Environment	Civic Engagement Organisation	Practice of Values	Perception of Impact	Environment		
Combine Resource Institution	2	2	2	1	2	1	1	0	0	1
Common Room	2	2	2	2	3
WARSİ - Komunitas Konservasi Indonesia	3	2	2	2	2	1	1	1	2	2
ELSAM	3	2	2	3	2	0	1	0	1	1
FIELD	2	2	2	2	2	1	1	0	1	0
Kantor Berita Radio	2	3	3	2	2	0	1	0	0	1
Koperasi Wana Lestari Menoreh (KWLM)	1	1	1	2	1
LPPSLH	2	2	2	2	1	1	1	1	2	1
Non-Timber Forest Products – Exchange Programme	2	1.5	2	1.5	1	0	0	0	0	1
Yayasan Ruangrupa	3	2	2	2	2	1	1	0	1	2
TOTAL	2.2	2	2.0	2.0	1.8	0.6	0.9	0.3	0.9	1.1

¹Change reflected on a scale of -2 to +2

Pakistan

Table 2.27 shows that on average, changes in the civil society arena around the Pakistan SPOs are not large. This shows considerable variation in scores, given on a 0-3 scale. For example, SPOs Awaz, Bedari and Shirkat Gah show improvements across the board, with the latter significantly increasing scores for Practice of Values and Perception of Impact (both +1). On the other hand, Lok Sanjh Foundation (LSF) and Madadgaar show lower scores on two dimensions each.

For Shirkat Gah, the evaluation team finds that within the dimension of Practice of Values, it has considerably improved the inclusion of target groups in its decision-making, and it is transparent toward all staff about financial information. Moreover, the team claims that Shirkat Gah is better able to deliver services that respond to the basic needs of its target group, and that it has become more successful in influencing government policy on women’s rights.⁸²

⁸² Pakistan Follow-up Report, pp. 120-2 [pages pdf file, report pages are not numbered].

Table 2.27 - CSI Changes in Pakistan

SPO	Baseline					Endline					Change				
	Civic Engagement	Organisation	Practice of Values	Perception of Impact	Environment	Civic Engagement	Organisation	Practice of Values	Perception of Impact	Environment	Civic Engagement	Organisation	Practice of Values	Perception of Impact	Environment
Awaz CDS	2	2.2	2	2.3	2	2	2.6	2.5	2.7	2	0	0.4	0.5	0.4	0
Bedari	0.5	2	2	2.3	1.5	1	2.2	2.5	2.67	1.5	0.5	0.2	0.5	0.37	0
LSF	2	2.2	2.5	2.3	1	2	1.6	2.5	1.7	1	0	-0.6	0	-0.6	0
PFF	2	2	2	2	1.5	2	2	2.5	3	1.5	0	0	0.5	1	0
Shirkat Gah	2	2.8	2	2	2	2.5	3	3	3	2	0.5	0.2	1	1	0
SCOPE	1	2.4	2.5	2	1	1.5	2.4	2.5	2	2	0.5	0	0	0	1
Madadgar LHRLA	2.5	2.4	3	3	2	2.5	2.4	2.5	3	2	0	0	-0.5	0	0
TOTAL	1.7	2.2	2.2	2.1	1.4	1.9	2.3	2.6	2.6	1.7	0.2	0.0	0.3	0.3	0.1

The evaluators explain Lok Sanjh's decrease in Level of Organisation indicators (-0.6) by pointing at its poorly maintained website; Lok Sanjh was actively using its website and social media at baseline, and its lower level of involvement in international networks. In addition, a lower score is recorded for Perception of Impact (-0.6). The SPO did not fully base changes in its MFS II project on an assessment of the needs of beneficiaries, and was unable to influence government or international organisation policy during the evaluation period.⁸³

Table 2.28 - CSI Changes in Uganda

SPO	<i>Relative change from Baseline to Endline¹</i>			
	Civic Engagement	Level of Organisation	Practice of Values	Perception of Impact
GWED-G	1	2	5	-2
KRC	1	4	1	2
Mango Tree	1	2	3	3
SEATINI	3	2	1	3
SACU	1	-1	0	3
UCMB	0	4	1	0
UGMP	0	1	0	2
VECO Uganda	1	0	2	3
Total	1.0	1.8	1.6	1.8

¹Cumulative scores of relative changes (on a scale from -3 to +3) per component.

⁸³ Pakistan Follow-up Report, Lok Sanjh Foundation, pp. 46-8.

Uganda

In Uganda, eight projects were evaluated on their Civil Society Strengthening components. On average, all show significant improvements. The GWED project, which champions human rights and builds capacity among partner organisations, saw a major jump in its score for Practice of Values over the two-year evaluation period. The KRC project actually has three different components: a major role in the development of a Regional Framework for CSOs, a rural communities programme, and leadership building among partner organisations. Overall the project shows improved performance, especially for Level of Organisation. With the exception of GWED (-1) and UCMB (no change), most programmes show significant improvements in Perception of Impact (+3).

2.3.2 Attribution of outcomes

The tables in Annex B provide an overview of the outcomes the evaluation teams analysed, to be able to attribute changes ascribed to MFS II support for SPOs. Teams used either process tracing to determine pathways or causal mechanisms that could explain the observed outcomes, or contribution analysis to assess the relative contribution of the SPO and the role played by other actors and factors. Evidence that would allow to confirm or reject these hypothesized explanations and contributions was gathered later. As discussed, there is considerable variation in the extent to which teams explicitly considered alternative explanations.

The evaluations for India and Indonesia applied the process-tracing methodology most consistently. Alternative explanations for observed outcomes are assessed explicitly and systematically. To a large extent, this also holds for the study carried out in Ethiopia. However, the analysis there is less focused. The DRC team applied contribution analysis. It is clear that they considered alternative explanations, but this is less explicitly documented than in the aforementioned countries. The depth of the Pakistan analysis varies: the evaluation of the SPO Bedari presents a good example of process tracing; the analysis of the remaining two SPOs is less convincing due to the absence of a systematic investigation of alternative explanations. The Bangladesh team assessed the entire theory of change of the SPO, rather than a specific outcome. An overview of other actors is presented in a table, but not always discussed explicitly in the analysis. The analysis of Ugandan SPOs is largely descriptive: there is no systematic investigation of alternative mechanisms that could also explain the observed outcomes. There seems to be a focus on linking SPO interventions to the outcome, rather than first defining a number of plausible explanations and only then examining which is most plausible.

This section only includes those countries for which alternatives were considered systematically and described in the reports. However, the tables in the Annex display for all countries the extent to which teams attribute the outcome to SPO interventions, or rather whether they judge it plausible that these interventions contributed to the outcome. The last column contains information on how the team links SPO actions to MFS II support and funds.

A recurring theme in the outcomes analysed is the support SPOs provide to community-based organisations and cooperatives. This can either consist of setting up or strengthening internal

organisation, connections to public and private sector or the market position of SPOs. The organisations supported are diverse, ranging from Block-level Development Committees in Rajasthan, India, to *Comités d'Alerte pour la Paix* in DRC and community-based forest management groups in Indonesia. In addition, CWM supports labour unions in India. There is a clear link to MDG 1 when agricultural cooperatives are supported. Three examples are discussed in the DRC study, among others with rice farmers, and in two studies in Indonesia, namely crystal sugar producers and rattan weavers). In all these cases, SPOs function as an intermediary between Dutch civil society organisations and local grassroots organisations.

Other SPOs were closely involved in the adoption of legislation. In DRC, IFDP played an important role in the adoption of a new environmental law by the provincial assembly in South Kivu. The sponsor of the bill in the assembly acknowledges the role played by IFDP.⁸⁴ In Indonesia, the SPO ELSAM played a necessary role in the realisation of a revised Law on Witness and Victim Protection.⁸⁵ Bedari, an SPO in Pakistan, played a lead role in drafting the Child Marriage Prohibition Bill, which was adopted by the Punjab Assembly in December 2014.⁸⁶

In other areas, organisations were also able to successfully influence policy. In India, CECOEDECON was instrumental in banning field trials of genetically modified seeds and increasing the quantity of produce farmers can sell at a minimum support price.⁸⁷ Whether these outcomes are indeed conducive to food security and economic development is questionable. Also in India, CWM helped a labour union influence minimum wages and dearness allowances.⁸⁸ The SPO JeCCDO played a considerable role in influencing the specific allocation of increased funds for education. However, the increase itself is explained by Ethiopian federal government policy rather than interventions by SPOs.⁸⁹ The actions of WARSI, working together with another NGO and the West Sumatra government, explain how community-based forest management has become an integral part of provincial forestry policy.⁹⁰

In many cases, the presence of certain external conditions is identified as a pre-condition to attaining outcomes. If so, SPO actions offer sufficient explanation for the changes. However, these conditions by themselves do not explain the observed outcome. For DRC, the general security situation is obviously an overriding concern when trying to achieve any civil society outcomes at all. Another relevant example is the SPO OSSA in Ethiopia. Its target group, people living with HIV, needs to have access to effective antiretroviral therapy in order for the SPO to be able to work on improving their social and economic capital.⁹¹ An example in India is the improved engagement of women in CSA farmer cooperatives; two pre-conditions are that Andhra Pradesh has a high concentration of women self-help groups, and the general trend that men are abandoning agriculture, giving women agricultural responsibilities beyond the traditional division of work. These two conditions enable CSA interventions—

⁸⁴ DRC Technical Report, p. 362.

⁸⁵ ELSAM Endline Report, pp. 26-9.

⁸⁶ Pakistan Follow-up Report, Bedari, p. 19.

⁸⁷ CECOEDECON Endline report, pp. 28-9.

⁸⁸ Centre for Workers' Management Endline report, p.31.

⁸⁹ JeCCDO Endline report, pp. 23-5.

⁹⁰ WARSI Endline report, pp. 29-31.

⁹¹ OSSA Endline report, pp. 21-4.

awareness raising, training of women, introduction of norms for the number of women on cooperative boards– to improve the engagement of women in the cooperatives.⁹²

The researchers emphasise that, in this context, interventions by other actors have led to the same result, and that therefore, multiple pathways exist that lead to the same outcome. This also holds for the interventions by NNET in India, which resulted in improved access to government schemes and programmes for Adivasi communities.⁹³ The DRC team also observes that in many cases there are other NGOs engaging in similar activities in the area. In the case of the rice cooperative in Sange, which was successfully established by SPO ADI, the researchers conclude that it is quite likely that another NGO would have targeted the area had ADI not done so first.⁹⁴

In many other cases, the evaluation teams consider it highly plausible that an SPO contributed to an outcome, but cannot disentangle the precise contribution of the SPO in relation to other actors. The India team termed this a ‘causal package’ if it considered SPO actions a necessary element, but not sufficient to explain the outcome. In India, this holds for CECOEDCON and the two outcomes analysed for CWM. Other examples include JeCCDO (Ethiopia) and the two outcomes for WARSI (Indonesia). In DRC, the population being more knowledgeable about their rights (Groupe Jérémie) and the development of *Mutuelles de Solidarité* (CEPROF) are two cases in point.⁹⁵

Another confounding factor in attributing outcomes to SPOs is the presence of pre-MFS II interventions (both MFS I and non-MFS). This is an issue throughout the MFS evaluations, beyond the CS component. For instance, the cooperative COOSOPRODA, strengthened through the interventions of the SPO VECO in DRC, already had a basic infrastructure in place due to participation in previous projects.⁹⁶

Researchers concluded that the SPO had not played a convincing or plausible role in achieving the researched outcome in a few cases only. However, it is notable that this is the case for three of the seven outcomes evaluated at Ethiopian SPOs. For example, EKHC is not convincingly helping enhance food support to vulnerable groups, due to the low nutritional value of foods provided. In addition, the idir coalition mobilised by OSSA does not explain improved referral to health services.⁹⁷ The DRC team evaluated twelve outcomes. No clear link was found in two of those: there were concerns about the impact of the community meetings organised by Réseau Haki na Amani, and the team judged it unlikely that the VICO project would lead to sustained improvements in the societal position of victims of armed conflict in Walungu.⁹⁸

⁹² Centre for Sustainable Agriculture Endline report, p. 25.

⁹³ Network of Northeast Tribes Endline report, p. 26.

⁹⁴ DRC Technical Reports, p. 289.

⁹⁵ DRC Technical Reports, pp. 171 and 289.

⁹⁶ DRC Technical Reports, p. 117.

⁹⁷ EKHC Endline Report, p. 28; OSSA Endline Report, p. 27.

⁹⁸ Annexes MFS II Evaluation DRC, pp. 434 and 614 of the pdf,

2.4 Summary

In sum, the evaluation results for the various MDG components show a very positive picture, with some notable exceptions, especially in the DRC and in Ethiopia. Moreover, these favourable scores are highly credible thanks to the use of well-established and verifiable, evaluation techniques and, in many cases secondary sources.

The picture for the CD component is more mixed. In the DRC the country team finds limited results and even some negative ones. In Ethiopia and India the results are mostly positive. In India, Indonesia and Liberia changes in capabilities are positive, but modest.

The synthesis team cannot take all these results at face value because it continues to worry about the validity of the evaluation method. Especially where the country teams rely almost entirely on the opinions of SPO staff without reference to outside experts, often without referring to a credible theory of change or alternative explanations for the observed changes in capabilities, the causal relations between the observed results and the outcomes are hard to determine.

The evaluations of the projects that aim to strengthen civil society arguably face the most problems. First, processes that improve the functioning of civil society take a long time, and the evaluations only covered a two-year period. Second, changes in the functioning of civil society are hard to measure and rely mostly on subjective impressions. Still, as it turns out, much can be learnt, especially when country teams carefully complement the changes in performance scores with additional information based on process tracing and other (qualitative) evaluation techniques. The resulting picture is generally positive, especially in countries where the civil society climate is conducive to the work of the southern organisations that are partners of Dutch NGOs.

3 Synthesis and analysis of results

3.1 Countries

When evaluation results are compared across countries it is important to keep in mind that evaluations were carried out by different teams. From the very beginning there was an enormous effort to ensure comparability across countries, as well as within countries and across time. An important outcome of this process was the adoption of a common methodology for most capacity development and civil society components of the MFS evaluation. For the MDG part there was already considerable methodological agreement at the outset. Nevertheless there may have been differences between country teams in the way common methodologies were applied in all three components, in particular since there was considerable room for interpretation in the scoring. What appear to be differences between countries may therefore in fact reflect differences between teams. In technical terms: a country/fixed effect cannot be identified.

India

The MDG component of the India evaluation covered a wide range of activities including micro-finance, livelihood of tribal women, and vocational training. It was affected by the fact that the activities to be evaluated had often started well before the 2012 baseline. In such cases there is no particular reason to expect any impact: the intervention may already have achieved its greatest impact and endline-baseline comparison only measure the extent to which that impact is sustained, if at all. Alternatively, the comparison underestimates the effectiveness of the intervention since it only picks up impact in addition to what had already been achieved at the time of the baseline. Three other issues noted in the India study also affect the other evaluations. First, the control group is sometimes affected by similar interventions so MFS impact is underestimated. Second, as already noted in the baseline synthesis report, the sample size may be too small to detect impact; then, an evaluation can falsely conclude that the intervention was not effective.⁹⁹ Finally, in three of the ten cases there were serious levels of attrition.¹⁰⁰

However, the evaluation does find significant effects in seven of the ten cases including all four cases where double-differencing was feasible. The other two cases often involved propensity score matching and an RCT. The MDG projects in India were well, or even extremely well, designed and well implemented. Comparison with benchmark cost data indicates that most of the projects were not only effective but also efficient.¹⁰¹ The project on community awareness in Bihar scored poorly both in terms of its design as in terms of reaching its objectives.¹⁰² The design problem was that the SPO tried to cover too many villages, a mistake which presumably could and should have been spotted early on.

⁹⁹ However, this has a serious effect for the main indicator only in one project (D6). In another project (D8) there is a power issue for an experimental component, but not for the main part of the evaluation.

¹⁰⁰ India Narrative Report, Table 3, p. 20.

¹⁰¹ The India Narrative Report correctly warns (p. 22) that the efficiency analysis is, inevitably, rather crude.

¹⁰² India Narrative Report, Table 4, p. 21.

In the CD component of the India evaluation, the evidence is mixed and in some respects puzzling.¹⁰³ For Samarthak Samiti, three of the five capabilities suffered slight deterioration. Many of the positive changes for the other SPOs are rather large (as much as 0.9) considering the short interval considered.¹⁰⁴ The scores for the capability to deliver on development objectives are sometimes difficult to reconcile with the scores for the other four capabilities which are considered intermediates for the purposes of this report. At one SPO the capability to deliver actually declines while other capabilities improve while for another, the capability to deliver improves much more than would be expected on the basis of the scores for the intermediate capabilities. A possible explanation is that once an organisation has achieved intermediate capabilities it may take some time before it is able to translate this into effectiveness in terms of development objectives.

During the endline workshop, SPO staff were asked for their view on the key changes in capacity at their organisation. The answers range from diversification of funding, to better monitoring and increased visibility. All SPOs except SDS, thought MFS played “a role in bringing about these changes”.¹⁰⁵

This is rather vague but the attribution question is tackled directly later, in section 5.2.3. There, process tracing is used to consider the attribution for changes in two of the five capabilities. This is done very carefully and convincingly. Alternative explanations for observed changes are considered for each SPO. The team is quite careful in attributing some changes entirely to MFS, some only partly and others not at all.¹⁰⁶ However, for the two capabilities selected for process tracing, the MFS contribution was a key determinant in the observed improvements. Here, as elsewhere, the evaluation team somewhat understates its case by noting that for some SPOs MFS was only one among many funders. It would have been clearer to attribute these changes to MFS in proportion to its financial contribution.

The Civil Society component of the India evaluation is also a good example of the attribution question. One of the six MDG outcomes can be attributed entirely to the activities of one SPO in the areas of water and sanitation, while in other cases there are multiple determinants. Similarly, under the theme of governance the outcome of greater political awareness can be attributed entirely to one SPO, Ninasam. Some SPOs lack critical capabilities and are therefore ineffective, which suggests that MFS support could have been better focused.

The India evaluation also reports on the correlation between the findings in the MDG and Capability Development components for SPOs covered in both. The results are striking, suggesting a remarkable disconnect.¹⁰⁷ Correlations between the capability scores and the MDG scoring are usually low or negative. In terms of ensuring that projects are well designed, that they are implemented as intended and that they reach their objectives the two capabilities, namely to Act and Commit, and to

¹⁰³ India Narrative Report, p. 36.

¹⁰⁴ For one SPO (BVHA) the score for the capability to Act and Commit is 0.5. (We consider this a substantial rather than a “slight” change.) Why salary increases are seen as an improvement in this capability (p. 29) is not entirely clear, although they might of course well contribute to such an improvement.

¹⁰⁵ India Narrative Report, p. 40.

¹⁰⁶ India Narrative Report, p. 46.

¹⁰⁷ India Narrative Report, p. 54.

Adapt and Self-renew are simply irrelevant.¹⁰⁸ For the synthesis team this is one of the most important findings of the evaluation.

Indonesia

In Indonesia the MDG and themes part of the evaluation covered five types of projects, focusing on: (1) reforestation, (2) poverty alleviation, (3) empowerment, (4) sexual education and (5) gender-based violence. In the case of reforestation there was little evidence of change regarding communities protect forests. This is not surprising since the project started well before the baseline was set, but there is clear evidence that the project increased the number of seedlings substantially and efficiently. However, the project had little effect on the general population in terms of disaster preparedness: the effects remained limited to a small bio-rights group. The projects that sought poverty alleviation ranged from reducing farmers' reliance on fertilisers and pesticides, to support for female micro-entrepreneurs. The empowerment projects helped communities to claim their rights through participatory mapping of their environment.

