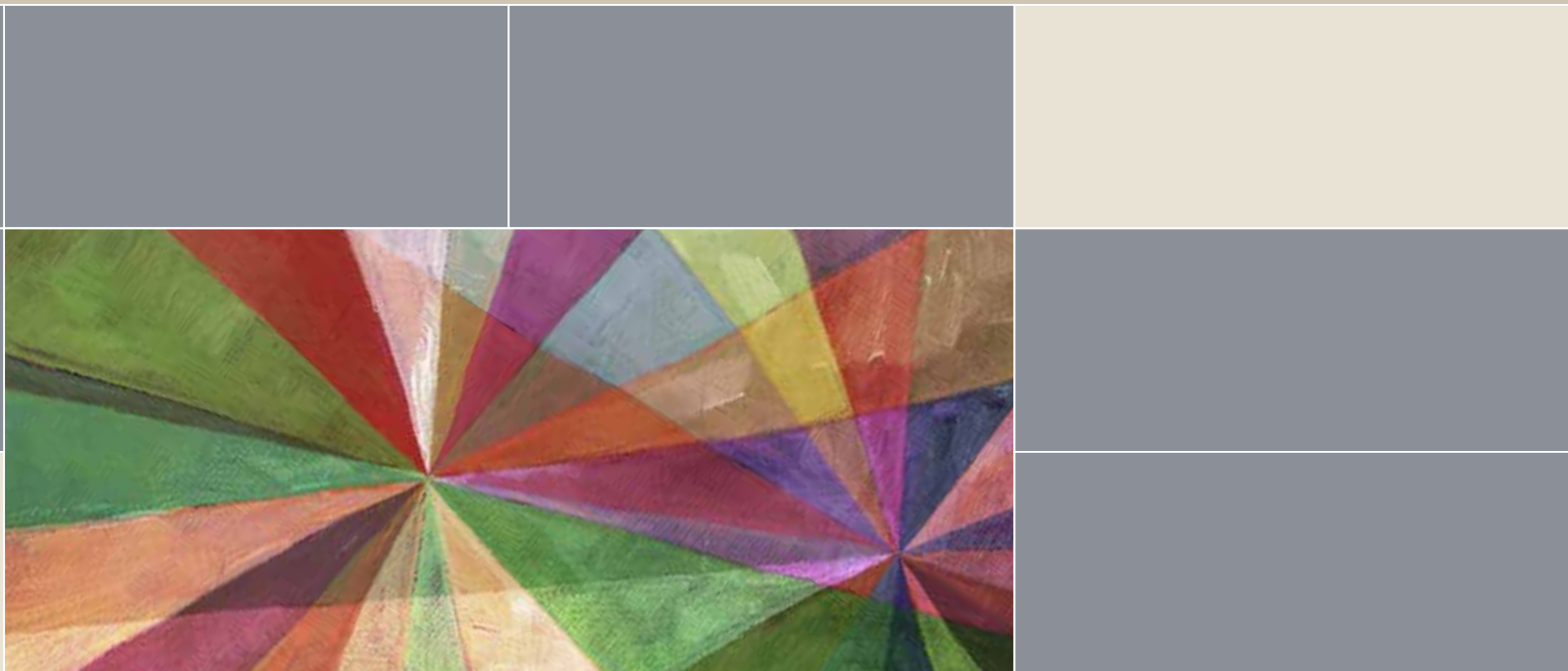




A Study of Monitoring and Evaluation in Six Norwegian Civil Society Organisations

Report 7/2012





Norad
Norwegian Agency for
Development Cooperation
P.O.Box 8034 Dep, NO-0030 Oslo
Ruseløkkveien 26, Oslo, Norway

Tel: +47 23 98 00 00
Fax: +47 23 98 00 99

Photo: istockphoto
Design: Siste Skrik Kommunikasjon
Print: Network Broking/Møklegaards Trykkeri
ISBN: 978-82-7548-682-8

A Study of Monitoring and Evaluation in Six Norwegian Civil Society Organisations

March 2013

Andante - tools for thinking AB

Core Team:

Kim **Forss** (team leader), Barbara **Befani** and Stein-Erik **Kruse**

Quality Assurors:

Roger **Riddell** and Nicoletta **Stame**

Disclaimer:

The report is the product of its authors, and responsibility for the accuracy of data included in this report rests with the authors. The findings, interpretations and conclusions presented do not necessarily reflect the views of Norad Evaluation Department.

Note on layout and language

The layout of the document has tried to conform to guidelines for accessibility and ease of reading, which require Arial font and left (not full) justification of the text.

The report has tried to avoid unnecessary use of acronyms and abbreviations.

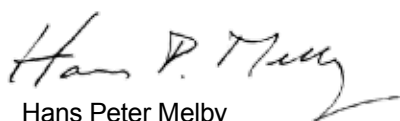
Preface

Official development assistance shall achieve results in countries far away from where its financiers – the taxpayers – live. The long way the money has to travel makes documentation of results more difficult. In developing countries, as well as in the north, however, people have a right to know the effects of aid.

Norwegian non-governmental organisations manage around a fifth of the Norwegian bilateral aid funds. Their results on the ground are important, as are their systems for knowing them. This study gives an overview and assessment of monitoring and evaluation in six of the largest Norwegian civil society organisations: Norwegian People's Aid, Norwegian Red Cross, Norwegian Church Aid, Save the Children Norway, Norwegian Refugee Council and the umbrella organisation Digni. The study presents findings across the organisations as well as differences between them.

The study was commissioned and managed by the Evaluation Department of the Norwegian Agency for Development Cooperation (Norad) and carried out by consultants lead by the consultancy company Andante – tools for thinking AB. The company is responsible for the content of the report, including the findings, conclusions and recommendations.