The evaluation found no impact of the four mapping projects in terms of the common quantitative indicators, possibly because these were not project-specific, but more likely because the evaluation took place too early to expect much effect. The evaluators noted that SPOs were more concerned with the map itself than with what the process of mapping was supposed to achieve in terms of awareness and knowledge. Villagers reported that they did not know what to do with the maps and that they looked at the SPOs for continued support. In this case, the projects seemed poorly designed.

The two projects on sexual education showed quite limited effects on attitudes and behaviour. Again, this may be because the project started well before the baseline was established: changes in attitudes may well have been achieved before the baseline took place. The projects on gender based violence involved counselling and legal assistance. The beneficiaries were self-selected so that the evaluation restricted itself to a before-after comparison. The projects appear to have been successful in terms of improved mental health and reduced domestic violence. The Rifka Annisa project, however, was extremely expensive, a point that was not mentioned in the report.¹⁰⁹ Of the 12 MDG projects one, project E2, was a disaster, see the Table on p. 51 of the country report; the text does not discuss this.

The Indonesia evaluation covered 12 SPOs related to the CD component, five of which were also covered in the MDG component. Their objectives ranged from stopping child prostitution to managing natural resources. The picture in terms of observed changes in capabilities over the two-year period is mixed: there were a few cases of deterioration, a substantial percentage of instances of no change (20

¹⁰⁸ The positive result is that the capability to Act and Commit is positively related to the observed results being relevant for the beneficiaries; the correlation coefficient for this relation is almost 0.6.

¹⁰⁹ The evaluation report notes that projects often received support from many sources and that the "interventions which could be traced back to MSF [sic] sometimes turned out to be very small" (Indonesia narrative country Report, p. 50). This point is not well taken: if an outcome can be attributed to an intervention financed from multiple sources it is reasonable to assign the impact to these sources in proportion to their contribution. Whether that contribution was small is irrelevant in this context.

percent) and improvement in the vast majority of cases.¹¹⁰ It is not expected that most organisations register large improvements in capabilities. Indeed there are only two SPOs with large (0.5 or more) changes in scores. Better human resource management (job security, health insurance) was considered important at ASB, better networking and an improved financial situation at WIIP. However, the latter was not selected for process tracing so that little can be said about what explains its success.

In the endline workshops the staff were quite clear in the contribution of MFS, e.g. in the case of Institut Dayakologi, supported by Cordaid, the SPO benefited from Cordaid support for rebuilding offices and for documenting and spreading information. In another case MFS support led to an organisational scan and subsequent improvements in the organisation, or to training activities. SPO staff were therefore quite explicit in what they considered was the impact of MFS. In general MFS is credited with improving staff capacity for planning, monitoring and evaluation.

The causal analysis is sometimes hard to follow, such as when *consequences* of capacity development such as improved case handling or more effective advocacy are presented as the underlying *reasons* for capacity changes.¹¹¹ Process tracing took place at five SPOs, focusing on two selected capabilities: to Act and Commit and to Adapt and Self-renew. In the case of ASB, a Hivos-initiated strategic workshop was identified as the trigger for key changes. The process tracing did not go beyond staff opinions. The report notes that for three of the five SPOs key changes can be attributed to MFS.

The Civil Society part of the study shows that NGOs have come under increased government scrutiny since a legal change took place in 2013. Different explanations are considered here and this is done quite well.¹¹² One finding is that an SPO (CRI) cannot be credited with an important policy change by the Ministry of Health because it did not succeed in generating public pressure. Similarly, revision of the law on witness and victim protection is credited to the activities of a coalition to which an MFS-supported CSO (ELSAM) belonged. The evaluation rejects alternative explanations. The activities are relevant both in terms of the theory of change as articulated in 2012 and in whether there is still a need for these activities. The evaluation shows that much is replicable, so organisations can learn from these experiences. An important contribution of three of the SPOs is an increase in cooperative structures and the success of another one in inducing government support for forest management.

Pakistan

The Pakistan evaluation is the smallest of the eight country evaluations: it covers three SPOs in the MDG component, four in the CD component, and seven SPOs in total. This number is obviously too small to allow any generalisations about MFS-effectiveness in Pakistan. It is worth stressing that the effectiveness of SPOs in this country is affected by government regulation and distrust, by extensive violence and by the country's strong patriarchal traditions reflected in the social exclusion of women. Legislation

¹¹⁰ Indonesia narrative country report, Table 3, p. 71.

¹¹¹ See, for example the Indonesia Narrative Country Report, Table 4, p. 81.

¹¹² Indonesia narrative country report, p. 109.

introduced in 2013 enables government regulation of international NGOs and domestic NGOs receiving external support.

The first project (LSF), aims at food security and adaptation to climate change. The current three-year activity (2011-2014) represents the continuation of two activities that started in 2004 and 2007, respectively. As such this is an extreme example of a project where the baseline was held long after the activity was initiated. The second project (JSL: Just and Sustainable Livelihoods) is an activity of the Pakistan Fisherfolk Forum (PFF) aimed at organising and mobilising fisherfolk and at carrying out advocacy to abolish the contract system of fishing. The third project (Awaz CDS) addresses gender-based violence in part through better informed decisions on sexual relations. This project is the continuation of an activity which started in 2006, long before the baseline.

Propensity score matching was used to establish the MDG contribution of these three projects.¹¹³ The LSF evaluation interviewed only 79 direct beneficiaries.¹¹⁴ The evaluation team found some differences between treatment and control villages but the project does not appear to have been successful: trainings were either not organised as intended or were poorly attended since farmers did not know about them. The project had no significant impact on household incomes, hunger yields or access to credit. Since the overwhelming majority of the farmers were not interviewed in the endline survey, results would have been meaningless even if they had been statistically significant. While the evaluators do not make this point they do assign the project a very low score (3) for the extent to which observed results can be attributed to MFS activity. Cost per beneficiary was much (70 percent) higher than intended, but the project's efficiency was not investigated. The evaluators gave this project a low score for both project design (5) and the extent to which it reached its objectives (4).

For the PFF project a new and larger sample was drawn for the endline survey with retrospective questions, so changes over time could be assessed. PSM estimates suggest that the project had no effect on fish caught, income, assets, or income diversification. However, the distribution of catch across households appears to have been improved as a result of the project in the sense that in PFF villages significantly fewer households experienced a reduction in fish caught over the period. This is an interesting finding although the mechanism to explain this result remains unclear.

For the evaluation of the MDG component of the Awaz CDS project the team abandoned the initial evaluation design. Rather than comparing treatment and control schools they compared pupils exposed to the LSBE curriculum, the key innovation in the project, with other pupils at the same project schools. PSM evidences that the project did not affect health-seeking behaviour, but that it was quite successful in terms of improving HIV/AIDS knowledge. Efficiency was, again, not considered.

All in all, the MDG effect of two of the three projects appears to have been positive whereas the results for the first project cannot be used.

¹¹³ Pakistan Narrative Report, p. 31 of the pdf (the report has no pagination).

¹¹⁴ Pakistan Narrative Report, p. 32 of the pdf (the report has no pagination). The report notes this and then continues: "However, the evaluation will retrieve the Intent-to-Treat (ITT) which is the impact of the project on direct and indirect beneficiaries together."

A fourth SPO was considered for the CD component: SCOPE, an environmental NGO. The improvements in the scores for the five capabilities in Pakistan were often quite large for the two-year period, as was also the case in other countries. For example, LSF experienced changes of 0.7 and even 1.0.¹¹⁵ In the case of LSF there is indeed evidence of very substantial progress: the organisation was able to formulate a four-year strategic plan with clear views on its mission, objectives, and core value. Oxfam Novib had requested such a plan, which undoubtedly contributed to this achievement. The evaluation report indicates that the organisation was flexible, renewing or redesigning its activities when changes seemed called for, often in response to a request from Oxfam Novib. LSF appears to be a good example that large changes in capability can indeed be achieved in a short period.

The key positive outcome is the ability of LSF to innovate, the key negative result its inability to retain staff. These two outcomes were selected for process tracing. The explanation for staff turnover appears simple. The competitive position of LSF is poor, partly because of its weak financial position: other organisations not only offer substantially higher salaries, but also better career options, tenure and secondary benefits. However, the two process tracing exercises basically used workshops to identify the staff explanations. There was no theory of change, nor a systematic exploration of alternative explanations. In that sense the process tracing did not achieve its intended purpose for this SPO.

For the Awaz CDS the team also found substantial improvements in capabilities, except for the capability to Commit and Act which deteriorated substantially. Here, too, the inability to retain staff seems to be a major problem. Staff identified non-competitive salaries as the main reason. Unlike the case of LSF, alternative explanations were considered.¹¹⁶ This part of the evaluation is very thoughtful and useful. It is noted, for example, that the SPO has a policy of hiring and then training weak candidates; this could explain high turnover rates.

For PFF many capabilities improved, notably that of achieving development results. The cost is considered very low but there is no comparison with a benchmark. Changes are attributed to Oxfam Novib.

SCOPE also registered improvements including some success in getting land reform on the political agenda. It achieved large progress in terms of its capability to relate by uniting a large number of organisations under an umbrella organisation. The evaluation notes that these changes cannot be attributed to MFS since “it did not provide funding for capacity building of SCOPE within the evaluated intervention.”¹¹⁷ This statement reflects a serious misunderstanding. Clearly, an MFS relationship can have causal effects other than through funding SPO activities. In fact Oxfam Novib was credited with causing changes in capability by insisting on a strategic plan. The attribution statement in the SCOPE

¹¹⁵ How the scoring framework was applied is not entirely clear. For example, “the involvement of a number of universities and research institutes in the debate about climate change” (Pakistan Narrative Report, p. 53 of the pdf, the report has no pagination) would seem to be an intermediate outcome rather than a result indicating the capability to achieve development results. The synthesis team is also puzzled by the remark that no theory of change was given since capacity development was not an objective of the project. This argument is also used to give no scores in three cases (Pakistan Narrative Report, p. 63 of the pdf, the report has no pagination).

¹¹⁶ It is not clear why this was not done systematically.

¹¹⁷ Pakistan Narrative Report, p. 86 of the pdf (the report has no pagination).

section is therefore not only erroneous in itself; it also implies that the methodology for attribution differed across SPOs, even within the same country.

Turning to the Civil Society component there were minor changes for PFF. Much of the analysis overlaps with that for CD. Therefore, a question arises on the usefulness of having both analyses. The claim that the test is double-decisive is nonsensical: alternatives have not been considered so it cannot be concluded that A was necessary and sufficient for B.¹¹⁸ No significant changes other than those in the fifth index were recorded for SCOPE. There is no clear story here on the MFS-attribution. For another SPO, Shirkat Gah (SG), process tracing was used to assess its impact on policy. Policy did change and this is credited to the activities of SG and similar organisations, basically by noting that their role “is well documented.”¹¹⁹ On the crucial question of MFS attribution the evaluation simply notes arguments on either side.

For Bedari (another CSO not covered in the other components) the civil society indicators improved considerably during the period. It developed, for example, an effective website for victims of violence. It achieved considerable policy impact by working with parliamentarians and government officials on issues such as the drafting of a bill prohibiting child marriages. Process tracing for Bedari was done at the Level of Organisation and for Perception of Impact. The report notes that while Bedari assumes that lack of knowledge of the human rights of women and children leads to violence against them this has not produced a testable theory of change. This is one of the best process tracing examples, a model for how it can be done.¹²⁰ Bedari worked through alliances, quickly acquired a leading role among organisations in this field and benefited from the liberalisation of the media. It was very influential at the stage when bills were drafted, having built very strong contacts with the administration at various levels and having responded effectively to queries and comments from politicians. Bedari’s achievement is particularly impressive given the hostile environment for SPOs in Pakistan and as a result their unimpressive effectiveness. The report considers and convincingly rejects alternative explanations.

That Bedari’s success is related to MFS funding is fairly clear, although other organisations state that the working in alliances was a reaction against donor influence.

Finally, no progress has been made on the scores at the SPO Madadgaar LHRLA. It seems to function well but the report is quite superficial.

Ethiopia

In Ethiopia the MDG component includes 13 projects with a wide coverage of MDGs.¹²¹ The set of project is very heterogeneous, aiming at raising the incomes of farmers, improving access to education, increasing the power of girls, reducing the use of firewood, building communities’ disaster resilience and improving sanitation. Many of the projects started long ago, well before the baseline was established so that no

¹¹⁸ Pakistan Narrative Report, p. 109 of the pdf (the report has no pagination).

¹¹⁹ Pakistan Narrative Report, pp. 126-7 of the pdf (the report has no pagination).

¹²⁰ It is unfortunate that this example was not followed for the other two CS process tracing analyses (PFF and Shirkat Gah).

¹²¹ Ethiopia Narrative Report (Endline Report MFS II Evaluation), Table 2, p. 12.

substantial impact should be expected. As the report stresses correctly, the double-differencing methodology can only measure the incremental impact of the intervention since the baseline was established. There is no reason to expect this *incremental* effect to be substantial or even positive: there may be some decline in a project's impact over time. In addition, the MDG evaluation was hampered by the fact that the control groups were affected by very similar interventions. For both reasons the results underestimate the true effect of the MFS support.¹²² Not surprisingly, in seven of the thirteen cases the evaluation does not find a significant effect. While lack of statistical power is an issue in four cases it can explain the absence of a significant effect in only one case (project C4).¹²³

In Ethiopia almost all projects were well, indeed often extremely, well designed.¹²⁴ Most of them were implemented as designed but many did not achieve all of their objectives. Three experienced serious problems (C4, C9 and C12). The education project (C4) was not implemented as planned; for the energy project (C9) the evaluation found no impact (possibly because not enough time had elapsed). The last of the three projects (C12, a water and sanitation project) would have been rated a great success with conventional evaluation methods since all relevant indicators improved: the incidence of diarrhoea fell, the time spent fetching water was reduced, and people started using better water containers and better latrines. However, the evaluation established that all such improvements also occurred (indeed more so) in the comparison areas. The lack of success of this project remains somewhat puzzling. The evaluation team suggests that local preferences were not considered¹²⁵ and that the project required behavioural change so that the endline came too early.

For three of the projects the efficiency analysis, an admittedly crude comparison with benchmark data, suggests that the projects were not cost effective.

The CD component of the evaluation covered nine organisations. In the case of the FSCE organisation which experienced many problems and got a new director only shortly before the endline survey, capabilities did not improve, they deteriorated. Very large improvements were recorded for AMREF, apparently as a result of staff training and greater engagement by the advisory council and the international board. As with most evaluations, the scores are based on the views of the SPO staff, expressed during the endline workshop. On the basis of their comments a "general causal map" was developed to explain the changes in the organisations capabilities since the baseline. However, this did not happen in all countries. In many cases the improvements can plausibly be attributed to MFS, as in the case of CARE which participated in workshops given by the mother organisation. The MFS contribution, however, remains rather vague: the report states that "MFS II funded capacity development interventions were mentioned as playing a role in the observed changes."¹²⁶ The summary on the attribution question

¹²² The first problem occurs in many of the evaluations and is likely to be identified correctly in all three components of the study. The second problem (technically: contamination of the control group) will not be picked up in the capacity development and civil society components since these do not use an explicit counterfactual. To that extent they may overstate the impact of the interventions.

¹²³ See the very useful summary: Ethiopia Narrative Report, Table 3, p. 18. See also Table 4, p. 20: project C4 received a very low score (3) for whether it was possible to attribute the observed changes to MFS.

¹²⁴ The notable exception is the dairy project (C10) which lacked a marketing component.

¹²⁵ Ethiopia Narrative Report, p. 19; the report offers no explanation for this point.

¹²⁶ Ethiopia Narrative Report, p. 45.

frequently indicates that MFS, internal factors and external factors all played a role.¹²⁷ In the Ethiopia evaluation there is, unlike in other reports, no explicit consideration of a theory of change and alternative explanations.

The civil society component of the evaluation shows that for three of the seven evaluated outcomes, there was no convincing link between SPO interventions and observed outcomes. The evaluation team follows the same process-tracing method as the India and Indonesia teams, but is less structured. In a number of cases, results cannot be disentangled.

The correlation matrix shows very little relation between the five capabilities and the scores in the MDG component. There is some correlation between the capability to Achieve Coherence and the extent to which a project was implemented as designed, also between the capability to Adapt and Self-renew and the extent to which a project reached all of its objectives.¹²⁸ The only substantial correlation (0.7) is between the capability to Act and Commit and the relevance of the outcomes for the intended beneficiaries.

Liberia

The Liberia evaluation was affected by the outbreak of Ebola in late 2013. This made the endline study for the civil society component of the study impossible, and dramatically affected the functioning of the SPOs concerned if only because they focused on the outbreak rather than their original objectives, and understandably so. The country's development efforts are seriously hampered by the legacy of its long civil war, which left a dysfunctional civil service in its wake.

Twelve SPOs were involved in the evaluation, many with a focus on women or children.¹²⁹ Four of the SPOs were covered by the MDG component of the study. Three of these (DEN-L, LSGCE and FOHRD) collaborated on one large project, PAMOJA. The evaluation of the remaining project (Business start-up Centre, BSC, supporting start-ups by young entrepreneurs with training and finance) suffered from serious attrition rates: 25 percent in the two-year period. The evaluation also suffered from low statistical power. No significant effects were found for the main indicators, possibly because of the statistical power issue and the short evaluation period. An alternative explanation is that the study could not control for selection effects since participants self-selected into the program. Double-differencing controls for observable characteristics, but the key difference between the two groups, the better ability of the successful applicants to write a business plan, cannot so be controlled for.¹³⁰ Also, and this would seem to be the obvious explanation although the authors are quite circumspect, start-up capital was not paid out so that an essential part of the intervention was not implemented.

¹²⁷ Ethiopia Narrative Report, pp. 53-54.

¹²⁸ Ethiopia MDG and CD report, p. 60.

¹²⁹ See the summary: Liberia Narrative Report, table 1.1, p. 8.

¹³⁰ The report correctly notes (p. 30) that a regression discontinuity design in this case would be appropriate, but this would require a much larger sample.

The capacity development component covers five organisations, including BSC. The evaluation recorded positive changes in the vast majority of cases.¹³¹ The changes were large in the case of BSC. This presents a puzzle: if the score for the organisation's capability to Act and Commit improved by 0.7, which is certainly a very large change, and its capability to deliver on development objectives by 0.25 then how can a critical part of its programme not be implemented, and how could there be no significant effects according to the MDG evaluation? The disconnect may be explained since it appears that SPO staff seem to have had a fairly narrow view of capabilities, emphasising improved networking skills and expansion as the key organisation changes.¹³²

Staff at the SPO identified reduced funding, notably from ICCO, as one of the key changes. Two SPOs were selected for process tracing: BSC and the human rights organisation RHRAP. For BSC the two capability changes selected for analysis could be attributed to MFS, notably because of MFS-supported training. ICCO support was also considered crucial for the observed changes in capabilities of RHRAP. In this case, reduced funding forced the organisation to improve its fundraising capability.

The team reflected on the CD evaluation noting that "it is questionable to what extent indicators can be compared across SPOs since they need to be seen in context, for them to make meaning" (p. 31).¹³³ This would seem to invalidate the entire exercise.

DRC

The evaluation for the Democratic Republic of the Congo covers nine SPOs, of which five under the heading MDGs and themes. The five projects range from ecosystem management to value chain development. In all five cases the project started before the baseline was done. Sample sizes were small and attrition rates quite large.¹³⁴ For the first SPO, AGIR, there is no evidence of impact on resource use and good governance. Trust in politicians, already low at the start, dropped significantly in the villages where treatment was carried out. This may reflect poor project design: the project generated resentment by taking plots close to villages. Indeed, all scores for this project, including the one for design, are poor.¹³⁵

A second project, the Salvation Army construction of six schools, had no effect on enrolment. The evaluators note that the high cost of schooling was the binding constraint on enrolment, and not the lack of school facilities. In this case the project was ill conceived. This project has poor scores as well, notably for design and for (not) reaching its objectives. The third project, SOBIBEF, aiming at empowerment of women and girls, had no impact other than a negative one on cassava yields. Most importantly, there is no evidence that the bargaining power of the intended beneficiaries improved. This is consistent with the evidence from focus group discussions where women claimed they are better off, but not because of SOBIBEF. For example rape victims noted some positive changes but did not attribute these to the SOBIBEF support homes.