Oslo, March 2013



Hans Peter Melby
Acting Head, Evaluation Department

Acknowledgements

The study of Monitoring and Evaluation Systems in six Norwegian Civil Society Organisations has been a complex task, which has involved several persons on the evaluation team as well as stakeholders interacting with the process, and respondents to interviews and surveys. The evaluation was conducted by a team of 5 persons, organised in:

- Core Evaluation Team (Kim Forss (team leader), Barbara Befani and Stein-Erik Kruse), responsible for design of the evaluation and choice of methods, data collection, analysis, writing the final report, and presenting findings.
- Quality Assurance Team (Roger Riddell and Nicoletta Stame), providing advice on evaluation design, data collection instruments, and following the process of study and providing comments on the final report.

Collectively we in the evaluation team would like to thank all those we have interacted with for their assistance throughout the process of study. The evaluation has interacted with a Reference Group with members from each of the six organisations and we are grateful to its members for their assistance in providing information on the organisations they represented, and their help in establishing contacts for interviews and disseminating the survey. We would also like to express our gratitude to Norad for the support given to the evaluation.

Table of Contents

	Preface	iii
	Acknowledgements	v
	Table of Contents	vii
	List of Acronyms	ix
	Executive Summary	xi
1	Introduction	3
	Background	3
	Purpose	5
	Definitions and Methods	5
	Limitations of the Study	8
	A Guide to the Reader	8
2	Background Characteristics of Organisations	11
	Purpose and Objectives of the Organisations	11
	Size and Funding Sources	12
	Organisational Context	13
	Membership, Governance and Management	14
	National Territory and Country Presence	15
	Conclusion: The Influence of Context	17
3	Monitoring and Evaluation Systems	19
	Introduction	19
	Policy Guidance	19
	Extent of Formalised Policies	19
	Defining Monitoring and Evaluation	20
	Purposes of Monitoring and Evaluation	21
	Disseminating the Policy	22
	Organisational Structures of the M&E Systems	23
	Creating Evaluation Units	23
	Competence and Experience	24
	Independence	24
	Position of M&E Function and Reporting	26
	Budgets and Financial Information on M&E	27
	Institutional Memory	29
	Conclusion on Structures	29
	Process Characteristics of M&E Systems	31
	Decisions to Evaluate	31
	Decisions on Evaluation Parameters	33
	Developing Terms of Reference	34
	Management Response Systems	34

Quality Assurance	35
Competence Development	36
Professional Networks	36
The Role of Norad	37
Differences and Similarities	39
Conclusion	41
4 Analysis of Evaluation Cultures	43
Introduction	43
Support for Evaluations	45
Who Benefits from Evaluation	46
Beliefs About Evaluation	47
Use of Evaluation	51
Changing Evaluation Cultures	53
Conclusion	55
5 The Quality of Evaluation Reports	57
Introduction	57
A Quick Overview	59
Assessment of Utility	62
Assessment of Propriety	65
Assessment of Feasibility	67
Assessment of Accuracy	69
Conclusion	74
6 Evaluation; Use and Learning	76
Introduction	76
Use of Evaluations	77
Learning as Changes in Knowledge Structures	80
Incentives and Obstacles	81
Process Use of Monitoring and Evaluation	83
Conclusion	86
7 Conclusions and Recommendations	87
First Objective: Map the Evaluation Function in Selected Norwegian CSOs	87
Second Objective: Assessment	88
Third Objective: Recommendations	90
References	94
List of Annexes	
Annex 1. Terms of Reference	101
Annex 2. List of Persons Interviewed	107
Annex 3. Data Collection Instruments	109
Annex 4. Format for Quality Assessment	117
Annex 5. Evaluations Studied in the Quality Assessment	125
Annex 6. Review of Past Evaluations	127
Annex 7. Review of Context And M&E Systems; A Qualitative Comparative Analysis	134

List of Acronyms

ALNAP	Active Learning Network for Accountability and Performance in Humanitarian Action
Cida	Canadian International Development Agency
CSO	Civil Society Organisation
DAC	Development Assistance Committee (of the OECD)
Danida	Danish International Development Agency
IFRC	International Federation of the Red Cross
M&E	Monitoring and Evaluation
MFA	Ministry of Foreign Affairs
NCA	Norwegian Church Aid
NGO	Non Governmental Organisation
NOK	Norwegian crowns
NorCross	Norwegian Red Cross
NPA	Norwegian People's Aid
NRC	Norwegian Refugee Council
Norad	Norwegian Agency for Development Cooperation
OECD	Organisation for Economic Cooperation and Development
PME	Planning, Monitoring and Evaluation
PMER	Planning, Monitoring, Evaluation and Reporting
QCA	Comparative Qualitative Analysis
RBM	Results Based Management
SADEV	Swedish Agency for Development Evaluation
Sida	Swedish International Development Cooperation Agency
TOR	Term(s) of Reference
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Scientific and Cultural Organisation

Executive Summary



Executive Summary

Background

Civil Society Organisations (CSOs) have always been leading participants in development and humanitarian activities, in channelling large amounts of development cooperation money from both government funds and other sources into a range of projects and programmes. Like others, they are expected to show the results of their activities and for that purpose they have developed systems for monitoring and evaluation.

Purpose

This study provides an overview and assessment of the role of monitoring and evaluation in six of the largest Norwegian CSOs. It has the following objectives; (1) to map the evaluation function in this group of Norwegian CSOs; (2) to assess the evaluation function, most notably in regard to the relevance and quality of the reports produced and the systems in place for making use of the results and ensuring evaluations are also used for learning; and (3) to make recommendations which could strengthen the evaluation function to enhance its role in relation to quality assurance and results, and for knowledge sharing and learning, both internally and externally.

Methods

The report summarises the activities and conclusions drawn from different components of the study. These included interviews with a range of staff and management in the six selected CSOs (Norwegian People's Aid, Norwegian Red Cross, Norwegian Church Aid, Save the Children Norway, the Norwegian Refugee Council and Digni); a survey sent to hundreds of staff members in Norway and abroad; and a quality assessment analysis made from a sample of 30 evaluation reports. These three core elements of the study were undertaken against the backdrop of a review of relevant studies from a wider literature on evaluation and the evaluation function, giving particular prominence to the literature on the culture of evaluation and the relationship between evaluation, knowledge structures and learning processes.