¹³¹ See summary table on p. 26.

¹³² Liberia Narrative Report, p. 27.

¹³³ The phrasing in this section is often so peculiar that its meaning is unclear.

¹³⁴ DRC Narrative Report, table 3.2, p. 20.

¹³⁵ DRC Narrative Report, table 4.5, p. 30.

The situation is different for the value chain (VECO) project. Here the focus group consensus was that the project improved access. The quantitative analysis, however, finds some perverse effects on the use of fertilizer and improved seeds while the effect on yields is positive but not significant. The opinion of the evaluation teams is that the attribution problem was not addressed satisfactorily, possibly because of low statistical power. There are also no positive effects for the CEPROF projects. In summary there is no convincing evidence of positive MDG projects, either because the projects were badly designed and implemented, or because samples were too small to detect positive impacts.

The capacity development component of the evaluation mentions the theories of change which were articulated in the workshops. These theories are not described so it is not clear what is being tested. Many of the changes were negative: capabilities deteriorated substantially during the evaluation period.¹³⁶ Unlike in other countries, SPOs in the DRC are highly dependent on international NGOs for training their staff and during the evaluation period donors moved away from such training. Donors apparently prefer selecting strong SPOs over staff development. Many of the MFS donors, however, have made great efforts to improve the capabilities of SPOs. Unfortunately, the DRC report makes little effort to indicate what can and what cannot be attributed to MFS support. Alternative explanations are not discussed at all.

Seventeen SPOs were considered for the civil society analysis. The evidence, summarised in Table 6.3 in the report, is mixed. Changes were generally positive for five organisations, negative for seven and a mixture for seven others. Contribution analysis was used to answer the attribution question. Unlike for other countries this relied heavily on interviews with beneficiaries rather than with SPO staff. However, the way in which attribution was established is not clear from the text. The conclusion that donor funding was a crucial factor seems to have no clear basis in the analysis. At times the report seems to confuse categories: it reports that there is no convincing evidence of improved conditions, which is clearly not an issue for the civil society component.¹³⁷ Overall, however, the contribution analysis is quite useful. It details how MFS support matters through the SPO activities it stimulated, and not just through funding, which was reduced in this period. An important point is that SPOs have not collaborated effectively yet. Also, CD is often conceived as staff training, yet the capability that DRC SPOs most need is that for diversified fundraising.

Uganda

In Uganda, the country team evaluated eight projects that have a focus on one or more MDGs. They range from health interventions to projects with education components, to environmental and disaster relief, to post conflict projects. All projects are being implemented in the aftermath of a long period of civil strife and violence in the country. Most MDG projects show very favourable scores in Uganda. However, the two projects that explicitly address the consequences of civil strife show lower scores. The War Child project focuses on the psychological and social wellbeing of children and youth, and on protection against violence. Overall, the project shows very little impact on an array of indicators. The only exception is the

¹³⁶ DRC Narrative Report, table 5.2, p. 46.

¹³⁷ DRC Narrative Report, p. 73.

higher employment level of participants in the projects with vocational skills component. The second project is FOKAPAWA, a women and youth-focused project with income-generation and peace-building components. Again, the country team finds very few positive effects. The team cautions that those results are not conclusive because the project was already underway before the start of the evaluation period, and the two year evaluation period may have been too short to expect measurable results. Nevertheless, these examples do underscore the difficulty of making progress in post-conflict areas where violence has not been completely eradicated yet.

In addition, eight projects were evaluated on capacity development and nine on civil society strengthening. With few exceptions all projects saw significant improvements over the two-year evaluation period. Despite its troublesome past (with civil strife and violence in the North), the overall picture for Uganda is remarkably positive, with very little doubt that these developments are at least partly attributable to MFSII funding of the SPOs that implement the projects.

3.2 Priority areas

3.2.1 MDGs and themes

Table 3.1 summarises the performance scores of the 53 projects that had one or more MDGs as objectives. Six of those projects also addressed good governance issues, and four dealt with objectives relevant to fragile states. A few did both. Overall, the scores show that country teams judge these projects very favourably on almost all performance measures. In particular, they receive high scores for Design, Implementation, Relevance and Efficiency. Clearly, Dutch NGOs and their SPOs know what they are doing. But that does not automatically mean that all objectives are achieved in all cases.

Table 3.1 - Average evaluation scores per theme¹³⁸

Theme	Number of Observations	Average MDG Evaluation Scores					
		Design	Implementation	Objectives	Attribution	Relevance	Efficiency
MDG 1	21	7.3	7.7	6.0	5.7	6.9	6.8
MDG 2	10	7.2	7.0	5.6	5.6	7.2	7.2
MDG 3	5	7.4	7.2	6.5	5.4	8.4	6.8
MDG 4, 5 & 6	7	7.1	7.9	7.1	7.1	9.0	9.0
MDG 7a,b	7	7.1	5.9	4.0	5.9	6.4	5.0
MDG 7c	2	9.0	8.5	3.5	7.0	8.5	7.5
Good Governance	6	7.3	7.6	6.0	5.5	6.4	8.0
Fragile States	5	6.4	6.8	5.2	4.2	6.5	6.5
Total	53	7.5	7.5	5.9	6.1	7.4	6.8

¹³⁸ Some projects are included in several MDGs. Therefore, the sum of the number of observations in this table is larger than the total number of MDG projects in the MFS II evaluation, 53. The 'Total' row gives the average for the 53 projects, counting each project only once.

For MDG 1, eradicate extreme poverty and hunger, the average score for objectives achieved is 6. A closer look at the MDG 1 projects shows that the negative scores are concentrated in the DRC (3 projects) and Ethiopia (3 projects). In the DRC one project was badly designed, but the implementation of the two well-designed projects was hampered by continuing violence in the country. The projects in Ethiopia all showed mixed results, some goals were achieved, others were not. In one case it is clear that ambitions were too high. The project in Pakistan that scored low on objectives achieved, suffered from a string of implementation and budget problems. It had to be redesigned mid-way, and implementation was limited to one year. The funding Dutch NGO has reduced its activities in the region.

Only two of the eight projects that address MDG 2, education, score low for objectives achieved. One is in the DRC, where schools have been built and the children who attend get a school lunch. But children are not enrolling in higher numbers, because the cost of attending those schools is too high for the very poor target population. The other project is in Ethiopia. This teacher training project suffered from a myriad of problems, ranging from high teacher turnover (e.g. only one trained teacher is still at the project site), to schools where computer classrooms were supposed to be set up lacking electric power,, and high migration out of the area since this is a community of pastoralists; and security issues also played a role. A project reaching its objectives under those conditions would have been surprising. Yet, some of these problems appear to be highly predictable.

Two projects addressing MDG 3 related to gender scored low on objectives achieved. Again, one was located in the DRC and another in Ethiopia. The DRC project also deals with MDG 1 and is the same as described above. The Ethiopia project, a helpline for girls, and more generally children, and part of a nationwide "Girl Power" programme, was well designed but suffered from implementation problems. For instance, the project only worked with landlines rather than mobile phones, and received about one call per day. It was also hard to define a control group for the evaluation.

Of the three projects that address MDGs 4, 5, and/or 6 regarding health, only one scored low on objectives achieved. This project in India, also scored low on design. It is mainly a CD project, which seeks to increase awareness on women's reproductive health issues and rights. The project suffered from poor design but is deemed very relevant.

No fewer than four of the seven projects that address MDG 7a and 7b, on environmental sustainability, score low and, in some cases very low, on achieving objectives. One in the DRC, two in Ethiopia, and one in Indonesia. The project in the DRC was already discussed in the section of good governance, where it also had low scores. It also promotes the use of sustainable agroforestry techniques but the evaluation shows that these techniques are equally unpopular in the treatment and the control area. The projects in Ethiopia were mainly capacity development projects, and scored well, but did not show achievements on MDG 7a and 7b. For the project in Indonesia, to enhance the ability of indigenous communities to properly manage natural resources, the country team noted improved awareness of the importance of sustainability, but little change in practice.

One of the two Ethiopian projects on MDG 7c scores low on objectives achieved. This is the same project that had a low score on achieving results for MDG 7a and b. Clearly, CD was the main focus of this project.

Two of the six projects on the theme good governance score low on objectives achieved. Four of the five projects addressing fragile states issues score low on objectives achieved. Three of those are, unsurprisingly, in the DRC, and the other is in Liberia. All four projects have MDG issues as their main focus point, and have already been discussed above. The project in Liberia is a relatively small business development centre for start-ups. While it can reasonably be argued that improving the business climate can, over time, reduce the fragility of a state, the scope of the project does not match the challenge of state building in Liberia in this respect.

In sum, the scores given by the country teams on projects addressing the MDGs and themes are overwhelming positive, especially for Design, Implementation, Relevance, and Efficiency. Projects that do not achieve all of their objectives have a few things in common. First, and not surprisingly, it appears to be very difficult to work in the DRC or Ethiopia, where violence and lack of security make it difficult for even well-designed and well-implemented projects to live up to their ambitions. The labour of NGOs and the SPO who have to work under such stressful circumstances and yet are capable of obtaining any results at all is worthy of compliments. Second, projects that do not focus on one or more of the MDGs as their main objective (mostly CD projects), often score low on MDG achievements. It may just be a stress to expect that the long term objectives of capacity Building already show better results on MDG goals in the short term. Third, a few projects are badly designed, and do not take local circumstances sufficiently into account. Finally, some projects show mixed results and get a low score because their ambitions seemed to be unrealistically high.

In a number of cases projects can be easily improved if there is a better focus on the main objectives. And it is fully understandable that projects in countries that suffer from civil strife or its aftermath will not always be successful. But it is worth noting that in most of these cases the projects score high on relevance. Especially under such difficult circumstances, abolishing the project is not an option.

The scores of the Attribution question are addressed in chapter 5, together with an analysis of the evaluation process in general.

3.2.2 Capacity development

What can we learn from the results of the evaluations of Capacity Development? The scores the teams gave for the 5Cs (Table 2.15 - Table 2.21) suggest a positive finding. First, with few exceptions, the scores were already reasonably high (3 or higher) from the outset. It appears that the MFS organisations selected SPOs that had already achieved reasonable competence so that they could be expected to already be, or quickly become, effective in terms of development objectives. Second, while the evaluation period was very short, the changes in capabilities were generally positive and often substantial, certainly in view of the already high baseline scores. It is plausible that the heavy reliance on the views of SPO staff introduced a positive bias, both in the baseline levels and in the changes in 2012-2014. Nevertheless, with few exceptions, notably the DRC, the evidence on the 5Cs is positive.

To what extent can the positive changes be attributed to MFS? Here the evidence is mixed. For Congo the answer is simple: there was no clear evidence of MFS support causing a noticeable

improvement or preventing a greater deterioration in any of the five cases. In most other cases, however, the story which emerges is more complex. Outcomes almost always had multiple causes, and the causal stories developed in the workshops made clear that many of the channels had nothing to do with MFS support. Nevertheless, in many cases improvements in capabilities clearly *had* been triggered by MFS activities: training activities, a change in leadership induced by donor action, instruction in financial management, planning workshops, or insistence by the donor that a clear strategic vision should be developed. A very valuable outcome of this evaluation is that these channels and their relative importance are now clearly documented

However, quantifying this impact is not possible and the attribution itself is usually challenging. It is not impossible. This evaluation impressively shows is that informal theory testing can be taken further. When done correctly, the detailed process tracing was very careful in setting out a theory of change, teasing out its implications, confronting these with the evidence, and systematically asking whether other explanations would have fit the data better. It also allowed to move beyond the opinions of SPO staff who obviously will, even unconsciously, favour specific explanations. If these two conditions are not satisfied, then the evaluation still is valuable as a detailed and insightful description of an SPO in action, but not as an answer to the attribution question. Thus, this glass is half full: the best CD evaluations in this study set a standard for future evaluations.

For many countries, process tracing was limited to two of the five capabilities. In retrospect this is unsatisfactory. The capabilities selected for process tracing are not necessarily the ones with the most pressing questions on what drove capacity development for each SPO. In that sense some of the evaluation effort was wasted. Also, it is not clear whether there is any room between no CD evaluation and process tracing, which is costly and time consuming. Answering the attribution question without using process tracing is rarely convincing, partly because doing so relies heavily on the views of staff. The lesson seems to be that process tracing should not be selective but should become standard in CD evaluations.

As noted in Section 2.2, a surprising finding is that in many cases the end, or even the announcement of the end, of donor funding had a very healthy effect on SPO capabilities. This is important, particularly because evidence for this was seen in many evaluations. The implication is that donors should either limit the length of their involvement or make sure that SPOs do not heavily rely on a single donor.

An important issue for future evaluations is whether the assessment of the 5Cs is considered important in itself or is seen as instrumental. In the latter case, capabilities are relevant only to the extent that they enable the SPO to achieve its objectives. This connection is by no means self-evident. For example, in one Ethiopian case an SPO had quite good 5C scores in the baseline evaluation and also realised improvements in the following two years. Nevertheless it failed to achieve much in terms of MDG contributions.¹³⁹ Such a disconnect between capabilities and results begs the questions whether capabilities are measured correctly and the extent to which they are relevant.

¹³⁹ This was the example of HOAREC, discussed in Section 3.2.

3.2.3 Civil society strengthening

Perhaps the main lessons learnt from the evaluations of the MFS-sponsored projects that focus on Civil Society Strengthening is that their success heavily depends of the general environment in which they operate. This is very clear in the DRC where, due to continuing violence, SPOs work under very difficult circumstances. Not surprisingly, the country team found little evidence of improvements in civil society activities during the evaluation period. It is probably also the case that the two year evaluation period is too short to find significant improvements.

However, in areas where the political climate is conducive to civil engagement, significant positive results were found. In India for example, SPOs have been able to establish new organisations and defend the rights of their target populations, and social inclusion has improved. In Indonesia, SPOs have also been able to cope well with the challenges and opportunities that the environment presents. For instance, a community forest project has been successful in influencing provincial forest policy in West Sumatra.

For Ethiopia the picture is mixed. A 2009 law regulating civil society organisations makes it difficult for those organisations to operate, especially when foreign funding is involved. Still, the country team concludes that Civic Engagement and the Perception of Impact have improved in some cases. In one case, a project was able to succeed by building up good rapport with the local government.

For a variety of reasons it has proved difficult to attribute results to the MFS interventions. One problem is that in some cases, like in the DRC, many SPOs operate in the same areas and focus on the same objectives. Therefore, the observed good results may not be better than those obtained by other similar organisations. The subjective scoring of changes over time is also problematic. In some cases team bias that occurs by focusing on positive outcomes cannot be ruled out. This is especially of concern when the measurement of changes over time is not being complemented with more in depth analyses such as process tracing.

3.3 Efficiency

Table 3.2- Overview MDG efficiency scores

Countries	Number of Observations	Efficiency
DRC	.	.
ETH	13	6.7
IND	9	8.2
IDN	6	5.8
LBR	2	6.5
PAK	2	6.5
UGA	3	5.7
TOTAL	35	6.8

During the course of the evaluation it was decided that the efficiency analysis would be limited to the MDG components: it would have been too difficult to disentangle the budget for CD and CS activities from

general project and organisation budgets (see also Section 5.2 in Annex A). In addition, finding benchmarks for the objectives of these activities, for example changing an environmental law, was considered problematic. Therefore there was no assessment of CD and CS activities for this endline report, although there are certainly reasons for concern. The Uganda team, for example, notes that “not any organisation did have a clear definition of efficiency, and systems of efficiency measurement were not existing.”¹⁴⁰

The synthesis team prepared an overview of benchmarks for the unit costs of project outputs, collected from existing literature, as starting point for assessing the efficiency of the MDG projects. Country teams were invited to submit up to four outputs per project. The survey did not find benchmarks for all suggested outputs; in those cases, reference values for similar projects were given wherever possible. In the end, country teams had to decide on the suitability of the benchmark proposed. The scores are summarised in Table 3.2. There are only observations for a subset of 35 activities and this set is not random. The efficiency scores are subjective and therefore not fully comparable. However, the average score (6.8) appears reasonable. It should be noted that the differences between countries are substantial: mean country scores range from 5.7 (Uganda) to 8.2 (India).

The Uganda evaluation covered eight MDG activities, but efficiency was only assessed for three of them.¹⁴¹ The scores (7, 5, 5) are not encouraging. Unfortunately, they are not discussed in the narrative report or the technical papers.

The cost estimates in the Pakistan report compare favourably with the benchmark data. However, it is not entirely clear how these comparisons were used for the efficiency scores: for the first project the comparison is quite favourable but the score is a 5, for the second project the score is an 8, and for the third project no score is listed.¹⁴²

In the Business Start-up Centre Liberia, the results for efficiency are mixed: the cost of training are quite low compared to a benchmark study, but the cost of matchmaking is far higher than that of the benchmark estimate. The efficiency score (7) therefore aggregates two different results. For the PAMOJA project the study report that all activities are costly compared to the available benchmarks. Therefore, the score given, a 6, seems high.

For the DRC the evaluators attempted to answer the efficiency question but encountered formidable obstacles. Detailed budget information could only be obtained for two projects and it proved impossible to find meaningful benchmarks.¹⁴³ The DRC report notes that with the exception of VECO, SPOs seemed unconcerned with efficiency, likely as a result of the nature of their relationship with donor

¹⁴⁰ Uganda Endline Report Joint MFS-II Evaluation 2015, p. 71.

¹⁴¹ Joint MFS-II Evaluation Uganda Endline Report 2015, Table 17, p.30. One of these was Kampabits for which section 5.4 in the technical report presents data on cost per student. The text notes that these are higher than the benchmark. However, it does not indicate how this results in a score of 5.

¹⁴² Pakistan Follow-up Report MFS II Joint Evaluations, Narrative, pp. 53, 58, 63.

¹⁴³ DRC Synthesis Report, p. 81.

organisations: once a donor has approved a set of activities, SPOs feel no need to economise.¹⁴⁴ It should be noted that this attitude is less likely to occur the more diversified the funding of an SPO is.

The DRC team considered efficiency in a limited way, and only for two organisations. For the Salvation Army schools it noted that construction costs are very difficult to compare across settings. The cost of training appeared to be unnecessarily high as a result of avoidably high accommodation costs and the high cost of flying in trainers.¹⁴⁵ For VECO, efficiency was considered in some detail in the evaluation.¹⁴⁶ The cost for land rehabilitation appeared to be quite modest, but training in rice production was costly, at € 92 per farmer compared to benchmark estimates of about half that amount. Training in rice processing was even costlier: € 268 per participant. The prices paid for rice were deemed to be quite excessive. The evidence for VECO, the organisation which is described as atypically cost-conscious, is therefore not at all encouraging.