Organisation of M&E systems

The study finds that **evaluation policies** are weakly developed even though all six organisations either have an explicit policy on evaluation or policy-like documents such as handbooks, or similar documents. The problem is that these documents tend to be rather vague and most do not clearly articulate strategic choices; they tend to be mainly descriptive and they omit important policy issues

such as how the agency should respond institutionally to the findings and recommendations of evaluations.

The structure of M&E is similar in the six organisations. Each has a distinct unit with responsibilities for M&E, usually 2 – 3 levels down in the organisational hierarchy. Their functions differ somewhat across the agencies examined but they tend to include systems development, overarching planning and help-desks for decentralised evaluations. Some units have extensive training functions for the organisation itself and its partner organisations. Monitoring and evaluation is structurally connected to planning and reporting. It is not much linked to governing Boards. There is little to no institutional memory and it is at times difficult to get an overview of the function – in particular the monitoring part of M&E.

The processes surrounding evaluation are relatively well established but tend to be informal and tailor-made to each evaluation. Evaluations are usually initiated from the bottom-up and the work to be undertaken then developed and consolidated into an overall evaluation plan. Budgets are drawn up separately for each evaluation, the amounts to be spent determined by estimates of what will be involved in each evaluation. Evaluation is invariably conducted by teams, the members of which are mostly contacted directly, but for larger evaluations there is a call for tender proposals. Reference groups are set up to monitor the process, and while only two of the six organisations have developed a formal management response system to respond to evaluation findings and recommendations, the deliberations of the reference groups cover much of the same ground as is achieved by a management response system, but without the transparency that the formal system generates.

Evaluation Cultures

“Evaluation culture” is the term used to describe the norms, values and attitudes that people in an organisation have to evaluation. The study found that these six organisations have similar evaluation cultures and that these cultures provide very strong support for the evaluation function. People have high expectations of evaluation processes and findings, and evaluation is associated with positive attributes (interesting, fun, and useful). Many within the agencies have both training and practical experiences of evaluation; yet they also want to learn more and to become even more involved. It is broadly recognized that evaluation requires a specific skill-set, which involves knowing about people and relations, as well as about measurement and results. While many take part of evaluations, their involvement is focused mostly on what is happening within their own organisation; there is little networking or interaction with the wider evaluation community or with other development cooperation agencies. While evaluation is seen as a specific and distinct activity with its own processes, it is also seen as a process that should be more closely integrated within and across other parts of the organisation. Independence is valued highly and there is a strong belief that evaluators need independence to do their job properly. At the same time, people want to be engaged in the process and to learn from evaluations; and they expect partners and implementing agencies to be similarly engaged. This is

an up-to-date understanding of evaluation practice, well aware of the complex nature of evaluations.

Quality of Evaluations

The study found that quality of the evaluations undertaken by these six CSOs does not differ markedly from what is known about the quality of aid evaluations undertaken by other development organisations, including official aid agencies. The evaluation reports produced, or more often commissioned by the agencies are assessed as good on substantive content, and they treat the evaluated projects and programmes with clarity and depth of understanding. However, our analysis suggests that the reports are weak in terms of the methodologies used: the ways they are constructed are often rather conventional and different options in respect of design are seldom considered; the methods of data collection and analysis are at times neither described nor are the strengths and weaknesses of the approach selected discussed against possible alternatives. Analysis of causal relationships is usually weak. In terms of coverage, while the analysis of results invariably focuses on relevance and effectiveness; impact, sustainability and efficiency are not covered as often. The analysis of implementation focuses mostly on planning, coordination and networks; coverage of financial management, governance and leadership is far weaker. The reports are well structured and readable, and they have clear recommendations that build on the analysis. The analysis of the 30 reports indicate considerable involvement of partner organisations: partners are involved in the design of the evaluation, the methods used tend to be participatory, and the evaluation teams often include participants drawn from both international as well as local settings.

Use and Learning

The study finds that monitoring and evaluation processes are used in part to generate learning. However, the extent to which they make sufficient use of the potential for learning is difficult to judge: it is very much a question of whether the glass is half full or half empty. The monitoring and evaluation systems in the six organisations generate strong instrumental use and process use, but far less conceptual use and enlightenment; additionally, the learning processes occur predominantly through assimilation of knowledge. Learning takes place in 'zones of comfort' and by way of gradual increases of knowledge. Still, evaluation use is not explicitly connected to other systems; in particular there are few incentives but also relatively few obstacles. Organisational learning is a huge subject and evaluation is affected by and plays a role in learning, but this study has not looked at 'the bigger picture'.

Conclusions

The study concludes that in many ways the monitoring and evaluation functions across the organisations perform well: there is a strongly supportive evaluation culture; many staff members are interested in evaluation and have practical experiences of monitoring and evaluation; and they have been exposed to training. The quality of evaluation reports is by and large satisfactory, even though (as with other agencies) there is certainly scope for improvement and for more extensive coverage. Policies, structures and functions have evolved over

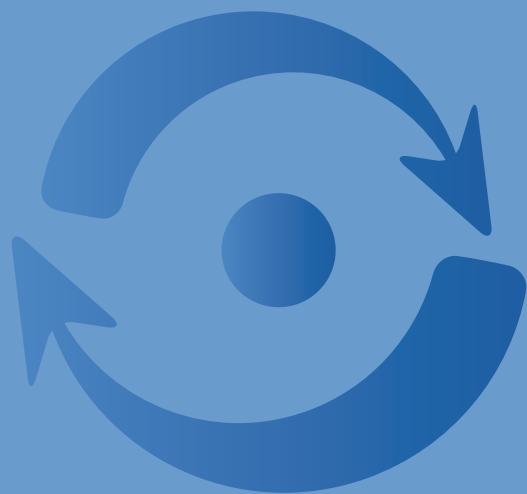
recent years and the organisations have all invested in systems development and in human resources for monitoring and evaluation, though many remain quite informal. There are of course significant differences between these six organisations in respect of the detailed design of structures and processes for monitoring and evaluation. But the similarities are more striking and the M&E functions of the six CSOs are quite different from the M&E functions in, for example, domestic public administration, the private sector, and in international agencies.

Recommendations

It is essential that all these positive developments are maintained and built-upon and we have found willingness for this to happen. Our recommendations focus on six areas of action:

1. Policy guidance that can be made more strategic, clearer and more comprehensive
2. Engaging the governing Boards, to increase the visibility and status of the monitoring and evaluation functions and to provide the Boards with results information.
3. Review of the functions of M&E units, which have developed organically, but may need to be reformed to make sure there are not roles and responsibilities around monitoring and evaluation that 'fall between the chairs'.
4. Establishing budgets and budget follow-up as financial information on the M&E functions is not available but will be necessary to assess also the value for money of this function.
5. Developing and standardising key processes, in particular management response, quality assurance, and the better anchoring of the evaluation process within the organisation.
6. Review of the incentives and obstacles to learning and use, which is an approach that must take a holistic perspective of the organisation.

Main Report



1. Introduction

Background

Civil Society Organisations (CSOs) are well-established actors on the international development arena. Most government donors channel official development assistance through CSOs, meaning that the organisations receive grants directly from the donor government. These amounts have increased considerably over the last decades, and so has the number of registered CSOs. Working through CSOs is often seen as modality of cooperation that offers opportunities different from state-to-state cooperation, and it is often seen to be a way of addressing poverty that has more immediately visible effects¹. But as the reform of development cooperation has gained momentum in the 21st century, there is a new critique of CSOs. With the impetus of the Paris Declaration of 2005, many have felt that the operations of CSOs should be critically analysed, bearing in mind the need for local ownership, alignment and harmonisation in the support given to civil society.

In 2010, Norwegian development assistance channelled through CSOs amounted to 3.6 billion NOK. 47% of this was channelled through the Ministry of Foreign Affairs and 7% through the Norwegian embassies. The rest, 46% or close to 1.7 billion NOK, was channelled through different budget lines managed by Norad. The single largest budget line - called "Support to Civil Society and Democratisation in the South" - provided slightly more than 1.2 billion NOK that was channelled through Norwegian CSOs to their partners in 74 countries for long-term development projects.²

These resources are provided through multi-year agreements with Norad. An important premise for such flexible long-term support is the organisations' ability to manage funds efficiently and effectively, and to deliver and document results. In order to meet such expectations, the organisations must have solid systems for quality assurance, and for monitoring and evaluation as components of an overall approach to results based management. Their systems must also be able to generate high quality products and processes that can be used to support internal organisational learning and external accountability.

There is a changing climate for civil society organisations in Norway. Over the last twenty years and in particular during the 1990s, the profile, number and budgets of CSOs expanded dramatically – based on the premise that the value

1 OECD (2011) *How DAC Members Work with Civil Society Organisations: An Overview*. Paris: OECD (October) <http://www.oecd.org/dataoecd/3/27/48843465.pdf>.

2 Norad, *Tracking Impact – An exploratory study of the wider effects of Norwegian CSOs*, 2011.

of their interventions was different from that of governments. These organisations were supposed to be people-centred and participatory in their approach, and their projects built on partnerships. Yet studies of CSO performance gave rise to increasing scepticism about their assumed comparative advantages. By the end of the twentieth century, robust and conclusive evidence was simply not available to confirm this proposition and the high expectations that the CSOs' unique contribution to development and poverty reduction was being realised in practice.

Evaluations and progress reports often show that short-term project objectives have been achieved with positive, but often scattered micro-results. Yet, these studies have repeatedly said little beyond the more immediate effects. In spite of a growth in the number of impact studies, little is still known about the long-term impact and the wider effects of CSO development interventions, though there is some knowledge about the results of Norwegian CSO projects from external studies and also from reviews and evaluations commissioned by the organisations themselves.

Underlying all discussions about impact - be it short-term and immediate, or long-term and wider effects - is the reliability of the evidence upon which conclusions are drawn. A central question is what do we know about the nature, quality, robustness and reliability of evaluations and the evaluation function in the Norwegian organisations? What do the systems for monitoring and evaluation, management and learning look like?

The literature draws attention to problems relating to monitoring, evaluation and knowledge of results. Efforts to assess accurately the impact of discrete projects have often been hindered by the cumulative effect of a number of common weaknesses. These include a lack of clarity concerning the precise objectives of projects and how they might best be assessed; poor or non-existent base-line data; inadequate monitoring and project completion reports; the low priority given to assessment and the related problems of inadequate in-house skills (Riddell et al., 1997).

These concerns persist. For example, one of the key conclusions of the recent Norad evaluation of NGOs in East Africa was that "most projects lacked the data and information required to be able to measure changes in indicators for key results accurately" (Ternström Consulting AB, 2011: xvii, 50-66 and 76-7). All too often, attention is focused on what evaluations tell us; far less attention is given to examining - no less to assessing - the quality of the methods used to draw these conclusions. "Evidence" ought to be treated with scepticism - arguably even discounted entirely - unless one is sure that the methods used are robust enough to guarantee that sound conclusions can be drawn.

A related problem is that assessments of the wider impact of CSO development activities require not merely information on inputs and outputs, but also information on outcomes and the relationship between these. The CSO sector is currently awash with debates and discussion about what methods to use and how appropriate and costly they are (see Woolcock and Karlan, 2009). Norad

has commissioned reviews of the organisations that have framework agreements, and these reviews have often also covered monitoring and evaluation systems, albeit with different focus. At the same time, many of the organisations have worked hard to improve their results-based management and this has included an increasing focus on monitoring and evaluation. Some of the criticisms and observations made especially in the older studies quoted above are probably out-dated now, as CSOs have introduced new and different methods and approaches as part of their wider work. Hence there's a need to focus more closely on what the monitoring and evaluation systems look like now and how they are changing.