In Ethiopia, efficiency could not be assessed for all projects but the report contains estimates for most. The NVEA education project had suitable financial data for an analysis of efficiency. The data were used to estimate the cost of constructing a school. A comparison with World Bank data suggests that in this respect the project was quite efficient. The same conclusion was reached for the cost of teacher training.¹⁴⁷ For the ECFA project, the efficiency analysis focused on the cost of community conversations, which seemed to be quite effective in changing social values; these costs appear to be very high.¹⁴⁸

An overall efficiency analysis was not possible in the case of the FSCE project since it covered a range of different activities and for many of these there are no benchmarks available. The evaluation produced a rough estimate of € 12.50 per unit of community conversation; this appears to be reasonable, and falls within the range of benchmark estimates provided by AIID.¹⁴⁹ In the teacher training project (TTCA) there were no benchmarks for several project activities. Where a comparison was possible the project cost seemed to be much higher than in the benchmark.¹⁵⁰ In the CPDRR project the establishment of savings and credit groups was found to have been much more expensive than benchmark estimates. This applies *a fortiori* for the cost of planting trees (around \$17,000 per hectare).¹⁵¹

In contrast, the language training in the SILL project appears to have been very efficient.¹⁵² The cost of the solar home system for energy provided by HOAREC was estimated at € 660, whereas the Solar Energy Foundation Ethiopia offered something similar for € 167 in 2009. This large difference is obviously a cause for concern.¹⁵³ For the AI project the evaluation used benchmarks but doubts about comparability make

¹⁴⁴ DRC Synthesis Report, p. 49. A striking example is that travel is not planned in such a way that visits to different projects can be combined in a single trip.

¹⁴⁵ DRC Technical Reports, p. 30.

¹⁴⁶ DRC Technical Reports, p. 99, Table 19.

¹⁴⁷ Ethiopia Technical Reports, Table 14, p. 27 of PDF.

¹⁴⁸ Ethiopia Technical Reports, pp. 69-70 of PDF. Since very few children called the child helpline, the cost per call for the helpline was astronomical but the evaluation team decided, correctly in our view, that this indicator was not meaningful in this context.

¹⁴⁹ Ethiopia Technical Reports, pp. 125-126 of PDF.

¹⁵⁰ Ethiopia Technical Reports, pp. 187-189 of PDF.

¹⁵¹ Ethiopia Technical Reports, Table 5, p. 249 of PDF.

¹⁵² Ethiopia Technical Reports, Table 13, p. 314 of PDF.

¹⁵³ Ethiopia Technical Reports, p. 345 of PDF.

it difficult to draw a firm conclusion on efficiency. In the WASH project, however, the cost of wells dug is clearly far below the benchmark. Because this involves Ethiopia it is more useful than is the case for most benchmarks.

In India crude estimates for efficiency could be constructed for nine of the ten SPOs. Seven of these nine cases received excellent scores (8, 9 or 10) for efficiency. In the other two cases the score was a 6.¹⁵⁴

The Indonesia report includes the most detailed analyses of project costs. They are often supplemented by a cost-benefit analysis. While the team clearly made enormous efforts, they were not always able to obtain the relevant cost or benchmark data.

In the case of the WIIP reforestation project, the report includes a cost-benefit analysis, with favourable results. Comparing unit costs to benchmarks suggests, however, that the project was not efficient: cost per hectare was about \$ 1,200 compared to benchmark estimates ranging from \$ 399 to \$ 939.¹⁵⁵ An efficiency analysis of the Pt.PPMA project was impossible because the team could not obtain the relevant data on activities and beneficiaries for the MFS II period, which is in itself a worrisome finding.¹⁵⁶ The mapping activities of NTPF-EP appear to have been very cost-effective compared to a Tanzanian benchmark.¹⁵⁷ There is no efficiency analysis yet for the FIELD project. The SwissContact project was not cost effective when compared to benchmark data.¹⁵⁸ By contrast the DAKU project, implemented by YPI, appears to be quite cost-effective.¹⁵⁹

For the GREEN project, the evaluation team analysed the cost data in detail but no meaningful benchmark on which to base an efficiency analysis was available. This was also true for the KSP-QT, LRC-KJHAM Rifka Annisa and HuMa projects. In the case of YRBI benchmark information was found but the analysis was apparently still incomplete.

In conclusion, the evaluation fails to shed light on the efficiency question.¹⁶⁰ Two issues become apparent. One concerns the management of financial data. Many SPOs do not collect such data or do not have data systems which allow them to produce reports for various periods, sources of funding or types of activity. Bookkeeping is apparently often elementary and there appear to be no accounting systems in place which allow financial data to be used as a management tool, rather than something that is a requirement of accountability. The second issue is whether donor or recipient organisations are themselves focused on efficiency. This is clearly not the case at many organisations: with rare exceptions the evaluation team could not get financial data in a form suitable for efficiency analysis, and organisations

¹⁵⁴ India Country Report, Table 4, p. 21.

¹⁵⁵ Indonesia MDG Endline Report, p. 46 of PDF.

¹⁵⁶ Indonesia MDG Endline Report, p. 192 of PDF.

¹⁵⁷ Indonesia MDG Endline Report, p. 281 of PDF.

¹⁵⁸ Indonesia MDG Endline Report, p. 778 of PDF.

¹⁵⁹ Indonesia MDG Endline Report, p. 918 of PDF.

¹⁶⁰ The extent to which results could have been furthered is not clear. That the Ethiopia, Indonesia and India evaluations provided much more information on efficiency than the other reports is remarkable. As of now, no satisfactory explanation has been found for this discrepancy.

lacked the benchmark data against which to compare their performance. Clearly, this is a disturbing state of affairs.

A lesson for future evaluations is that efficiency analysis deserves much more attention, both in terms of the data that need to be available, and often is not, and teams analysing them in detail which in this evaluation was the exception rather than the rule.

Some of the results are quite striking; the costs of MFS supported activities are either far above or below benchmark estimates. This should not be so quickly dismissed as reflecting the inappropriateness of the benchmarks used. Rather, such striking results call for more research. If they turn out to be robust, then these results have important policy implications: some types of activities should be redesigned or simply abandoned, while others should be expanded and implemented in other contexts.

Cost considerations have become something of an afterthought in modern evaluation practice. This is entirely mistaken. That it was often difficult to obtain the data for an efficiency analysis indicates that the importance of such work has rarely been sufficiently internalised by the SPOs or indeed by their MFS donors.

The question whether it is sensible to exclude CD and CS activities from efficiency analyses needs to be revisited. The MDG evaluations were in many cases able to analyse the cost of training activities and to find relevant benchmarks. The CD and CS evaluations often covered training activities as well. In retrospect it is not at all clear why these should be excluded from an efficiency analysis. On the contrary, capacity development is often concerned with capabilities that are quite similar across organisations even if they pursue very different objectives. In this context efficiency analysis must be feasible.

3.4 Project size (budget and duration)

Table 3.3 shows the performance scores of the projects ordered by duration. There is a very clear pattern: without exception, projects 54 months or longer receive the highest scores. This is not surprising, because many of the projects focus on issues that, by their very nature, are difficult to solve. In addition, many are being implemented in very difficult environments. It takes time to get results. Also, in many cases the organisation that needs to implement these projects needs to expand, recruit and train staff, and procure equipment. With funding assured over longer periods of time, implementing agencies can learn and correct mistakes in design and implementing strategies. Of course, when taking a long term focus, the need for rigorous impact evaluations at regular intervals becomes even stronger.

Table 3.3 - Average scores based on project duration

Project Duration	# of projects	Design	Implementation	Objectives	Attribution	Relevance	Efficiency
Less than 36 months	16	7.7	7.2	5.9	5.3	6.9	6.1
Between 36-56 months	23	6.8	7.1	5.3	6.5	7.3	6.8
More than 56 months	14	8.5	8.4	6.8	6.3	8.2	7.4
All Projects	53	7.5	7.5	5.9	6.1	7.5	6.8

This pattern, that size matters, seems to be repeated when projects are sorted by budget size (see Table 3.4). With one exception, projects with a budget of one million Euros or more receive the highest scores. Especially on implementation, large projects do a lot better than small ones. The exception is on efficiency, where projects in the € 200, 000 to 400,000 range do slightly better than the largest projects.

Table 3.4 - Average scores based on budget size

Budget size	Number of Projects	Design	Implementation	Objectives	Attribution	Relevance	Efficiency
<200,000 EUR	18	7.5	6.8	6.1	5.9	7.6	6.6
200,000-400,000 EUR	17	7.5	7.5	5.7	5.9	7.1	7.3
400,000-1,000,000 EUR	8	7.8	8.3	6.3	6.0	7.8	5.8
>1,000,000 EUR	9	7.8	8.4	5.8	6.4	7.6	7.4
All Projects	52 ¹⁶¹	7.5	7.5	5.9	6.1	7.5	6.8

While the intuition that well-funded projects implemented over longer periods of time are likely to do better than projects with small budgets and short time horizons, a closer look at the underlying mechanisms is warranted. For instance, the correlation may be different for MDG projects, where results cannot be achieved in a short period of time, than for projects that focus on capacity development, where short training courses may make a measurable difference within a year. Of course, that size as measured by budget or duration is a proxy for, or is highly correlated with other project characteristics cannot be ruled out. However, a first look at project characteristics by project size does not reveal any clear patterns.

¹⁶¹ There is no budget information for the project CPDRR Uganda from the Ugandan Red Cross Society (URCS).

3.5 Synergies among the three components

The question of whether synergies exist between the MDG, capacity development and civil society strengthening, was put forth in sections 2.1 and 3.2.2. For instance, in section 2.1, on MDGs and themes, the synthesis team questioned the impact of projects that were mainly focused on MDGs, on good governance and improving the conditions in fragile states.

Conversely, section 3.2.2 presents questions on the rationale for evaluating changes in SPO capacities when there is no convincing evidence that these capacities contribute to development outcomes:

The most important question about the evaluation's three components is whether it makes sense to consider changes in capabilities separately from the SPO's impact on MDG or CS outcomes. In principle the question is whether MFS activities help the SPO to achieve its objectives. Whether that happened through changes in capabilities (because of training activities, for example), or through other means seems less relevant. Conversely, it is not clear how an organisation that becomes much more competent, thanks to MFS support, but that fails to translate this into MDG or CS outcomes should be evaluated.

The synergy question was not addressed extensively by the country teams, with the exception of India and Ethiopia. Both country reports produced correlation matrices between the five capabilities and the scores on MDGs.

The country team for India states: *"it is remarkable that pairwise correlation coefficients between capability to deliver on development objectives and each of the MDG factors are extremely low and non-significant"*. Nevertheless, they acknowledge that the number of observations (7) is very small.

The Ethiopian team finds *"Some correlation between the scores on MDGs and CD"*. For instance, there is a significant correlation between the capacity to Act and Commit and the Relevance of project results. There is also a correlation, albeit somewhat weaker, between the capacity to Achieve Coherence and implementation of the project as it was originally designed. The team acknowledges that the analysis is elementary and that more research is needed to better understand the underlying mechanisms driving these correlations.¹⁶²

Since 2011, MFS II has been the main programme through which the Dutch government aims to strengthen civil society in developing countries. Yet the Dutch Ministry of Foreign Affairs has developed a new grant framework to support civil society in developing countries, which runs parallel to implementation of the MFS II projects and their evaluation. This new framework focuses on lobbying and advocacy in a broader sense and emphasises the role of NGOs as society's watchdogs.¹⁶³ To strengthen and support this NGO role, the Ministry will form strategic partnerships with NGOs and alliances. Significantly, the Ministry has chosen not to fund activities towards the provision of basic services through this framework. The total amount of grants available will be considerably lower than MFS II.

¹⁶² The synthesis team has calculated correlations between MDG evaluation scores and 5C capability scores (both in levels and changes) for the entire available sample. Correlations in this case were low.

¹⁶³ See the policy framework *Samenspraak en tegenspraak – Strategische partnerschappen voor pleiten en beïnvloeden* (Ministerie van Buitenlandse Zaken, 2014).

In retrospect, the decision not to fund the provision of basic services through the MFS framework that focuses on civil society strengthening appears to be wise, or at least one that anticipates the apparent lack of synergy between MFS II capacity development interventions and MDG outcomes. However, it still begs the question of why focus on Capacity Building for SPOs if the link between their capacities and the delivery of outcomes on the ground is, at best, tenuous.

4 Reflection on evaluation process, lessons for future programme evaluation

Aiming for rigour in evaluations

This evaluation was unusual from its very start, both in its ambition and in its scope. Its ambition was to apply modern impact evaluation methods to a very large sample of MFS-supported activities. This was certainly unusual: activities of Dutch NGOs and their partner organisations had until then typically been evaluated by consultants using simple before-after comparisons and working closely with the NGOs. Such methodologies have increasingly come under attack in the last decade. Aid critics and academic researchers alike have pointed out that changes over time are poor indicators of impact: they may reflect factors completely unrelated to the aid-supported intervention which is the subject of the evaluation. The traditional way of showing that an NGO is successful, and thus deserving of MFS support, is slowly losing credibility.

At the same time there is increasing pressure to make evaluations as independent as possible: the organisation evaluated should not choose the evaluation team, should not attempt to influence its approach and should avoid even the appearance of any such attempt.

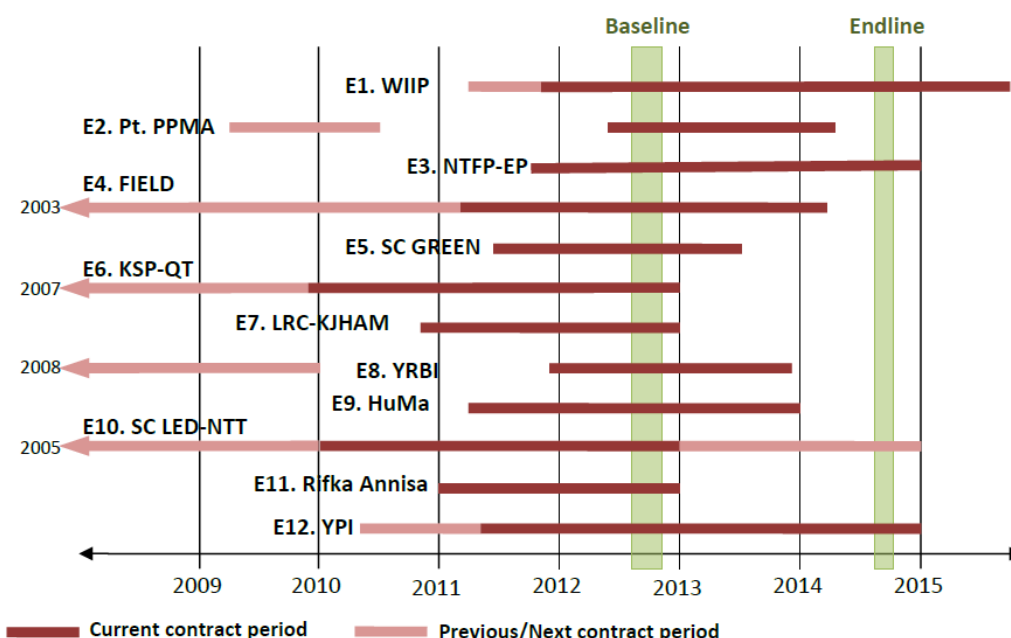
Modern evaluation methods are characterised by a rigorous attempt to construct a counterfactual. At one extreme, randomised controlled trials (RCTs) became popular in development economics as the preferred method, or as its proponents say, the gold standard, for assessing the impact of interventions. It is important to stress that RCTs are designed for a clearly defined activity or set of activities such as a conditional cash transfer programme aimed at raising school enrolment of children from poor households. Using RCTs also presupposes that it is feasible to have separate treatment and control groups and that the impact is evaluated only when the activity has run its full course. The MFS evaluation presented in this document faced serious limitations on all three scores.

The MFS evaluation sought to be as rigorous as possible. Therefore, the decision to ask research teams and not consultancy firms to conduct the evaluation was crucial in this endeavour, and helped to ensure independence. While it was recognised at the outset that in RCTs would not be feasible or even appropriate in many if not most of the contexts in which the evaluation took place, the standard that the evaluation teams should use the best possible methods was agreed upon, and maintained.

The most important threat to the quality of the evaluation came from the decision to use a common evaluation period. All activities in the sample were assessed twice: a baseline was set in 2012 and the endline in 2014. Changes, positive or negative, were therefore measured over the same two-year period. While this common period was obviously attractive from a bureaucratic point of view, on the ground it was not always the most appropriate periodization.

In numerous cases the baseline did not measure the situation prior to the activities under evaluation. On the contrary, the activities had often started before the baseline, either formally or in terms of substance, such as when a contract which started after the baseline covered a continuation of earlier activities. There were many examples where the baseline took place long after the project had started (See Figure 4.1).

Figure 4.1 - MFS II evaluation & contracting periods (Indonesia)¹⁶⁴



Dutch NGOs generally maintain long-term relationships with an SPO partner. These are formalised in a series of successive contracts. To consider the activities covered by one of these contracts separately is misleading, even if the contract only enters into force after the baseline period and is completed before the endline. The measured changes will partly reflect the earlier MFS relationship. Such complications are treated with care and clearly signalled in the evaluation report.

However, the common evaluation period makes comparisons across the various SPOs extremely difficult. This reduces the usefulness of the evaluation in an unnecessary manner. It would be strongly advisable to abandon the principle of a common evaluation period for future evaluations. Instead, the evaluation of programmes such as MFS should be set up as continual exercise. A selection could be made of activities which are about to begin, and this would ensure that a baseline is completed first. Endlines would be timed in such a way that the project runs its full course. A follow-up survey could be set for a later period, to determine whether the project's impact had been sustained over time.

This would make evaluation periods project-specific, doing justice to the enormous differences in project periods. While some activities, like ICT training, can be evaluated almost instantaneously, others (notably those seeking behavioural change) typically require years. Nothing is to be gained from waiting for two years in the former case, while a great deal can be lost as a result of attrition or other condition. In the latter case an evaluation after two years might miss a large part of the project's eventual impact, thus underestimating its success.

Clearly, if evaluations were organised in this way, a report on a small number of evaluations for recently completed projects could be made for any given year. Then, separate studies could be set up to draw lessons across projects within a particular country, or across countries for a particular type of

¹⁶⁴ This figure is taken from the Narrative country report Indonesia, p. 43.

activity. The Evaluation Department of the Ministry of Foreign Affairs (IOB) has been doing this for years. For example, it set up a series of evaluations of drinking water and sanitation programmes in developing countries supported by Dutch bilateral aid. Once this series was completed IOB commissioned a separate review to draw lessons learnt from these studies. The same approach should be adopted for MFS-type activities.

Continual evaluations would have another important advantage: it would be easier to guarantee the quality of the evaluation teams. For the current MFS evaluation NWO-WOTRO had to reissue its call for evaluation in some countries since the initial applications were judged to be of insufficient quality. The number of applicants is likely to be larger when calls go out at different times. Also, the cost of a new call is lower since evaluations do not have to start at the same time.

Accountability or learning? Implications for sample selection

The scope of the evaluation was also unusual: the MFS evaluation assessed the achievements of a very large number of organisations, in many different countries, pursuing a very wide range of objectives. This made the MFS evaluation a colossal and very complicated operation with an extremely heterogenous sample, across many dimensions.

The key question here is whether the evaluation is largely seen in terms of accountability or as a basis for learning. The choice is necessary since these two views have different implications for sample selection. A sample stratified only in term of the size of MFS support would be entirely appropriate if accountability is the key consideration. The IOB would then report each year on a random sample of recently completed aid-supported activities, controlling for size. If the evaluation scores were always positive, this would provide enough information.

The heterogeneity of the present sample is daunting. For example, in the case of Indonesia the projects in the sample are spread out over the enormous archipelago, and are also quite different in nature: there are projects for organic farming, reforestation, community empowerment, handicraft training, and sexual education, to name some. Comparing the results of the evaluations of such diverse activities within the same country does not necessarily lead to better understanding of the issues.

This is also true, although not to the same extent, for comparisons across countries. In the present sample there simply are not enough cases of, for example, handicraft training, to make a meaningful comparison about the effectiveness of different forms of handicraft training.

The lesson is not that the sample should be larger but rather that, if learning is an important objective then the sample design should reflect this. Comparability would then be enhanced by limiting the sample to a small number of particular types of projects. This would be very much in line with what IOB already practices in other areas.