Purpose

Against this background, the purpose of this study is to provide an overview and assessment of the role of the evaluation function in Norwegian CSOs, to assess the extent to which the knowledge and evidence produced is of high quality, and to analyse whether and how it is shared and leads to learning. The terms of reference for the study are reproduced as Annex 1. The study is also expected to provide insights into and lay out recommendations that will contribute to knowledge-oriented monitoring and evaluation system better suited to improve quality assurance of the organisation's work and results. The study has the following objectives:

- To map the evaluation function in selected Norwegian CSOs.
- To assess the evaluation function, as regards to the relevance and quality of reports and the system for transferring results, ensuring use of evaluations and learning.
- To provide recommendations to promote an evaluation function adequate for quality assurance of the organisations' work and results, for knowledge sharing and for learning, both internally and externally.

The Terms of Reference list the following organisations as the ones which are to be used as case studies: the Norwegian Refugee Council (NRC), the Norwegian Red Cross (NorCross), Norwegian Church Aid (NCA), Norwegian People's Aid (NPA), Save the Children Norway (SCN) and Digni. In total there are some 80 organisations that have framework agreements with Norad, and while these six are the largest, there are many others that are quite large too. As this study does not include any data from these other organisations, we are not able to judge the extent to which - or whether at all - the conclusions that we draw regarding these six organisations have relevance to the others. This is a study solely of these six organisations. It is not a wider study of the evaluation function across all Norwegian CSOs.

Definitions and Methods

Monitoring and evaluation are theoretical concepts and practical activities that relate closely to each other. We see them both as important aspects of evaluative information; much as the thinking on monitoring data rapidly involves evaluative inquiry, so does evaluation imply a use of monitoring data. It has also been suggested that in time streams of monitoring data may supplement the

traditional dominance of evaluation reports³. In the text we sometimes use the phrase “M&E systems”, and at other times “evaluation systems”, but when we use the latter we do so in the sense of a wider system of evaluative information, which includes monitoring data. We should also add that in general we use and apply the OECD/DAC definitions of the key terms commonly found in evaluation discourse including efficiency, effectiveness, impact, sustainability, and the more general evaluation-related terms, including the two overarching concepts, monitoring and evaluation⁴.

When we talk of **M&E systems**, we refer to the structures and processes across and through which these activities take place. Structures and processes are clear and well defined and amenable to analysis by the conventional tools of organisational research. However, there is a need to make clear the boundaries within which our analysis is conducted in particular in relation to other organisational processes such as planning and learning. While both planning and learning interact with evaluation and are of central relevance to this study, they are also much larger subjects: for the purposes of this study, we do examine them more narrowly, not treating them apart from their interface with monitoring and evaluation.

Our study also explicitly examines the cultural aspects of evaluation. By **‘evaluation culture’ we mean the values, norms and attitudes** that are associated with the evaluation function⁵. This is a relatively new area of investigation. However, we would argue that it is not merely an interesting but it is an increasingly necessary element of analysis in helping to understand the evaluation function, especially when analysing how and understanding why different organisational designs actually work. In particular comparative studies can often help to bring in the concept of culture in order to understand more fully the factors that lie behind the strengths and weaknesses of organisational designs.

The analysis builds on data from three different sources; interviews, document analysis, and a cross-agency survey. Their relevance for different parts of the study is shown in Box 1.1. In quantitative terms the study’s component parts comprised the following activities

- Interviews conducted with 60 persons from the CSOs and from Norad, using the interview question check-list reproduced in Annex 3.
- A survey sent to 284 respondents, resulting in a response rate of 73%, using the survey format reproduced in Annex 3.
- A quality assessment of 30 evaluations using the assessment format in Annex 4, and the sample of evaluation reports reproduced in Annex 5.

3 Rist, R. and Stame, N. (2006) *From Studies to Streams: Managing Evaluative Systems*. Transaction: New Brunswick.

4 OECD (2010) *Evaluating Development Co-operation: Summary of Key Norms and Standards*. Paris: Network on Development Evaluation, Development Assistance Committee, OECD <http://www.oecd.org/dataoecd/12/56/41612905.pdf>.

5 Organisational culture is defined as the norms, values and attitudes related to the organisation, see for example Hofstede (1980, p. 25 – 28) and more recently Barbier and Hawkins (2012, p. 3 – 13, and 181 – 185).

Triangulation in this study means that three or more theories, sources or types of information, or types of analysis are used to verify and substantiate the validity of an assessment made⁶. The purpose is to overcome bias by combining several data sources, methods and analyses. We were not able to gather three or more sources of information to verify each judgement made, but as the table shows, we did have at least three sources or types of information in respect of the systems used, the use to which evaluations are put and the quality of reports produced.

When we started the study we did not know how large the differences would turn out to be between the six organisations and we planned to try to trace the background to the differences through a Qualitative Comparative Analysis. It turns out that there are indeed differences, but not as systematic and widespread as we had expected.

Box 1.1 Triangulation of data sources

Phase of the study	Interviews	Document analysis	Survey
Analysis of M&E systems, structures and processes. Chapter 2	In-depth individual interviews with 6 – 10 respondents in each CSO. Focus group interview, and stakeholder seminar in 2 CSOs	Annual reports, Norad applications, policy documents, guidelines and handbooks on M&E, etc.	Questions on evaluation background, training and practical experience in part 1 and 3 of survey
Assessing evaluation culture. Chapter 3	Qualitative information on evaluation culture, extending the survey data	-	50 survey questions on values, norms and attitudes in respect to M&E
The quality of evaluation reports. Chapter 4	Opinions on quality	Detailed analysis of quality of 30 evaluation reports, 5 from each organisation	Opinions on quality
Use and learning. Chapter 5	Interview questions on use of reports, practical example followed through with a process description of use	Analysis of policies, rules and regulations, job descriptions that relate to use and learning	Questions in part 3 of survey

⁶ Cohen and Manion (2000) define triangulation as an “attempt to map out, or explain more fully, the richness and complexity of human behaviour by studying it from more than one standpoint. The definition above comes from Sida (2004, p. 114).