Accountability pays the price in the current setup: the evaluation will not use a sample that is representative for MFS supported activities. However, that is also true for the present MFS evaluation if only because it covers a small number of countries.

It might make sense to use two separate designs in future evaluations. One would aim at accountability and would use a sample reflecting the full diversity of MFS-supported activities. This would achieve the learning objective through a series of evaluations for relatively homogeneous activities, like specific types of women empowerment or vocational training.

The main lesson here is Jan Tinbergen's fundamental point that the number of objectives should not exceed the number of instruments: the attempt to achieve different objectives with the same sample is bound to fail.

Sample size

The sample size had implications for the quality of the evaluation. In the MDG component the sample sizes for the large number of individual evaluations was basically dictated by the budget available for the evaluation in each country. This led to low statistical power in a number of cases. In those cases a conclusion of positive impact could not be drawn since the standard error was too large for the impact to be considered statistically significant.

When the evaluation was initiated it was hardly possible to estimate standard errors to calculate desirable sample size, since there so little work had been carried out in this area. The results of the current evaluation can be used in the future to estimate the standard error for a particular type of activity and country. For a given overall budget it may turn out to be wise to reduce the number of countries, or the number of projects within a country.

Organisation of the evaluation

The history of this evaluation is complicated and the specifics do not need to be reproduced in this report. Suffice it to say that the MFS organisations were expected to set up their own evaluation at an early stage and through an organisation in which they all participated. For various reason this effort was abandoned and MFS organisations asked NWO-WOTRO to play a central role: this body was to issue the call for evaluations, it would select and contract the evaluation teams, monitor the process and maintain contact with all stakeholders throughout the process. While NWO-WOTRO had little experience with modern evaluation research and many of these roles were new to them, they quickly took charge of the process and played their roles admirably.

They could not foresee, however, that governance of the evaluation would become incredibly complicated.¹⁶⁵ A large number of parties, including the Ministry, IOB, individual MFS organisations, the consortiums, the country teams, the synthesis team, the steering committee, the advisory board, and others, were involved and their roles and responsibilities were not clearly demarcated. The synthesis team made clear from the very beginning that it would only deal with the steering committee and NWO-WOTRO. This practice worked well but had to be defended at various stages. Also, in many of the country teams the team members responsible for the capacity development and civil society components did not

¹⁶⁵ This is illustrated in Figure 1.1.

fully accept the role of the team leaders. Indeed, at one stage they insisted on a direct line to the synthesis team. This was not accommodated and tension within the country teams resurfaced on various occasions.

The MFS evaluation suggests two lessons for future evaluation

First, an evaluation must not only be independent, it must also be perceived as being independent. Not only should there be no interference by stakeholders, but even the appearance of such interference must be avoided. This is already international standard practice, and has been for a long time. In The Netherlands it has not yet been internalised: the need for distance between evaluators and the organisation they evaluate is recognised in theory but not always in practice. This MFS evaluation probably reflects this partly in its early history when the MFS organisations felt ownership of the evaluation. While there was never a problem that could not be resolved, it would be wise to indicate at the very outset what is and is not appropriate in an independent evaluation for future reference.

Second, it would be wise to use a much simpler governance structure and to ensure that responsibilities are clearly defined in future evaluations. A serious design flaw in this case was that country teams were basically free to do as they liked once they had signed a contract; it should be noted, however, that they did not abuse this situation. It also became clear that while the synthesis team was expected to play a coordinating role, it had not been given the power to do so. Fortunately, this did not become a serious problem: the team leaders were in frequent contact to coordinate their work and the strong support of the steering committee allowed the synthesis team to play a directive role in spite of its lack of formal authority.

Methodology

The evaluation had three components: MDG outcomes, capacity development (CD) and civil society (CS). The baseline report lists some concerns about the proposed methodologies, notably for the CD and CS elements. One concern was that the scores given might not be comparable over time (between the baseline and the endline), and across activities. To a certain degree this is unavoidable: the scores are applied to specific characteristics of the SPOs that, by their very nature, do not lend themselves to measurement.

It is easier to avoid bias resulting from basing the evaluation largely or even exclusively on the views of stakeholders. Many evaluations relied very heavily on the views of SPO staff. It would be expected that this led to staff exaggerating positive changes. An indication of this is that the changes are often large in spite of high baseline scores and the short intervention period.

It was recognised that there was little scope for a formal counterfactual for the CD and CS components: what the SPO did was often unique. This ruled out formal testing of theories of change. However, the informal testing (process tracing) which was adopted as an alternative, could help bridge this gap in particular if it involved the views of knowledgeable external parties rather than only stakeholders, and if alternative explanations were systematically considered. The lesson for future

evaluations is the need for greater discipline. This is much easier to implement in a system of continual evaluation where a revision does not hold up other evaluations.)

The India study stands as a model for a convincing qualitative evaluation. It was particularly well done and shows the possibilities of good evaluation models. It is regrettable that this potential was not fully exploited in the other CD and CS evaluations. As a rule, staff was the main source of information and alternative explanations were not seriously considered, if at all in many cases. Also, for example in some parts of the Pakistan evaluation, the theory of change was not specified and no alternatives were provided.¹⁶⁶

The WOTRO call specified the use of the Civicus framework.¹⁶⁷ As many teams pointed out, this was not meant for evaluation. The Ministry of Foreign Affairs had required the use of Civicus for standard monitoring and evaluation (M&E). Yet, as stated in a number of country reports, Civicus had never been used for rigorous evaluation. The teams had to disaggregate the SPO impact on civil society in various ways including households, public sector, and private sector. It is difficult to see how this taxonomy can be applied in a meaningful way.

The most important question about the evaluation's three components is whether it makes sense to consider changes in capabilities separately from the SPO's impact on MDG or CS outcomes. In principle the question is whether MFS activities help the SPO to achieve its objectives. Whether that happened through changes in capabilities (because of training activities, for example), or through other means seems less relevant. Conversely, it is not clear how an organisation that becomes much more competent, thanks to MFS support, but that fails to translate this into MDG or CS outcomes should be evaluated. The evaluation suggests that there are many such cases. The India report, for example, indicates that the two capabilities selected for process tracing are irrelevant in terms of MDG scores. This raises the question why these capabilities should be analysed. If CD is seen as instrumental rather than as an aim in itself, the chosen setup with three components is not necessarily the most logical. It is worth noting that many SPOs think of success in terms of their impact, not in terms of their internal organisation.

A related observation is that the three components are likely to be affected by different biases. The CD and CS evaluations have no explicit counterfactual and may therefore overstate impact. In the MDG evaluations there are cases where the control group was affected by similar interventions. In that case the evaluation will only show an incremental effect and may understate impact since it will identify the effect of the SPO over and above that of rival organisations.

¹⁶⁶ The evaluation team was unable to convene a workshop with staff at Shirkat Gah. The fact that MFS funding had stopped at the end of 2012 is given as explanation). See p. 19 Follow-up report Shirkat Gah - Women's Resource Centre.

¹⁶⁷ See Annex A for a short explanation.

5 Conclusion

5.1 MDGs and themes

The evaluations of projects that focus on the MDGs and themes is discussed in section 2.1. The picture is generally positive: the average scores for Project Design, Implementation, Relevance and Efficiency are 7.5, 7.5, 7.4 and 6.8, respectively for the 53 projects evaluated. These scores clearly show a successful programme. The important score for Objectives Achieved is lower, 5.9 on average. A closer look at those scores shows some obvious explanations. First, the low scoring projects are concentrated in two countries: the Democratic Republic of Congo, and Ethiopia. This can be explained because it is difficult to accomplish results in countries that suffer from civil strife and violence. Nevertheless these findings are worth stressing since it is often argued that aid should be focused on fragile countries and countries affected by conflict. The evidence in this evaluation for the DRC and Ethiopia does not support the suggestion that aid can be effective in such environments.¹⁶⁸ In fact, it is quite remarkable that some projects in these countries any positive results at all. The same is true for Liberia, which was hit by an Ebola epidemic during the evaluation period.

There is little evidence that projects that aim to improve the capacities of SPOs make those SPOs more successful in addressing the MDGs that are their focus, as can be seen in projects where both CD and MDG objectives were evaluated. Furthermore, MDG projects that claim to address good governance or fragile states, have little to show in terms of results. In the former case, this may just be a case of overreach: capacity development takes time, and translating increased capacities into better MDG outcomes takes even more time. Obviously, it may be hard to detect results in a two-year evaluation period. Still, the lack of results raises the question why to focus on capacity building in the first place.

The lack of impact on good governance or fragile states by MDG-focused projects is less surprising. The relatively small scale activities of many of the projects in, for example, DRC or Ethiopia are no match for the governance and security problems that plague those states. It is probably better to simply not pretend that these, most often very relevant and successful projects, also contribute to larger goals.

5.2 Capacity development

Outcomes for the result area capacity development are more mixed. The results in India are generally positive and in some cases substantial. Furthermore, there is clear indication that these positive results are directly linked to MFS-supported efforts. The results in the DRC are limited, or even negative. Ethiopia scored high to begin with, and even showed substantial improvements over the two year evaluation period. New leadership in some of the projects appears to have played a major role. The results for Liberia are modest at best, and while considerable improvements are recorded for CD projects in Pakistan, the evaluation is rather superficial and does not give a clear explanation for these results.

¹⁶⁸ There is no consensus in the literature on the effectiveness of aid in these environments. For example, Paul Collier has argued that there is no role for aid during a civil war but that it can be very effective in the early post-conflict period.

It should be noted that these results may have been at least partly impacted by the evaluation methodology which leaves a lot of room for subjective scores, especially if they are largely based on the views of SPO staff members themselves. Fortunately, a number of country teams have shown that it is possible to go beyond the mere scoring of capabilities at baseline and follow-up. Where detailed process tracing was well done, and a clear theory of change was developed, the evaluations reached solid conclusions, including the ability to address the difficult question of attribution. Given the value of such carefully conducted evaluations, the synthesis team concludes that process tracing should become a standard in CD evaluations.

5.3 Civil society strengthening

To measure the impact of MFS sponsored interventions on civil society was arguably the most difficult part of this large impact evaluation exercise. First, the evaluation took place over two years, which in most cases is not long enough to find measurable change. Second, the changes that occurred are hard to quantify. The country teams had to rely to a large extent on subjective judgments, although in many cases these were complemented by more in-depth studies using process tracing or contribution analyses to attribute these changes to the MFS intervention.

The overall picture is rather positive, but there are significant differences between countries. For the DRC positive and negative findings offset each other, leading to the conclusion that the civil society arena has seen little change. One particularly negative finding regards a micro-credit product which lacked the financial capabilities to be successful, received insufficient support from the Dutch NGO and thus, not surprisingly, was unable to improve the societal position of the victims of armed conflict in the project area, which was its CS objective. This is a clear example of overambitious objectives and lack of focus resulting in a failed project.

In Ethiopia, the country team found a general improvement in the dimensions of Civic Engagement and Perception of Impact. This is quite remarkable because due to a 2009 law on civil society organisations the SPOs work in a very difficult environment. In India, the civil society picture is also generally positive, as Indian society is generally more conducive to SPO activities. Despite the shadow of civil strife in the past, Uganda shows very favourable results across the board.

In Indonesia, the overall scores are positive, which is rewarding. However a certain evaluation bias in favour of positive outcomes cannot be excluded. As is the case of India, the civil society climate in Indonesia generally supports SPO activities, which contributes to their success.

The Pakistan team found only small changes in the scores for the SPOs. With the exception of one case, the Pakistan report does not include convincing evidence that the changes observed can be attributed to MFS interventions.

5.4 Synergy, efficiency and project size

The synthesis team looked at a number of additional characteristics of the MFS sponsored projects. The question whether there is merit in trying to obtain multiple objectives within the same project will be analysed first. Generally this situation arises when a project aims to improve people's livelihood (MGD1), while simultaneously increasing the security situation in the project area (fragile states). Although this seems plausible because a more prosperous society may be less likely to turn to violence than one that houses large groups of people living in extreme poverty, in practice, the evidence for such synergy is very limited. In addition, capacity development projects that also aim to contribute to one or more of the MDGs, may be overreaching. Even when the capabilities of the SPOs that benefit from the CD projects improve, the SPO success in achieving MDGs does not seem to change. Of course, this observation could be explained by the relatively short evaluation period. It takes time to increase the capabilities of the SPOs, and it takes even more time to see significant improvements in the MDGs.

The question whether MFS sponsored projects are implemented in an efficient way appears to be very hard to answer. Before the start of the endline assessments, it was already decided that this issue would only be addressed for MDG projects since it was considered too difficult to disentangle budgets for CD and CS activities. For the MDGs, the synthesis team provided an inventory of benchmarks found in the literature. But even in this case, reliable benchmarks are very scarce. Furthermore, not all countries addressed this question, and those that did only succeeded in getting results for a limited number of projects. Thus, these results are mixed.

The Congo team found that SPOs seem unconcerned with efficiency, and noted that the analysis is severely hampered by the lack of financial information. For the two organisations where an efficiency analysis could be carried out, the team found one case where the costs for training activities was very high, but the cost for land rehabilitation was quite modest. Training in rice production was costly, and the training cost for rice processing seems excessive.

The Ethiopia team found an education project to be quite efficient and concluded the same for a teacher training and a language training project. On the other hand, a savings and credit project was much more expensive than the benchmark estimates provided by the synthesis team.

The India team gave very high scores for efficiency for seven of the nine projects they analysed.

The Indonesia team, which went out of its way to collect relevant cost and benchmark data, did find some projects that are implemented in an efficient way and some negative effects. There is no clear pattern.

The synthesis team concludes that the evaluation fails to shed much light on the efficiency question. There are two significant problems. MFS-sponsored projects are very diverse, even if only the group of MDG projects is considered. Benchmarks for all these projects are hard to come by, although this problem should become less severe over time. A more serious problem, and one that needs to be addressed urgently, is the lack of sufficiently detailed financial information, not only on the side of the SPOs but also, perhaps surprisingly, on the side of Dutch NGOs. Clearly, the efficiency question is highly relevant, and there appears to be ample room for improvements. The fact that neither the recipient organisations nor their Dutch counterparts routinely collect financial data that allows for efficiency

analysis, and that neither use benchmarks against which to measure their own performance is a matter of grave concern.

Finally the synthesis team looked at the question of whether size matters for project performance. Size is measured by either budget or project duration. A clear pattern can be seen: projects with either larger budgets or longer implementation time, or both, score significantly better on all measures: design, implementation, achievement, attribution, efficiency and relevance. At this point it is not clear what drives this positive correlation between size and performance.

5.5 The evaluation process

This evaluation was unusual and perhaps even unprecedented in ambition and scope. While rigorous impact evaluation has gradually become integrated into many development activities, especially in large multilateral donor organisations and major bi-lateral organisations, impact evaluation of Dutch-sponsored development projects lagged behind in these developments. Therefore, there can be no doubt that this large evaluation project, covering eight countries and just under 200 projects is a major step forward. Indeed, the eight country reports that were the base of this synthesis report, together with their almost 200 project-by-project technical appendices contain a wealth of information and important lessons. Summarising them in this report does not do justice to them. Many of those lessons are very country and even project specific. We hope that these reports will be widely distributed and that the major stakeholders will take their lessons to heart.

In this section we will focus on the evaluation process. The ambition guiding the MFS evaluation was to make it as rigorous as possible. Modern evaluation methods are characterised by their efforts to construct a counterfactual: what would have happened in the absence of the intervention? It is often argued that randomised controlled experiments (RCTs) are the golden standard for such an approach. The synthesis team does not take that position. RCTs, by their nature, are restricted to evaluating clearly defined, often one-dimensional activities implemented in a single environment. External validity of RCT results is a major problem. And in many cases there also are significant ethical issues that must be addressed, as in projects that provide health care or other important social services to the target population. Many MFS-sponsored projects have multiple objectives and are implemented in vastly different environments.

The synthesis team together with the country teams addressed the choice of evaluation techniques in a number of workshops prior to the roll out of the evaluation. This led to a common approach for each group of projects: MDGs, capacity development and civil society strengthening (see Section 5 in Annex A). While this was a step forward, it did not solve all problems resulting from the original design of the overall evaluation. The most important problem derived from the common evaluation period of two years. Country teams were expected to do a baseline survey in 2012 and a follow-up survey in 2014 for every project. This caused serious problems that reduced the strength of the evaluations.

First, in most cases, the evaluated projects had started before the baseline survey, in some cases ten or more years before. Thus the baseline survey, which was supposed to measure the *status quo*, i.e.

the situation before any intervention, was not a proper baseline survey. A project that started five years before could have been highly successful in its first two or three years, and maintained that success in the years thereafter, but the evaluation team would find no impact under this setup. Second, some of the projects had already ended (or no longer received MFS funding) before the follow-up survey. Impact measurement under those conditions is almost impossible. For future evaluations it is strongly advised to abandon the common evaluation approach. Instead, evaluations should be built into (major) projects from the start, and should be continuous, thus becoming an integrated part of projects. That evaluation becomes an intrinsic part of a project will also help to strengthen the design and clarify the ultimate objectives, and the path (the theory of change), that leads to those objectives.

A second major issue is whether the evaluations aim at improving accountability, or whether learning is the main objective. Each objective calls for a different evaluation approach. For instance, if learning is the objective, it is most relevant to draw lessons across countries about a specific type of project, such as improving access to health care for low income individuals. The evaluation department of the Ministry of Foreign Affairs has already adopted this approach. However, if accountability is the objective, sampling projects on, for example, budget size without concern for the project objectives, would be more appropriate. In the current case, both objectives appear to be aimed at simultaneously.

Another problem that the country teams encountered was that the sample sizes of their surveys were almost entirely determined by the available budget. A more focused approach would start with the objectives of the project, determine, or at least estimate, the ideal sample size to determine any impact, and provide a budget accordingly. It could be that in many cases where the country teams did not find any impact, sample sizes were just too small.

Probably because of the size of the current evaluation, and in part because of its history, the organisation of this project was unusually and unnecessarily complicated. In the future, the organisation of such an evaluation should be streamlined, with fewer layers of management and clearer lines of responsibility. In addition, the independence of the evaluation needs to be guaranteed.

Finally, the synthesis team had, and still has, some doubts about some of the methods used for the various evaluations. In particular, it was recognised early on that it would be difficult to establish suitable counterfactuals for the CD and CS components, thus making it very difficult to attribute the outcomes measured to specific project intervention. It was difficult but not impossible. Some of the country teams went out of their way to augment the standard approach in a number of ways, for instance by developing detailed theories of change, by process tracing, by searching for alternative explanations for the obtained results, by interviewing external experts (rather than just staff of the SPOs), et cetera. Much can be learnt from these approaches and we recommend that these approaches are included and further developed in future evaluations.