Limitations of the Study

The most significant limitations, threats to the reliability and validity of this study are:

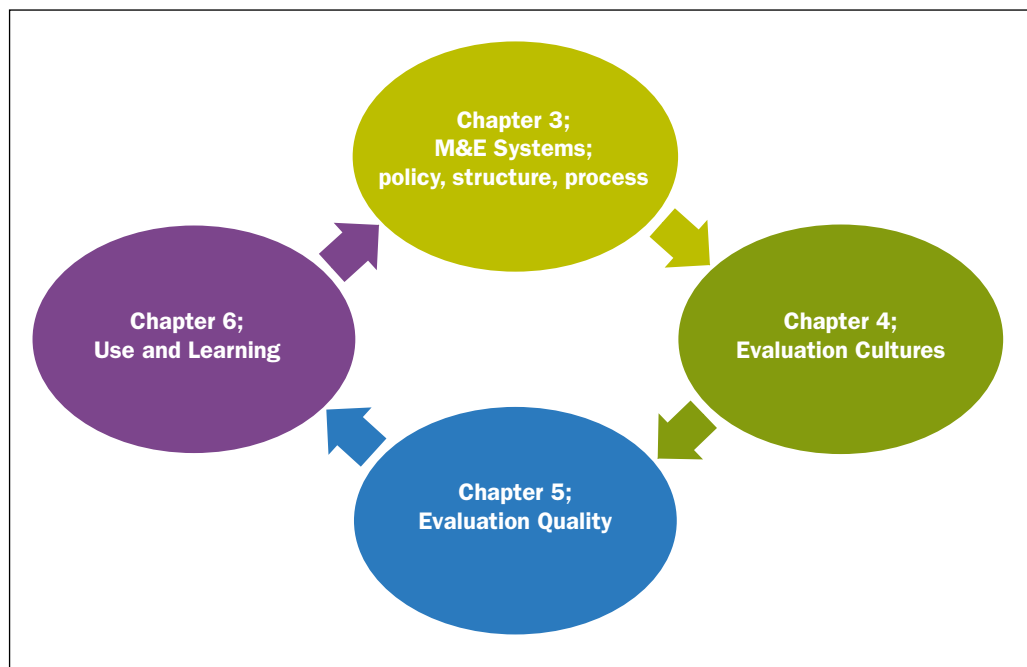
- Although called a study that focuses on “Norwegian CSOs”, in fact this study only makes use of information on six specific (and large) Norwegian CSOs. Thus the study’s findings should not be used to make generalisations beyond this group of agencies. We simply do not have the evidence to suggest that the study findings will necessarily apply to other Norwegian (or non-Norwegian) CSOs.
- The study discusses monitoring and evaluation, but it has an emphasis on evaluation. The reason for that is twofold;(1) the explicit purpose of the evaluation quoted above emphasizes evaluation, and (2) evaluation is more easily defined, visible and subject to analysis than is monitoring, which is a far wider, less precise, and dispersed activity.
- The role of the Ministry of Foreign Affairs in M&E of humanitarian aid has not been assessed in this report.
- The sample of interviewees and survey respondents is small and has a bias towards headquarters personnel. Although we have tried to compensate for this with phone interviews to field offices, they don’t represent more than 10% of the total.
- The sample of evaluation reports is biased towards reports commissioned from headquarters and to the most recent years, and could perhaps reflect an increasing concern for quality.
- Both monitoring and evaluation consist of formal and informal processes. Our approach and the information we have been able to gather tend to be more biased towards the formal structures and processes. We have tried to compensate for this through the survey data and also by questions asked during interviews, but the bias remains.
- The causal analysis is formalised but it may be based on too sharp a focus on recent events and pay insufficient attention to directions of change.
- It is difficult to assess how much influence Norad has had and what role the organisational reviews have played in shaping M and E systems and in influencing recent changes in approaches and attitudes to evaluation. Dialogue between Norad and the CSOs is clearly important, and we do have information on the interactive process, but the manner in which power is exercised and its importance is difficult to penetrate. There is probably more to the relationship than our account has been able to reveal.

A Guide to the Reader

The report is structured so that each of the main areas of study is treated in a chapter on its own, though there is a certain overlap between chapters. Thus, the analysis of structures and processes presented in chapter 3 contains elements of the analysis of use and learning found in chapter 6. Likewise, the analysis of quality, which is mainly discussed in chapter 5, has consequences for the process described in chapter 3, as well as for learning in chapter 6, and so on. The chapters are also structured in such a way that the chapters towards the end build more on the chapters in beginning, though it is not only a one-way relationship.

The report follows the overarching structure shown in Figure 1.1. However, before that chapter 2 presents the six organisations, pointing to differences between them in respect of funding, membership, governance and networks. This is a descriptive chapter that presents the organisations, and the reader who is familiar with them can go straight to chapter 3. Some of the chapters are more theoretical (chapter 6 in particular), and others more practically and empirically oriented (3 and 4). The subjects are closely interrelated though, and the overall purpose of the M&E systems described in chapters 3, 4 and 5 is to contribute to organisational learning and better development results through the use of monitoring and evaluation information.

Figure 1.1 The Narrative in this Report



The report goes beyond providing a factual account of what we found. We use theory and theoretical perspectives in all the chapters and make references to the relevant literature. However, we have deliberately tried to keep the theoretical elements as brief as possible, though the interested reader should be able to follow up on the analysis and discussion presented by making use of the sources cited in the footnotes and references. Most evaluations build on preceding evaluations and so does this one. Annex 6 contains a summary of what past evaluations have had to say on this subject. The causal analysis with the Qualitative Comparative Analysis (QCA) method connects a set of variables on culture, processes and structures to background information on the organisations. This is presented in Annex 7. The detailed questions in the TOR are treated within the structure shown in Figure 1.1, but to facilitate the reading for those who are looking for answers to a specific question, Table 1.1 points to where such answers can be located.