6 References

- AIID (2014). *MFS II Joint Evaluations – Literature Survey Efficiency: Unit cost benchmarks*.
- Beach, Derek and Rasmus Brun Pedersen (2013). *Process-tracing Methods – Foundations and guidelines*. Ann Arbor, Michigan: Michigan University Press.
- Deaton, Angus (2010). 'Instruments, Randomization, and Learning about Development', *Journal of Economic Literature*, vol. 28, pp. 424-455.
- Elbers, Chris and Jan Willem Gunning (2014), 'Evaluation of Development Programs: Randomized Controlled Trials or Regressions', *World Bank Economic Review*, vol. 28, pp. 432-445.
- Elbers, Chris and Jan Willem Gunning (2014a), 'What Do Development NGOs Achieve?', forthcoming in J.Y. Lin and C. Monga (eds.), *The Oxford Handbook of Africa and Economics*, Oxford: Oxford University Press.
- Gunning, Jan Willem (2012). 'Evaluating Development NGOs', Policy Brief nr. 56, Clermont-Ferrand: FERDI.
- Gunning, Jan Willem (2014). 'The brass standard? The scope for RCTs in impact evaluation', the 2014 Michael Lipton Lecture, University of Sussex. Retrieved on June 4, 2015 from <https://www.youtube.com/watch?v=7lb6Vfnz8ko>.
- Gunning, Jan Willem, Jacques van der Gaag and Gerton Rongen (2013). *MFS II Joint Evaluations - Synthesis Report on the Baseline Country Studies*. Retrieved on April 3, 2015 from <http://aiid.org/page.php?id=40&project=25>.
- Keijzer, Niels, Eunike Spierings, Geert Phlix and Alan Fowler (2011). *Bringing the invisible into perspective: Reference document for using the 5Cs framework to plan, monitor and evaluate capacity and results of capacity development processes*. Maastricht: ECDPM.
- Mati, Jacob M., Federico Silva and Tracy Anderson (2010). *Assessing and Strengthening Civil Society Worldwide: An updated programme description of the CIVICUS Civil Society Index: Phase 2008 to 2010*. CIVICUS, April. Retrieved on April 3, 2015 from <http://www.civicus.org/view/media/Assessing%20and%20Strenghtening%20CS%20Worldwide%202008-2010.pdf>
- Mayne, J. (2011). Contribution analysis: addressing cause and effect, in K. Forss, M. Marra and R. Schwartz (eds.) *Evaluating the Complex*. New Brunswick, NJ: Transaction Publishers.
- Ministerie van Buitenlandse Zaken (2009). *Subsidiebeidskader Medefinancieringsstelsel II 2011-2015*. Retrieved on January 14, 2013 from <http://www.rijksoverheid.nl/documenten-en-publicaties/richtlijnen/2010/01/01/medefinancieringsstelsel-ii-subsidiebeidskader.html>

Ministerie van Buitenlandse Zaken (2014). *Samenspraak en tegenspraak – Strategische partnerschappen voor pleiten en beïnvloeden*. Retrieved on April 9, 2015 from <http://www.rijksoverheid.nl/documenten-en-publicaties/regelingen/2014/05/13/samenspraak-en-tegenspraak.html>.

Palenberg, M. (2011). Tools and Methods for Evaluating the Efficiency of Development Interventions. Evaluation Working Papers. Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung.

Ravallion, Martin (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*, vol. 50, pp. 103-114.

Rodrik, Dani (2008). 'The New Development Economics: We Shall Experiment But How Shall We Learn?', John F. Kennedy School of Government, Harvard University, HKS Working Paper RWP 08-055.

White, Howard and Daniel Phillips (2012). *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework*. International Initiative for Impact Evaluation Working Paper 15.

Van der Meer, Frans-Bauke and Michiel Kort (2014). *Assessment of the Process of Joint MFS II Evaluations*. Unpublished report. Erasmus University Rotterdam.

7 Annex

A. The setup of the evaluation

1. Evaluation framework

A central element of the MFS II grant framework is the civil society result chain designed by the Ministry of Foreign Affairs. This tool defines priority result areas based on outcomes, are situated at four levels. The results of the MFS II programmes should be assessed in terms of these priority result areas. The four levels of outcomes are:

1. Attainment of Millennium Development Goals (MDGs) and themes;
2. Capacity development of Southern Partner Organisations (SPOs);
3. Civil society strengthening;
4. International lobbying and advocacy.¹⁶⁹

The first three levels are relevant to the country and synthesis teams; they form the different components of the evaluation and have separate, although closely related, sets of research questions. For the synthesis team, synergies between these components should be assessed too, to evaluate the assumptions behind the result chain of the Ministry. The outcomes on these three levels were subdivided in the Call for proposals; Table A.1 gives an overview of these subdivisions.

Table A.1 Overview of evaluation components and subdivisions

MDGs and themes	
MDG 1:	Poverty and hunger
MDG 2:	Education
MDG 3:	Gender
MDG 4,5,6:	Health
MDG 7a,b:	Safeguards for a sustainable living environment & forests and biodiversity
MDG 7c:	Drinking water and sanitation
Theme:	Good governance and civil society building
Theme:	Fragile states
Capacity development	
Capability to:	Act and commit
	Deliver on Development Objectives
	Adapt and Self-renew
	Relate to External Stakeholders
	Achieve Coherence
Civil society strengthening	
Dimensions:	Civic Engagement
	Level of Organisation
	Practice of Values
	Perception of impact
	Environment

¹⁶⁹ The priority result area International lobbying and advocacy falls outside the remit of the country and synthesis studies. It is the subject of the separate ILA evaluation.

The capabilities defined for the capacity development evaluation stem from the 5 capabilities (5C) model developed by the European Centre for Development Policy Management (ECDPM).¹⁷⁰ These form the basis for assessing the impact of MFS II on the capacity of SPOs. The component Civil society strengthening was also based on a pre-existing assessment tool, the Civil Society Index (CSI), developed by the CIVICUS World Alliance for Citizen Participation.¹⁷¹ The country teams had to adapt these methodologies to make them suitable for the purposes of this evaluation.

The Call for proposals stipulated a set of research questions to be answered in the evaluation. Table A.2 presents these questions.

Table A.2 Research questions

MDGs and themes	
1.	What are the changes under each MDG or theme at community/household level during the 2012-2014 period?
2.	To what degree are these changes at target group level attributable to the development interventions of Southern partners of the MFS II consortiums (i.e. measuring effectiveness)?
3.	What is the relevance of these changes?
4.	Were the development interventions of the Southern partner organisations efficient?
5.	What factors explain the findings drawn from the questions above?
Capacity development	
1.	What are the changes in partner organisations' capacity during the 2012-2014 period?
2.	To what degree are the changes identified in partner capacity attributable to development interventions undertaken by the MFS II consortiums (i.e. measuring effectiveness)?
3.	Were the efforts of the MFS II consortiums efficient?*
4.	What factors explain the findings drawn from the questions above?
Civil society strengthening	
1.	What are the changes in civil society in the 2012-2014 period, with particular focus on the relevant MDGs & themes in the selected country?
2.	To what degree are the changes identified attributable to the development interventions of the Southern partners of the MFS II consortiums (i.e. measuring effectiveness)?
3.	What is the relevance of these changes?
4.	Were the development interventions of the MFS II consortiums efficient?*
5.	What factors explain the findings drawn from the questions above?
* This question was dropped after the baseline assessments, see Section 2.5.	

2. The synthesis

The synthesis of MFS II evaluations centres on two different questions. First, the goal of the synthesis study is to assess the impact of the MFS II interventions as a whole by drawing on the results of the country teams. This requires summarising these studies and analysing the results obtained. Second, the synthesis study compares the methodologies employed by the different teams, assessing study design, data quality

¹⁷⁰ See Keijzer *et al.* (2011) for an overview and explanation of the 5C framework. The document can be accessed at www.ecdpm.org/5Cs.

¹⁷¹ See Mati, Silva and Anderson (2010) for a discussion of the Civil Society Index's methodology.

and other methodological issues. These two elements of the study complement each other because the aggregation of results takes into account the strength of the methodologies used to obtain them.

The country teams, in close collaboration with stakeholders, have attempted to describe the way in which the activities under evaluation led to the intended outcomes and how these outcomes were affected, for better or for worse, by other factors (confounding variables). These theories of change have played a central role in the qualitative or quantitative analysis: they have been tested against the collected evidence. It is not feasible to summarise all these theories of change. In some cases, the theories of change are defective in the sense that they do not provide a causal account. This was a major problem in the baseline reports, but was later rectified by most of the teams (see section 5.3 in Annex A). Typically, these are theories of great complexity. As an example, Figure A.1 shows the graphical representation of one such theory of change, one of the simpler ones. It is immediately obvious that such theories of change cannot be summarised.

One of the tasks of the synthesis teams was to coordinate between the individual country teams in order to ensure that the studies could be compared. Aggregation at the level of priority result areas requires results to be comparable. In order to facilitate this, the synthesis team produced reporting formats for both the baseline and follow-up reports.

Moreover, the synthesis team organised three workshops for the country teams. The first Joint Workshop was held on 15 and 16 May 2012 in Amsterdam and sought to streamline baseline assessments. Discussions took place on, among others, sampling for the civil society component, the revised timeline, uniform indicators, the report format and the scoring system for the studies. An important component of the workshop was to provide an opportunity for country teams and Co-Financing Agencies (CFAs) to discuss the evaluation.¹⁷² Country teams presented their revised evaluation methodology to the other teams and CFA evaluation and programme managers.

The second workshop, held on 17 and 18 June 2013, served to address the recommendations made in the baseline synthesis report. Decisions were made on how to approach the attribution problem for capacity development and civil society strengthening, as well as on how to evaluate the efficiency of the projects (see section 5 in Annex A for the outcomes of this discussion). Research challenges and suggested solutions were also presented to CFAs on June 18. In addition, teams requested a postponement of the follow-up assessment.

A third workshop was organised on 26 and 27 February 2014 to kick off the follow-up phase. Discussion took place on how to apply the previously agreed CD and CS methods, on timing of fieldwork and report deadlines and on a uniform reporting template. Moreover, a specific method on how to evaluate efficiency was presented and agreed upon.

The synthesis team took the lead in re-sampling projects and partners, as it was vital to do so in line with the original sampling criteria, and in the same way for all countries involved. Moreover, in the initial stages of the evaluation the synthesis teams has functioned as a first point of contact, introducing the country teams to the MFS consortiums contacts.

¹⁷² The consortia that were granted MFS II funds consist of two or more Co-Financing Agencies (CFAs).

3. Sampling

Sampling started with a selection of the countries that would be evaluated. Then, projects and SPOs were sampled for each of the three components of the evaluation: achievement of MDGs and themes, capacity development of SPOs and civil society strengthening. Sampling took place in a number of stages, carried out, in chronological order, by the *Koninklijk Instituut voor de Tropen*, the synthesis team and the country teams. The baseline synthesis report provides an overview of this process.¹⁷³ Some SPOs are present in multiple samples, reflecting the sample design. Table A.3 presents a break-down per component and country.

Table A.3 - Sample size by country and component at baseline

Country	MDGs & themes	Capacity development	Civil society
Bangladesh	4	5	16
DRC	5	5	19
Ethiopia	13	9	9
India	10	12	10
Indonesia	12	12	10
Liberia	4	5	12
Pakistan	3	4	8
Uganda	8	8	25
Total	59	60	109

¹⁷³ See Gunning, Van der Gaag and Rongen (2013), available at <http://aiid.org/page.php?id=73&publication=38>.

Since the baseline assessments, several developments led to a change in the sample. This synthesis report is based on the remaining projects and SPOs, as presented in Table A.4. The changes are the following:

MDG:

- Liberia: the PAMOJA project is a widespread rural project, implemented by the Dutch Consortium for Rehabilitation and local partners. Three of these, the SPOs DEN-L, LSGCE and FOHRD were sampled for evaluation. Since the organisations work on the same MDG project, there is only one evaluation, which, however, focuses on a number of different outcome indicators.
- Uganda: the project of the SPO Pozidep was not financed with MFS II funds in the end. It was dropped at a very late stage during the baseline phase.
- Bangladesh: results have not been included due to serious doubts about their reliability. There are errors in the indicated statistical significance of results. Also, the magnitude of many of the reported effects is difficult to believe and the country team has not provided a satisfactory explanation for these results.

CD:

- India: there is no endline assessment for two SPOs. The explanation provided is that the MFS II contract stopped during baseline for the SPO DRISTI, and that the grant for the SPO Women's Right Forum was stopped without the CFA being able to get in contact with them.
- The attribution question is only answered for a subsample.

CS:

- Liberia: the CS assessments were planned later than the MDG and CD evaluations. Due to the outbreak of Ebola, the team was unable to implement them.
- Ethiopia: the team was unable to do a follow-up assessment for RiPPLE. CFA WASTE and other members of the WASH alliance ended their partnership with the organisation due to unclear progress and financial reports, which made it impossible to link project activities to potential outcomes.¹⁷⁴
- India: the team was unable to do a follow-up assessment for the SPOs Prithvi Theatre and Cenderet. The relationship between CFA Hivos and Prithvi ended in October 2012; researchers were unable to interview Prithvi on the sustainability of results and achievements after the baseline.¹⁷⁵ Cenderet stopped implementing project activities after the baseline assessment, because funds transferred by CFA Cordaid did not arrive in its bank account. Efforts were made

¹⁷⁴ RiPPLE Endline report, p.16. The research team wrote a short report explaining the background to these developments.

¹⁷⁵ Prithvi Theatre Endline report, p. 14. A short report was written by the research team.

to get the project back on track, but the partnership between Cordaid and Cenderet was ended in December 2013.¹⁷⁶

- Pakistan: the SPO Roshni did not receive MFS II funds in the end. Therefore, no follow-up assessment was held.
- Uganda: initially, the team retained only 13 out of the 25 SPOs assessed at baseline, with the aim of carrying out process tracing for all 13 of them. However, the team did not do an endline assessment for three of them, due to the complexity of combining CD and CS evaluations (two SPOs) and the refusal of one SPO to collaborate because it had also been sampled for the MDG evaluation. For two of the remaining ten assessments, the collected data lacked the consistency necessary to be able to address the evaluation questions. This means evaluation reports were only produced for eight SPOs.
- The attribution question is only answered for a subsample.

Table A.4 - Sample size by country and component at endline

Country	MDG & Themes	Capacity Development	Civil Society
Bangladesh	.	5	16
DRC	5	5	19
Ethiopia	13	9	8
India	10	10	8
Indonesia	12	12	10
Liberia	2	5	.
Pakistan	3	4	7
Uganda	8	8	8
Total	53	58	76

4. Timeframe

The evaluation process started later than envisaged due to the late selection of three of the eight country teams. As a consequence, baseline assessments took place between May and December 2012. Baseline reports were submitted in January and February 2013. Interim visits and assessments took place in a small number of cases.

For some projects in the field of agriculture and education, the period in between preferred follow-up assessments and reporting deadlines was too short.¹⁷⁷ NWO-WOTRO therefore granted an extension of two months. Therefore, follow-up assessment took place between March and November 2014. Final country reports were submitted in April 2015, after a number of feedback rounds with SPOs and CFAs. This synthesis report is based on these reports.

¹⁷⁶ CENDERET Endline report, p. 13-5. A short report was written by the research team.

¹⁷⁷ Assessments of agricultural projects need to take place in about the same month of the year in order to minimise seasonal influences. For some educational projects, extension of the deadline meant that the evaluation could take into account an extra quarter of student grades.

In many cases, there was a mismatch between the timing of the evaluation and the duration of a project. This issue is revisited in chapter 4.

5. Agreed methodologies and choices country teams

5.1 Overview of employed methodologies

Table A.5 gives an overview of the methods used for evaluating the MDG projects. The method of choice, used 34 times, was a difference-in-difference design which required two measurements over time for both a treatment and a control group. Since most projects had already started before the baseline assessments, designs based on randomisation of the intervention were not applicable. Only the project in India was able to introduce random elements.

Table A.5 - Overview MDG methods per country

Country	Cross-sectional regressions	Diff-in-Diff	Focus groups	Propensity Score Matching	Randomised Control Trials (RCT)	Before-after comparison	Individual Interviews	Survey	Behavioural Experiment
DRC		5	5	3					3
ETH	1	13	2	8		1	2	3	
IND	5	4	1	5	1	1	5		
IDN	6	4	5			6	12	10	
LBR		2	1	2		1	2	2	1
PAK	1		2	3				3	
UGA	1	6		6		2	1	6	
TOTAL	14	34	16	27	1	11	22	24	4

In cases where it was not possible to construct a control group, teams relied on a before-after comparison for project beneficiaries only. In other cases, like India, the team relied on comparison of cross-sectional analyses. The difference-in-difference design was not considered appropriate in these cases. In 27 evaluations, the teams combined difference-in-difference or cross-sectional analysis with propensity score matching. Behavioural experiments were done for four projects.

Evaluation teams made use of surveys in which quantitative data was collected by administering a questionnaire at individual, household or community level, or at a combination of the three. In many cases, researchers combined this with qualitative data gathered through focus group discussions (16 times) or individual interviews (22 times reported, but probably applicable for all projects).

Table A.6 and Table A.7 provide an overview of the data collection methods used for the capacity development and civil society strengthening evaluations. Teams used the methodologies of process tracing or contribution analysis to make claims about attribution of results to MFS II interventions. These methods are discussed in section 5.3, below.

Table A.6 - Overview of CD methods per country

Country	Focus group discussions	Field Visits	Individual Interviews	Self-assessment	Workshop	Document analysis and triangulation	Observation
BGD	x		x	x	x		
DRC	x	x		x	x	x	x
ETH	x		x	x	x	x	x
IND	x		x	x	x	x	x
IDN	x		x	x	x	x	x
LBR	x		x	x	x	x	x
PAK	x		x		x	x	
UGA	x			x			x
Total	8	1	6	7	7	6	6

The central element in data collection is a workshop with SPO staff, and sometimes other stakeholders. All teams used this instrument, except for the Bangladesh CS researchers. Topics typically discussed during these workshops are the 5C model or Civil Society Index, an organisation’s theory of change, a historical timeline, and causal maps, contribution stories or impact grids. This information is then supplemented with individual interviews with SPO staff and external stakeholders, focus group discussions, self-assessments, observation and occasional field visits to beneficiaries. In addition, teams reviewed project and other documents. Triangulation of findings took place throughout this process.

Table A.7 - Overview CS methods per country

Country	Focus group discussions	Field trips/ Observation	Before-after comparison	Individual Interviews	Self-assessment/ self-report	Workshop	Document review and triangulation
BGD	x			x			x
DRC	x	x	x	x	x	x	x
ETH	x			x	x	x	x
IND	x			x	x	x	x
IDN	x			x	x	x	x
LBR
PAK	x		x	x		x	x
UGA				x		x	x
Total	6	1	2	7	4	6	7

5.2 Evaluating efficiency

The baseline country reports stressed the difficulty of obtaining financial data, notably for the capacity development and civil society strengthening components of MFS II interventions. Much of this information was absent from the baseline reports, in some cases due to a lack of cooperation from the CFA or SPO.¹⁷⁸ The synthesis team urged teams to redouble efforts to collect these data, and to establish a methodology for addressing these research questions, since obtaining cost data is only a first step in assessing efficiency.

This challenge was addressed during the June 2013 and February 2014 workshops organised by the synthesis team. In June 2013, teams emphasised once again that it was difficult to disentangle capacity development and civil society strengthening budgets: SPOs usually do not keep track of these activities separately, but include them in general project budgets. Therefore, it was agreed to focus on assessing the efficiency of the MDG projects only. NWO-WOTRO and the Ministry of Foreign Affairs did not object to narrowing down the scope of the evaluation questions on efficiency.

At the February 2014 workshop, teams agreed to focus on benchmarking unit costs of project outputs as a way of assessing the efficiency of the MDG projects.¹⁷⁹ This method provides a good balance between accuracy and resources required for the analysis. Nevertheless, it is only a partial analysis of project efficiency. The synthesis team facilitated assessments by conducting a literature survey of unit cost benchmarks for selected outcomes.¹⁸⁰ Section 3.3 discusses the results of the efficiency assessment.

5.3 Attribution of capacity development and civil society strengthening outcomes

A central question in the evaluations is whether observed changes in SPO capacity and at the civil society level can be attributed to MFS II-funded interventions. The methodology used in the baseline country reports was unable to answer this question convincingly. In response, the synthesis team made the following recommendations:

- To specify the organisations' theories of change in causal terms, and to develop clear hypotheses from these theories of change that can be tested against the evidence, if this had not already been done.
- To specify alternative theories of change, also in causal terms, both for the CS parts, where this was the intention anyway, and the CD parts.
- To indicate more clearly how alternative causal explanations are compared to answer the attribution question for the CD and CS parts.

The research teams have taken this challenge very seriously: improved methodologies were discussed and agreed upon during the two joint workshops. Teams in Ethiopia, India, Indonesia, Liberia and Pakistan opted to use process tracing, and contribution analysis was used by teams in Bangladesh and

¹⁷⁸ This was expressed by several teams.

¹⁷⁹ See Palenberg (2011) for a description of this and other methods to evaluate efficiency.

¹⁸⁰ AIID (2014).

the DRC).¹⁸¹ Both methods enable researchers to make causal claims about whether and how an outcome and intervention are connected. Process tracing, however, asks for the construction of alternative causal mechanisms more explicitly. In both methods, evidence is sought to confirm or reject the hypothesised mechanism at work. Evidence that provides both a necessary and sufficient explanation enables making strongest causal inferences.

Since process tracing and contribution analysis require more time than the methods originally proposed, it was too costly to implement them for the entire sample of SPOs. Therefore teams made a selection based on a number of criteria, among which the amount of MFS II funding for capacity development or civil society objectives, a focus on a certain MDG or theme, or ongoing MFS II support in 2014. This subsample was obviously selected before the start of fieldwork.

In addition, teams had to decide which outcomes to analyse by means of process tracing or contribution analysis, usually one or two per SPO. Per country, researchers identified important themes, such as strengthening intermediate organisations for CS, or a specific capability to focus on. Some teams selected outcomes before the actual fieldwork was carried out; others did so in the initial stages. An overview of all outcomes analysed is included in Annex C.¹⁸²

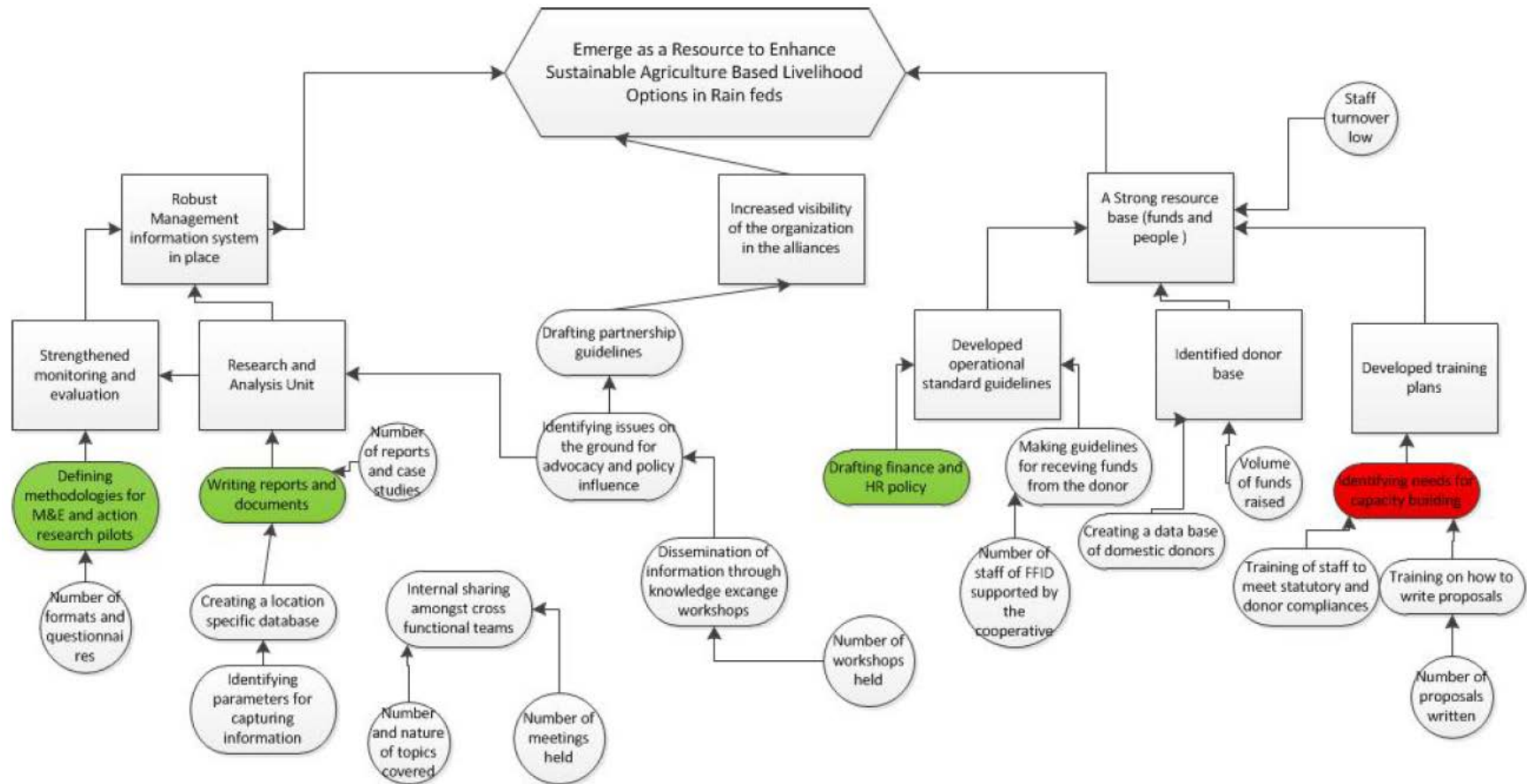
Teams have applied the agreed methodologies in different ways. Some teams are exemplary in using process tracing explicitly and systematically; others fail to clearly specify alternative pathways. This is discussed in more depth in the body of this report.

¹⁸¹ See White and Phillips (2012, pp. 40-4) for a short introduction to both methodologies. For a more extensive treatment, see Beach and Brun Pedersen (2013) for process tracing and Mayne (2011) for contribution analysis.

¹⁸² The overview will only be complete in the revised synthesis report.

Figure A.1 Example theory of change

Visual representation of the ToC on capacity development of FFID



Source: India baseline report, p. A 691.

B. Capacity development & civil society strengthening in Bangladesh

Capacity development

Table B.1 presents the 5C assessments of the five Bangladeshi SPOs in the sample. Overall, the organisations had well-developed capabilities at baseline, with scores ranging from 3.3 to 3.7. The country team observes that these are mostly mature organisations that have been working in development for a long time.¹⁸³ Generally, there were no explicitly agreed upon objectives for capacity development.¹⁸⁴ Organisations may not necessarily need MFS II funded capacity development support.

Nevertheless, 5C scores increased on average for all five capabilities, although the country team does not consider these increases to be significant. At the level of the individual SPO, some changes are worth noting, however. The Centre for Disability in Development (CDD) has seen improvements across the board, with a stronger relationship with the government and a larger funding base. These changes culminate in an increase of +1.2 on the capability to deliver on development objectives, illustrating that the other capabilities may stand in the service of the former. In addition, the researchers note that the strategic relationship between CDD and Light of the World, the CFA, led to capacity development at the CFA.¹⁸⁵

The SPO Sushasoner Jonny Procharavizan (SUPRO) registered rather large increases in the capabilities to Adapt and self-renew and to Relate to External Stakeholders (both +0.8). The team suggests that CFA Oxfam Novib has played a large role in bringing about these changes, since it has funded SUPRO since 2006, and contributed up to 95 percent of the SPO' budget. Specifically, SUPRO was introduced to the international tax justice coalition CRAFT.¹⁸⁶

Deteriorations in capacity were only observed for Practical Action Bangladesh (PAB), although the team contends these were not significant (in fact, they note that indicator descriptions showed improvements). They point out that the used methodology is weak when few of the SPO staff are involved in both rounds of assessments, in this case due to high staff turnover.¹⁸⁷

¹⁸³ Bangladesh Narrative Report, p. 10.

¹⁸⁴ Bangladesh Narrative Report, pp. 63-4.

¹⁸⁵ Bangladesh Technical Papers pp. 121-3.

¹⁸⁶ Bangladesh Technical Papers p. 190, Bangladesh Narrative Report, p. 62.

¹⁸⁷ Bangladesh Technical Papers p. 346.

Table B.1 - 5C Score Changes in Bangladesh

SPO	Baseline					Endline					Change				
	Act and commit Objectives	Adapt and self-renew	Relate external	Coherence		Act and commit Objectives	Adapt and self-renew	Relate external	Coherence		Act and commit Objectives	Adapt and self-renew	Relate external	Coherence	
AOSED	2.5	3	2.7	3.1	2.7	2.7	3.2	2.8	3.1	3.1	0.2	0.2	0.1	0	0.4
CDD	3.6	3.6	3.5	3.8	4	4.5	4.8	4	4.6	4.6	0.9	1.2	0.5	0.8	0.6
Practical Action Bangladesh	4.2	4.3	4.1	3.9	4.2	4.1	4	4	4	3.8	-0.1	-0.3	-0.1	0.1	-0.4
SUPRO	2.8	3.6	2.9	3.2	3.4	3.4	3.8	3.7	4	3.8	0.6	0.2	0.8	0.8	0.4
Village Education Resource Center	3.6	3.7	3.3	3.2	3.2	4.2	4	3.7	4	3.5	0.6	0.3	0.4	0.8	0.3
TOTAL	3.4	3.6	3.3	3.7	3.7	3.7	3.8	3.7	3.9	4.0	0.4	0.2	0.3	0.2	0.3

Civil society strengthening

Table B.2 gives the CSI scores for the 16 SPOs that were assessed by the Bangladesh team. The largest average changes are for the dimensions practice of values (+0.5) and environment (-0.6). Nevertheless, in the majority of cases, no significant changes were observed (for 46 of the 80 dimension scores, the change is 0).

The team reports a large deterioration (-2) on the environment dimension for four organisations: ADD, AOSED, DSK and VARD. SPOs have had to operate in a difficult political situation in the run up to elections, which was accompanied by violence. In order not to be associated with either the incumbent government or the opposition, organisations had to put up a difficult balancing act.¹⁸⁸

The SPO ASK noted improvements across the board (save environment). It has notably improved on practice of values (it is much more transparent in its code of conduct and human resource management) and Perception of Impact (it is perceived as being one of the few institutions defending human rights and the Bangladeshi constitution), although the team cautions against interpreting the latter improvement all too easily.¹⁸⁹ The SPO AOSED has seen much deterioration in its civil society arena, notably on level of organisation and Perception of Impact. It has experienced financial stress and organisational difficulties due to the withdrawal of MFS II funding, on which it was highly dependent.¹⁹⁰

¹⁸⁸ Bangladesh Narrative Report, p. 66.

¹⁸⁹ Bangladesh Technical Papers p. 246.

¹⁹⁰ Bangladesh Narrative Report, p. 62.

Table B.2 - CSI score changes in Bangladesh

SPO	Baseline					Endline					Change				
	Civic Engagement	Level of Organisation	Practice of Values	Perception of Impact	Environment	Civic Engagement	Level of Organisation	Practice of Values	Perception of Impact	Environment	Civic Engagement	Level of Organisation	Practice of Values	Perception of Impact	Environment
Association for Community Development	1	2	2	2	2	2	2	2	2	2	1	0	0	0	0
Action for Disability and Development	1	2	1	2	2	2	2	2	2	0	1	0	1	0	-2
An Organisation for Socio-Economic Development	1	1	1	2	2	1	0	2	1	0	0	-1	1	-1	-2
Aparajeyo Bangladesh	1	2	2	1	2	1	2	2	1	2	0	0	0	0	0
Ain O Salish Kendra	1	1	1	1	1	2	2	3	3	1	1	1	2	2	0
Bangladesh News Network	1	1	1	2	1	1	0	1	2	1	0	-1	0	0	0
Bangladesh National Women Lawyer's Association	1	1	2	2	2	2	1	3	3	1	1	0	1	1	-1
Caritas Bangladesh	1	1	2	2	1	2	1	2	2	2	1	0	0	0	1
Centre for Disability in Development	1	2	2	2	1	2	2	2	2	1	1	0	0	0	0
Christian Service Society	1	2	2	2	1	1	1	2	2	1	0	-1	0	0	0
Dushtha Shasthya Kendra	1	1	2	1	2	1	2	2	2	0	0	1	0	1	-2
Family Planning Association Bangladesh	2	2	1	2	2	1	2	2	2	1	-1	0	1	0	-1
Practical Action Bangladesh	1	2	2	1	2	2	2	2	2	1	1	0	0	1	-1
Sushasoner Jonny Procharavizan	2	1	2	2	2	2	1	2	2	2	0	0	0	0	0
Uttaran	2	2	1	2	2	2	2	2	2	2	0	0	1	0	0
Voluntary Association for Rural Development	2	2	1	2	2	2	2	2	2	0	0	0	1	0	-2
TOTAL	1.3	1.6	1.6	1.8	1.7	1.6	1.5	2.1	2.0	1.1	0.4	-0.1	0.5	0.3	-0.6

C. Civil society strengthening: outcomes (process tracing and contribution analysis)

Table C.1 - Civil Society Strengthening: Outcomes and Attribution (Process Tracing)

Country	SPO	Focus Area	Outcome evaluated	SPO Attribution	Link SPO - MFS II support
ETH	Ethiopian Kale Heywet Church	Enhancing civic engagement	Enhanced inclusion of orphans and vulnerable children into the school community through psychological and material support	Material support to the target group is a precondition to attain the outcome. Three socio-cultural interventions, among which MFS II-funded EKHC's peer education groups, and USAID-sponsored volunteers trained by EKHC contributed to the outcome.	The MFS II contribution seems rather limited. No material support is included in the MFS II project, whereas the USAID programme also provides some form of it. An advantage of the MFS II project, however, is that it is embedded in society.
		Strengthening intermediate organisations	Enhanced food and nutritional support to vulnerable groups	Several actors are identified, such as the forum of faith-based organisations and grain banks, both supported by EKHC. Other actors are the World Food Programme and women groups. All provide explanation for the outcome. However, EKHC's role is not convincing. For example: the nutritional value of food support given by grain banks needs to be reviewed.	EKHC's support to the faith-based organisation forum and the grain banks is financed through MFS II.
ETH	Catholic Church Gamo Gofa and South Omo	Strengthening intermediate organisations	The extent to which intermediate organisations improved food security of households in the Hamar district	No evidence was found that food security actually improved in the district. CCGG&SO has played a minimal role. The Saving and Internal Lending Communities and cooperatives did not have an effect. The effect of the health micro finance insurance schemes is unknown. Instead, income generating activities and the government's Productive Safety Net Programme may have improved food security for a very limited number of people.	The Saving and Internal Lending Communities, cooperatives and health microfinance insurance schemes were part of the MFS II funded project.

ETH	JECCDO	Strengthening intermediate organisations	Contribution of intermediate organisations to improved access to quality education for marginalised and disabled persons	Three factors are necessary but not sufficient individually to explain the outcome. These are: the availability of school buildings and adaptations for disabled children, the presence of teachers with the skills to create a child-centred learning environment, and a favourable community attitude towards the inclusion of marginalised and disabled children. JECCDO has taken action in all three areas, by training parent-teacher associations and education boards. Other NGOs and the local government have worked on these issues as well.	JECCDO work on all three areas has been financed with MFS II funds.
		Policy influencing	Increased government budget for education	The increase is mostly explained by federal government policy to make more funds available for education. The specific allocation of these funds at the regional level is influenced by experience-sharing visits, a regional education forum and the government-NGO network. Together these make up a causal package. JECCDO is a member of both. It plays a considerable role because it is a rather large NGO with a good reputation and linkages.	Networking and dialogue objectives are part of the MFS II funded programme.
ETH	Organisation for Social Service for AIDS (OSSA)	Enhancing civic engagement	The extent to which orphans, vulnerable children and people living with HIV are reintegrated into society and their social and economic capital improved	Access to effective ART treatment is a condition that needs to be satisfied. OSSA plays an important role by raising awareness and providing testing. Savings associations improve participants' social capital, but not their economic capital; participants lose money due to high inflation. The start-up capital distributed to beneficiaries to do business improves their livelihood, but success depends on personal entrepreneurial skills. OSSA organised the savings associations and provided start-up capital.	OSSA's activities are part of a project that is mainly MFS II funded.

		Strengthening intermediate organisations	Improved referral to health services through referral and care and support efforts	The idir coalition mobilised by OSSA does not explain the outcome. A number of other actors and factors do play a role, such as ART medication being free of charge and voluntary and mandatory testing.	
IND	CECOEDECON	Strengthening intermediate organisations	Five Rajasthan Block-level Development Committees (KSS) are able to influence public sector policies and practices. Specifically: Field trials for GM seeds stopped and decision reconfirmed; farmers now receive the minimum support price for four instead of two sacks of their harvest	A causal package: CECOEDECON and KSS actions are necessary, but by themselves not sufficient to explain outcome.	
IND	Centre for Sustainable Agriculture (CSA)	Enhancing civic engagement	Improvement in the engagement of women in CSA farmer cooperatives	CSA's role is a sufficient explanation for the improvement, but it is not a necessary element, multiple pathways exist. Certain external trends are necessary conditions for all pathways.	
		Strengthening intermediate organisations	Enhancement in the capacities of CSA's farmer cooperatives towards ensuring livelihood security of farmers	Given three favourable and necessary conditions, CSA activities provide sufficient explanation for attaining the outcome	
IND	Centre for Workers' Management (CWM)	Enhancing civic engagement	Two Chennai-based unions, GAFWU and PTS, have enhanced their membership, as well as the quality of women's engagement in trade unions	Three identified pathways, among which CWM support for unions since 2008, form a causal package; together they provide a minimally sufficient and necessary explanation.	Support had already started before the MFS II period.
		Strengthening intermediate organisations	The Garment and Textile Workers Union in Karnataka state has the capacity to influence policies and practices, in particular the minimum	A causal package which includes CWM's technical and negotiation support to GAFWU, offers a minimally sufficient and necessary explanation. CWM's efforts to support GAFWU internationally did not play a role.	Important technical support was a template developed by CWM to calculate a needs-based minimum wage. This was already done in 2007. Guidance and support by

			wages in the sector and the implementation of dearness allowance norms in Karnataka.		CWM have become more demand-driven under MFS II.
IND	NNET/Legal Cell for Human Rights	Enhancing civic engagement	Improved engagement of paralegal personnel so that individuals covered by the project access government schemes and programmes	A number of external conditions are necessary, but not sufficient, to ensure access. Given that these conditions are in place, NNET interventions are sufficient to explain the outcome. Interventions by other actors also provide sufficient explanation.	Training of the paralegal personnel is financed by MFS II funds.
IND	REDS-Tumkur	Enhancing civic engagement	Adijan people being better integrated and socially accepted in Tumkur district (it is questionable to what extent this outcome has actually been achieved)	Three pathways are identified, each of which is a sufficient but not necessary condition. These are REDS' support to Adijan people in the district since 1984, other factors include changes in the economic and social landscape, and REDS' solar lamp project for which Adijans are trained as technicians. The contribution of this last project does not seem to be substantial yet.	A more substantial contribution seems to come from interventions predating the MFS II period. These interventions only partially continued during MFS II. ICCO ended its financial support in 2012.
		Strengthening intermediate organisations	Adijan Panchayats are increasingly capable of claiming their rights, in particular with regards to reclaiming their land	Both REDS's efforts to create Adijan Panchayats and train them to reclaim land, and the favourable environment provided by the government are sufficient, but not necessary explanations for the outcome. REDS lobby and advocacy activities do not explain the outcome.	It is possible that REDS was more effective in achieving the outcome before the MFS II period. ICCO ended financial support in 2012.
IDN	Combine Resource Institution	Strengthening intermediate organisations	Increased participation of community radios and community journalists in the SK online platform	Both the creation of CRI's online platform and more people having access to the internet are necessary conditions, but by themselves not sufficient. Given that these are satisfied, capacity building efforts by CRI and other NGOs are sufficient to explain the outcome, but not necessary: it does not matter whether it is CRI or another party doing the capacity building.	CRI's contribution to changes in civil society is explained by Hivos' continued support to CRI's core programs since 2006. Hivos funds supported the development, operations and maintenance of the SK platform. In addition, there were interventions to strengthen the

		Influencing policies and practices	Ministry of Health is willing to validate national health insurance beneficiary data	Direct lobbying by various civil society elements is both sufficient and necessary to explain the outcome. However, enough evidence to suggest that CRI played a significant role in achieving this outcome is not available.	institutional management of SK member representatives.
IDN	WARSI - Komunitas Konservasi Indonesia	Strengthening intermediate organisations	Community-based forest management (CBFM) groups in 9 villages in 3 districts of 2 provinces have received full endorsement	Both the political will of government and WARSI efforts in guiding the CBFM proposals through bureaucracy are necessary to explain the outcome. Taken together, they also provide sufficient explanation: they are a causal package. Other NGOs have not been significant actors in this respect for these 9 villages.	There is a clear link to MFS II funding: this is one of the objectives of the MFS II-funded project.
		Influencing policies and practices	CBFM is mainstreamed into West Sumatra province forestry policy	The actions of three actors together explain the outcome, they constitute a causal package. These actors are provincial government, WARSI and another NGO.	WARSI's lobbying and advocacy efforts were supported by MFS II and REDD+
IDN	ELSAM Institute of Policy Research and Advocacy	Influencing policies and practices	Revised Law on Witness and Victim Protection	ELSAM's role was necessary but not sufficient. Based on its experience in working with victims, ELSAM provided analysis of regulatory gaps and drafted policy. With these technical inputs, the LPSK coalition of which ELSAM is a member successfully lobbied for revision of the law. Two other possible pathways were rejected.	ELSAM has benefitted from Hivos' long-term policy support and budgetary flexibility. However, agreed result indicators for this project were vague and difficult to measure, making it harder to attribute the observed changes to MFS II.
		Strengthening intermediate organisations	ELSAM's network organisations are more capable of data collection and analysis for use in local advocacy	The support that ELSAM provides is the most valid explanation. The data management framework it created is a necessary but not sufficient condition. Past support from other NGOs may have played a role too.	ELSAM has benefitted from Hivos' long-term policy support and budgetary flexibility.
IDN	LPPSLH	Level of Organisation	Small-scale crystal sugar producers are organised into three operational and profitable cooperatives	Product marketability and LPPSLH's approach to organise producers constitute a causal package - both are necessary and together they provide sufficient explanation. The approach taken by the government and actions by other civil society actors were not significant.	Hivos has supported LPPSLH since 2008. Outcomes can be attributed to MFS II support. The project did not have other donors.