Table 1.1 Guide to where the questions in the Terms of Reference are discussed

Question (summary by the evaluation team, for the full question see Annex 3)	Section in the report
Question 1. Does the organisation have a results based programme management? Is there an evaluation strategy? Is there an evaluation policy?	Discussed in chapter 3, where structures and processes, as well as the presence/absence and contents of policies and strategies are reviewed.
Question 2. Does the organisation have a monitoring system? Does it cover internal and external needs for information?	Discussed in chapter 3, connected to information needs (contents of reports in chapter 5 and use of monitoring and evaluation in chapter 6.)
Question 3. How is M&E organized in the organisation? How are evaluations decided and budgeted? Who conducts the evaluations? What kind of system exists for follow-up?	Discussed throughout chapter 3. Overview in tables 3.4. and 3.7.
Question 4. What kinds of evaluations are conducted in the organisation? Are joint evaluations considered when relevant? Is analysis of outcome/impact included (what kind of method is used to identify the counterfactual	This question contains several specifications about the content of evaluations. Chapter 5 describes evaluation coverage, methods, methodologies and types of questions addressed.
Question 5. What is the quality of the reports? What criteria are covered? To what extent do they look at relevance, effectiveness, efficiency, impact and sustainability?	The questions concerning the qualities of the evaluation reports are answered in chapter 5, but to this should be added the systems put in place for quality assurance that are presented in chapter 3, and the perceptions of quality that are presented in chapter 4.
Question 6. What kind of system for learning has the organisation established? How are results from M&E reports transmitted? How is the follow up of reports? When in the process of evaluations are learning and use taking place?	The different aspects of use, for example process use, are discussed in chapter 6. Learning is discussed in the same chapter. The structures put in place are discussed in chapter 3, and the policies around use and learning in chapter 3.
Question 6 and 7. To what extent does the organisation find the established system and the contact with Norad appropriate and to what extent has the organisation followed up recommendations from Norad's evaluations?	These questions are addressed in the final section of chapter 3.

2. Background Characteristics of Organisations

Purpose and Objectives of the organisations

This chapter presents the six organisations that are analysed in the report. They are different and to some extent that may explain differences in M&E systems. That being said, this background context does not translate into systematically different approaches to monitoring and evaluation. The organisations have relatively similar goals and principles but also specialize in different areas of work and rely on different instruments and processes⁷. Their broad goals are⁸.

- Eradicating poverty and injustice (Norwegian Church Aid, NCA)
- Promoting and protecting the rights of people who have been forced to flee their countries, or their homes within their countries (NRC)
- Achieving a just distribution of power and resources (NPA)
- Protecting life, health, human worth and equal rights for all, irrespective of sex, disability, ethnicity, religion, age, sexual preference or social status (NPA)
- Protecting human life and rights in order to work towards worldwide peace (NorCross,)
- A world in which every child attains the right to survival, protection, development and participation (SCN⁹)
- A just world where God's creation is looked after, all human beings experience Dignity and no one suffers because of poverty (Digni).

In order to reach the above goals, the organisations focus on different specific objectives that they attempt to achieve through different activities:

- International assistance and relief for the wounded, elderly, sick and lonely: anyone in need, irrespective of race, religion or ideology, including the weakest groups in society
- Emergency preparedness and response
- Assistance, protection and durable solutions for refugees, internally displaced persons and returnees
- Long-term development aid and cooperation
- Advocacy
- Clearance of mines and explosives

⁷ The information on the broad goals, specific objectives and activities reported in this introduction to the chapter has been found on the organisations' official websites.

⁸ These are not meant to be inclusive of all the work of organisations but are just a sample.

⁹ SCN's vision statement in the Global Strategy 2010-2013.

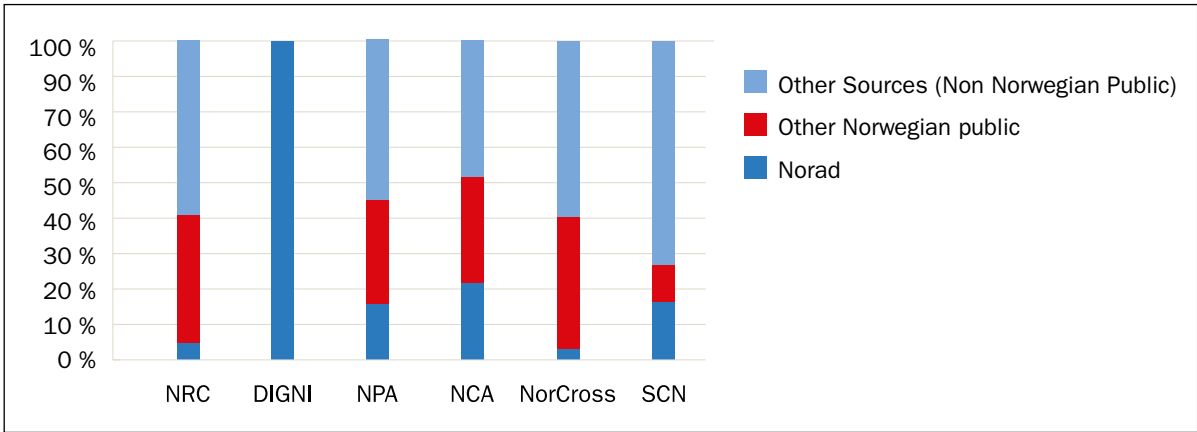
The organisations have different histories and as a consequence differ in several structural aspects, including size. In the rest of the chapter we will consider those characteristics that might presumably have an influence in shaping their evaluation systems.

Size and Funding Sources

For what concerns size¹⁰, organisations range from a total budget of as low as 150 MNOK (Digni) and 13 staff (Digni) to as high as 1.200 MNOK (NorCross and NRC) and 3.000 employees (NRC, NorCross has only 560 staff but it relies heavily on volunteers – 40.000 per year). Two organisations (NPA and NCA) draw from a similar budget amount (800 and 843 MNOK) despite one (NPA) having the double number of employees (2.400) when compared to the other (NCA, 1.127). Finally, SCN drew from a total budget of 557 MNOK in 2008, while employing a total of 910 people (note that the 2010 SCN total budget has increased to 654 MNOK).