		Strengthening intermediate organisations	Improved sugar producer position vis-à-vis middlemen	The new crystal sugar value chain provides necessary and sufficient explanation for the outcome. See the outcome above for LPPSLH's role in setting up this alternative value chain. The outcome cannot be achieved within the framework of the traditional value chain.	
IDN	Non-Timber Forest Products – Exchange Programme	Level of Organisation	Ensuring sustainable NTFP-based community livelihoods, in particular rattan, in Kutai Barat	Rattan products are also sold through other channels, each channel provides sufficient explanation for the outcome, but is not necessary.	Previous non-MFS II interventions also play an important role.
		Influencing policies and practices	Forest-dependent communities in Sintang are in a better position to claim their rights over forest-resources as a result of participatory mapping.	Explanations may differ per village. NTFP-EP's efforts alone are not enough to ensure government commitment NTFP-EP has built upon the past work of other civil society actors. Its lobbying activities have not yet resulted in local government including community maps in the local spatial plan.	
PAK	Bedari	Level of Organisation, Perception of Impact	Policies and legal frameworks which are more supportive of the rights of women and children, specifically strengthened and more influential civil society campaigns and networks.	Bedari initiated and has a leadership role in the Alliance Against Child Marriages (AACM), a forum that drafted the Child Marriage Prohibition Bill. Various “smoking gun” pieces of evidence are found of Bedari's learning, subsequent strategy, and its consequent growing role in making civil society networks influential, among which AACM. The growth of social media across Pakistan and a new modus operandi of the development sector in Pakistan (cluster approach) contributed to this.	There is a direct link between MFS II funding and the strengthening of Bedari's capacity, networks and alliances in South Punjab and Lahore. However, Bedari's partner organisations insist that the new move towards working from alliances has nothing to do with donor impetus.
		Level of Organisation, Perception of Impact	Policies and legal frameworks that are more supportive of the rights of women and children, specifically recognition of the rights of women and children by policymakers	Bedari's contribution passes “hoop tests”, which shows that its role may not have been sufficient to have fully affected policy but that there is enough evidence to suggest that it was a very necessary part of the process that led to this level of recognition and support by a critical mass of parliamentarians and the media.	MFS II funding made possible Bedari's expansion to Lahore, the provincial capital. This made it possible to work more effectively with the provincial bureaucracy and legislature.

PAK	Pakistan Fisherfolk Forum (PFF)	Policy impact	Abolishment of the contract system for fishing in Sindh and the implementation of a licencing system	The evidence of extensive capacity building at community level is neither necessary nor sufficient to prove contribution to a legislative change. The link between the intensified advocacy activity and the abolition of the contract system in January 2011 represents a hoop test, (necessary, but not sufficient to determine the outcome). Taken together, mass mobilisation, high media coverage, and effective advocacy with decision-makers, form a doubly-decisive test (necessary and sufficient) for affirming causal inference. Other factors: the impression is that the PPP-led government realised that the issue had become too sensitive to continue to postpone it. The provincial Fisheries Minister of the time, with whom PFF maintained good and frequent relations, played a crucial role in accelerating the adoption of the Fisheries Bill.	The Fisheries Bill was passed in 2011, so the MFS-II funded project did not initiate the process that led to the outcome, but contributed with considerable budget to PFF's advocacy objective. Although the outcome cannot be attributed solely to the project, MFS budget was considerable compared to the contribution of other donors/projects. MFS-budget related to civil society strengthening represented 39 percent in 2010-2011, 41in 2011- 2012 and 23in 2012-2013 of the total for this category of programme cost.
PAK	<i>Shirkat Gah</i> - Women's Resource Centre	Policy impact	Government is receptive to <i>Shirkat Gah's</i> input and critique to achieve women-friendly laws, policies and programmes	There is plenty of evidence that the increased strength and influence of CSOs like <i>Shirkat Gah</i> and various networks can be credited with the policy changes. SG has an impressive level of engagement with the state: an extremely significant piece of evidence is the invitation by the Sindh Government to educate a large group of legislators on domestic violence legislation. Partners repeatedly stressed that <i>Shirkat Gah</i> is a very proactive member of these networks, often takes on leadership roles, and represents the networks on government commissions and committees. An alternative explanation is that the political environment may have become more conducive to reforms over the last two years: this is refuted by a comparison between the provinces of Sindh and Punjab.	<i>Shirkat Gah's</i> partner organisations insist that the new move towards working through alliances has nothing to do with donor impetus. However, some obvious connections can be drawn between the MFS-II funded WESJP-Phase II programme and <i>Shirkat Gah's</i> growing profile in policy circles. The project ran from 2011-2014, with MFS II funding for the years 2011 and 2012. Other donors were also involved.

Table C.2 - Civil society strengthening: outcomes and attribution (contribution analysis)

Country	SPO	Contribution Focus	Outcome evaluated	SPO Attribution	Link SPO - MFS II support
BGD	Association for Community Development	Social impact	Reduced incidences of gender based violence against young women and girls	Results can be directly attributed to ACD but in order to be effective ACD relies on law enforcement and local government support. It is successful in engaging them. In addition, there are too many contexts where similar interventions do not take place and in which there is greater vulnerability for young women and girls (which serves as a counterfactual). Local government and law enforcement agencies make a medium-level contribution to the outcome. Their role is significant, but passive. Other NGOs also contribute, but ACD is seen as the regional lead NGO on this topic.	ACD is part of the Girl Power Project Bangladesh. The total value of the grant is approximately € 240,000. This is the largest grant held by ACD. MFSII provides 7 percent of the annual budget of ACD.
BGD	An Organisation for Socio-Economic Development	Strengthening intermediate organisations	Progress towards equal access to safe drinking water for all, specifically the creation of WASH Committees at the village level	One of AOSED's major accomplishments is the creation of WASH Committees at the village level, with village members who want to secure better access to drinking water and have received training on hygiene practices related to water use. Although AOSED has stopped providing support, committee members continue to collect small payments for maintenance of sand filters. Many other NGOs are active in this field, making a medium-level contribution to better access to safe drinking water.	AOSED was highly dependent on MFS II funding (80% of total budget) and is now experiencing severe financial and institutional stress because funding has been discontinued.
BGD	<i>Aparajeyo</i> Bangladesh	Social impact	Reduced gender based violence Reducing child marriages	<i>Aparajeyo</i> is successful at obtaining cooperation from relevant stakeholders. Moreover, there are too many contexts where similar interventions do not take place and in which there is greater vulnerability for young women and girls; which serves as a counterfactual. Three other actors (local committees, another NGO, mothers and other community members) make a medium-level contribution to the outcome. Of these, mothers are seen as a primary rival in explaining the outcome. The other two actors account for the outcome together with the actions of the SPO.	<i>Aparajeyo</i> is part of the Girl Power Project Bangladesh. MFSII finances 4.46% of <i>Aparajeyo's</i> annual budget.

BGD	Centre for Disability in Development	Social impact	Educating communities and the broader environment to make society more inclusive of people with disability whilst enhancing the capacity of people with a disability to participate in society	The team gives a score of 8 to results being attributable to the project interventions. Specifically, CDD is credited with increasing the number of members of the National Forum of Organizations Working with the Disabled. CDD also helped to create an enabling environment by changing the structural causes of exclusion of persons with disabilities. Other NGOs make a medium-level contribution to the outcome.	Since 2011 MFSII funds 8 percent of CDD's total budget. It is valued by senior management as continuous open-budget support, which can be strategically used for internal maturity and sustaining CDD's achievements. CDD and CFA Light for the World learn from each other, particularly at the organisational development and capacity building levels.
BGD	Sushasoner Jonny Procharavizan (SUPRO)	The research team writes: "The purpose of this section is not to assess the effectiveness of SUPRO's activities but to contextualise them within the broader institutional and political context in which SUPRO operates." If the effectiveness of the interventions has not been assessed, there is nothing to be attributed.			
COD	<i>Action pour le Développement Intégré au Kivu</i>	Agriculture	Setting up a functional cooperative, specifically the <i>Cooperative Agricole de Sange</i> that groups rice producers	ADI has clearly contributed to the establishment of the cooperative. More effort is needed for it to develop into a more mature cooperative. It is quite likely that the area would have been targeted by another NGO had ADI not done so first.	
COD	<i>Association des Femmes des Médias-Sud Kivu</i>	Women empowerment & gender	Strengthening the Noyaux Clubs d'Écoute (Listening Clubs), which allow members to approach local authorities and hold them accountable	Contribution of AFEM interventions to strengthening clubs is highly plausible. However, these are part of a larger rights-based discourse, and many other organisations work on this issue: changes cannot be wholly attributed to AFEM	AFEM interventions were funded by MFS II. However, AFEM has already been working with the Listening Clubs since 2006.
COD	Groupe Jérémie	Good governance Fragile States	The population is more knowledgeable about their rights. Better prison conditions and improvements in state policy and	It is plausible that GJ has contributed to better knowledge of rights, but many more actors engage in similar activities. It is certainly plausible that GJ has contributed to improvements in prison conditions: both state and non-state actors provide evidence	For knowledge-sharing activities, GJ works with grassroots groups; the source of funding is not always clear. GJ-induced changes in relation to policy and practice in prison conditions and governance can

			practice towards detainees.		mostly be attributed to the MFS-funded project. The project benefitted from pre-MFS II work by GJ.
COD	IFDP	Good governance Fragile States	Adoption of a new environmental law by the provincial assembly of South Kivu	IFDP has played and continues to play an important role in environmental protection legislation. The member of parliament who proposed the law acknowledges the role of IFDP throughout the whole process.	MFS funding can be seen as a contributing factor to both policy and practice of the environmental protection decree. Technical support by the CFA was also significant.
COD	<i>Réseau Haki na Amani</i>	Good governance Fragile States	Organisation and impact of community meetings (<i>barza</i>)	Collected evidence does not fully clarify the size, number and impact of the <i>barza</i> organised by RHA. Targets set in three-year plan were not met. There are concerns about the reduction of the number of participants in the <i>barza</i> .	RHA staff are greatly dissatisfied about cooperation with MFS donors. Delays in funding have an immediate effect and lead to interruption of activities.
COD	RFDP	Women empowerment & gender	Strengthening of the <i>Comités d'Alerte pour la Paix</i> (CAP)	It is plausible that RFDP strongly contributed to changes in terms of internal structure, capacity and level of independence of CAPs. Development of CAPs has in turn led to strengthening of civil society at the grassroots in Walungu. However, this cannot be attributed to the interventions by RFDP alone.	It is highly plausible that sustained financing by CFA has contributed to the outcome. MFS II funds are estimated to be two-thirds of the total CFA contribution.
COD	<i>Union Paysanne pour le Développement Integral</i>	Agriculture	Professionalising a cooperative, COOPABU. Functioning of the synergy network COS-PASAK (<i>COordination de la Synergie du Programme d'Appui à la Sécurité Alimentaire au Kivu</i>)	The impact of the cooperative was clearly visible: the cooperative was able to expand, new buildings were constructed, and more potatoes were produced and kept in storage. COOPABU received funds from another NGO; it is difficult to disentangle these contributions. Both NGOs provided material support and knowledge. UPDI has had a strong role in the realisations of COS-PASAK. The synergy contributes to more intense relations within civil society, but this is not the only initiative that brings together different parties. Also without COS-PASAK, UPDI would still be able to find entry points into the larger civil society and to collaborate with the aim of improving food security.	The synergy network was initiated by CFA with a view to Congolese partners gradually taking over. At present the synergy would have little chance of surviving without CFA funding.

COD	<i>Vredeseilanden Congo (VECO)</i>	Agriculture	Strengthening the cooperative COOSOPRODA	There have been important development in terms of infrastructure and production, group sales and increased revenues, but COOSOPRODA had already established a basic infrastructure through previous interventions. There was limited progress in terms of organisational capacity. Since no other interventions targeted COOSOPRODA during the evaluation period, it is plausible that developments are related to VECO's MFS II-funded project.	CFA provided institutional funding to VECO (salaries, office rent, and transportation). This was crucial for all VECO activities directed at the cooperatives, and thus contributed to strengthening civil society at the grassroots level.
COD	VICO - Villages Cobaye	Women empowerment & gender	Improvement of the societal position of victims of armed conflict in Walungu	It is unlikely that this project will achieve its planned outcome, because improvements in living circumstances were temporary and no direct evidence was found that beneficiaries' societal circumstances improved in a sustainable way. VICO lacked the necessary financial management skills to implement a microcredit project. The CFA knew about VICO's weaknesses but did not address them sufficiently	Many beneficiaries received a loan through the project, most failed to repay: it is likely that MFS II funding contributed to a temporary improvement in their living circumstances. No evidence was found that MFS support resulted in civil society strengthening outcomes beyond information exchange.
COD	CEPROF	Agriculture	Setting up and developing the <i>Mutuelles de Solidarité</i> (MUSOs)	It is plausible that the activities of CEPROF have contributed to the strong increase in number of MUSOs in the intervention area. However, development of the MUSOs cannot be only attributed to the intervention by CEPROF; there are previously-existing groups and other actors present in the area. It is too early to draw conclusions on the contribution of MUSOs to strengthening civil society.	Close involvement of CFA staff in the implementation of the project has contributed to the successful setting up of MUSOs. During the evaluation period, CFA was the only funding partner of CEPROF.
UGA	GWED-G	Strengthening intermediate organisations, Social impact	GWED–G contribution to the strengthening of youth groups for the empowerment for human rights	GWED-G has empowered the youth groups and has also built their capacity to be economically self-sustainable, as illustrated with several examples. The current leadership of the district is receptive to youth issues. GWED-G is a key partner of the district.	GWED-G is involved in three projects funded by MFS II through the Northern Uganda Human Rights Partnership.

UGA	Kabarole Research Centre (KRC)	Level of organisation	Establishment of the Rwenzori Regional Development Framework	KRC has had a significant role as initiator and incubator of the Framework. However, there has been little implementation of joint activities. Factors that negatively influenced the Framework are that financial institutions have not solved all problems related to access to credit for farmers, and that district level government cannot support the Framework's activities because these are not aligned with the country's National Development Plan.	In the MFS II period (2010-2014), KRC has received funding from Hivos on three projects.
UGA	Mango Tree	Strengthening intermediate organisations, Social impact	Development, Promotion and Propagation of the New Leblango Orthography and its use in Primary Education	The outcome can to a great extent be attributed to Mango Tree and its partner, the language board. It is a unique example of mother-tongue based education in Uganda. Other factors that contributed to the outcome are: increased attention by national policy makers and teacher training colleges, increased attention by district and municipal education officers, the interest of influential Lango individuals, and promotion by donors such as USAID.	Fifty percent of the funding for activities part of two projects is financed through MFS II funds.
UGA	SEATINI	Policy influencing	Tax justice campaign: reducing/removing tax incentives for multilateral [sic] corporations GMO campaign: Biotechnology and Biosafety Bill was referred back to committee	The two cases show that SEATINI has played a very important role in lobbying and advocacy campaigns in Uganda and acts as a lead organisation in coalitions of civil society organisations and other actors.	Oxfam Novib's contribution was 27 and 31 percent of SEATINI's 2012 and 2013 budgets respectively.
UGA	Send a Cow Uganda	Strengthening intermediate organisations, Social impact	Increased food security through community development	SACU's approach to community development has contributed significantly to the outcome. SACU's approach has received support from local leaders. The strategy is also supported by the government.	SACU projects are largely financed through MFS II funding.
UGA	Uganda Catholic Medical Bureau	Policy influencing	Development of the National Policy on Public-Private Partnership in Health	UCMB has played a pivotal role. Other faith-based organisations were unable to make substantial contributions. There is political will to involve the private sector; however, the policy was delayed between 2005 and 2010 because the Ministry of Health prioritised other issues.	MFS II funds were used to support UCMB's operational plan between 2012 and 2014.

		Social impact	ICT for Health Strengthening	The project addressed the need for accurate and timely information by establishing an information system and building staff ICT capacity. However, there were (and still are) many challenges during implementation.	It seems that the entire budget of this project is funded through MFS II.
UGA	Uganda Governance Monitoring Platform (UGMP)	Enhancing civic engagement	Contribution to the Citizens Manifesto and Black Monday Movement	UGMP has made a significant contribution to citizen empowerment and capacity building through these two outcomes. Since UGMP is a platform with very limited staff, it relies on its member and partner organisations for implementation. It is not clear whether all members contribute equally.	The programme is co-funded by five Dutch CFAs. The report does not make clear whether only MFS II funds are used.
UGA	<i>Vredeseilanden</i> Uganda	Strengthening intermediate organisations	Strengthening of Farmer Organisations in Eastern Uganda	VECO has played an important facilitating role. Sustainability of realised outcomes is currently not ensured.	MFS II funds totalled € 743,342 for three projects, over a period of 27 months (Oct 2012-2014).
		Policy influencing, Strengthening intermediate organisations	Elimination of counterfeit agro-inputs in Eastern Uganda	VECO played an important facilitating role in building the capacity of District Farmers' Associations (DFA) to advocate for laws that prohibit counterfeit agro-inputs. VECO also contributed to raising awareness of counterfeit inputs among farmers. However, the project ended in September 2013; DFAs have not been able to realise the necessary legal framework since.	