In respect of funding sources, there is a first distinction between Norad, the Norwegian public sector and other sources. A further breakdown addresses the composition of non-Norwegian public funding, which includes international public, and private or corporate sources, in addition to resources directly earned by the organisations through activities like lotteries and the funding provided by mother or sister organisations, when the CSO is a branch of a wider global network.

Diagram 2.1 Composition of funding sources for the 6 CSOs



For one organisation (Digni), Norad is the only funding source while all five others also receive funding from the Norwegian Ministry of Foreign Affairs (MFA). Norwegian public sector funding is essential for all the organisations under study, making up around half or more of the total available budget in all organisations except one (SCN): 44% for NorCross, 45% for NRC, 50% for NPA, and 57% for NCA. Norad usually accounts for less than half of Norwegian public

¹⁰ The sources for the data are; NORAD (2008), *Performance Organisational Review of the Norwegian Missionary Council Office for Development Cooperation*, www.Digni.no. Figures presented at closest '000, here and throughout the chapter.

funding, except in Digni's case where it is 100%¹¹ and in the SCN case where it is 61%. Norad contributes 43% of total Norwegian public funding for NCA, 34% for NPA, and only 11% for NorCross and NRC. SCN has asked Norad to contribute around 30% or less to its total income in the current planning period. This used to be the total Norwegian public contribution to SCN in 2008, when Norad's contribution amounted to a mere 21%. In 2010, however, Norad's and Norwegian public contribution to SCN have decreased to respectively 18% and 29% of total income. Diagram 2.1 depicts the relative contributions from Norad and other Norwegian public sector bodies (essentially the MFA) to the six organisations¹².

Organisational context

Some organisations have very well established international networks that allow them to obtain funding from a variety of international donors: these donors are in some cases public institutions and organisations like the UN and UN branches, national development agencies (like Sida, USAID, Danida, DFID, etc.) and the EU, as in the case of NRC. In other cases they are mostly private or corporate donors, as those that make up almost half of SCN's budget.

NRC and NPA obtain 45% and 25% respectively of their funding from international public institutions¹³, which fund the other three organisations to a considerably lower extent. At the same time, SCN and NCA receive respectively 17% and 13% from international "mother" and "sister" organisations, including NGOs¹⁴. A very high amount of SCN funds (43%) come from private and corporate donors¹⁵, which contribute 19% of NorCross total budget and only 4% to NPA's (including Trade Unions)¹⁶. NPA also manages to collect 4% of its funding respectively from lotteries and 13% from the refugee reception centres and other income-generating operational activities¹⁷. Diagram 2.2 illustrates the composition of funding sources for the six organisations. In particular, unlike Diagram 2.1, it breaks down non-Norwegian public funding into 5 categories: as coming from international public institutions, mother/sister organisations, private or corporate bodies, own resources and other sources. Some of the organisations under study are part of much wider, global international networks (like the Red Cross/Red Crescent, Save The Children and to some extent NCA). Figure 2.1 illustrates the "superstructure" of these organisations. Some of Digni's members¹⁸ are part of larger international organisations, like for example the Salvation Army.

11 However, Digni's 19 members are not fully financed by Norad. The Digni secretariat is fully financed by Norad, the development work supported by the members are not.

12 Some data for NorCross are estimated.

13 NRC (2011), *Annual Report and NPA* (2010), *Annual Accounts*.

14 NCA (2010), *Approved Budget for International Cooperation and SC* (2009) *Norad Multi-Year Application for Organisations with a Core Funding Agreement, 2010-2014*; and <http://www.innsamlingskontrollen.no>.

15 Save The Children Norway (2009) *Norad Multi-Year Application for Organisations with a Core Funding Agreement, 2010-2014* and <http://www.innsamlingskontrollen.no>.

16 Norwegian People's Aid (2010), *Annual Accounts*.

17 *idem* (see previous footnote).

18 Digni's members have other funding sources, most have also sister-organisations with independent funding.

Diagram 2.2 Breakdown of funding sources for the 6 CSOs, including breakdown of non-Norwegian public funding

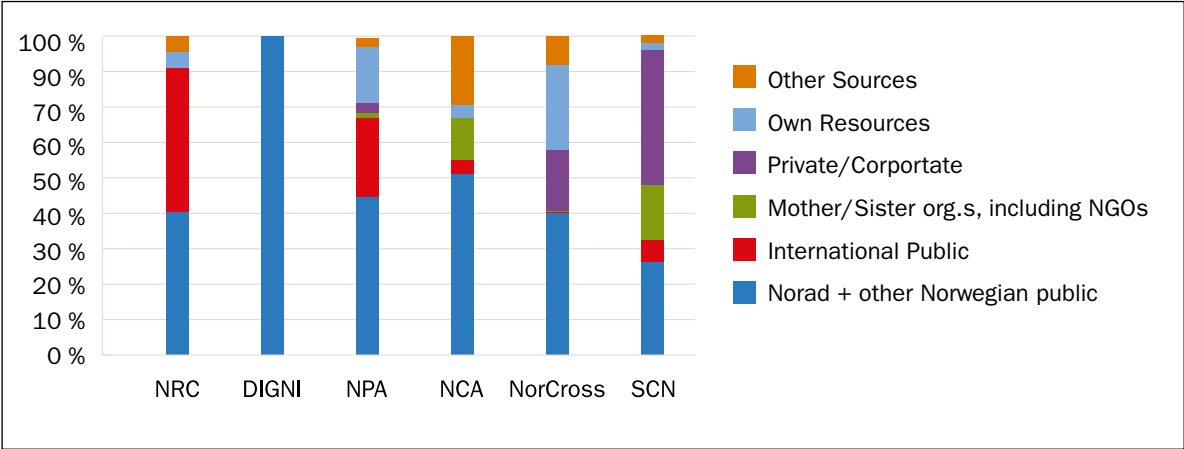
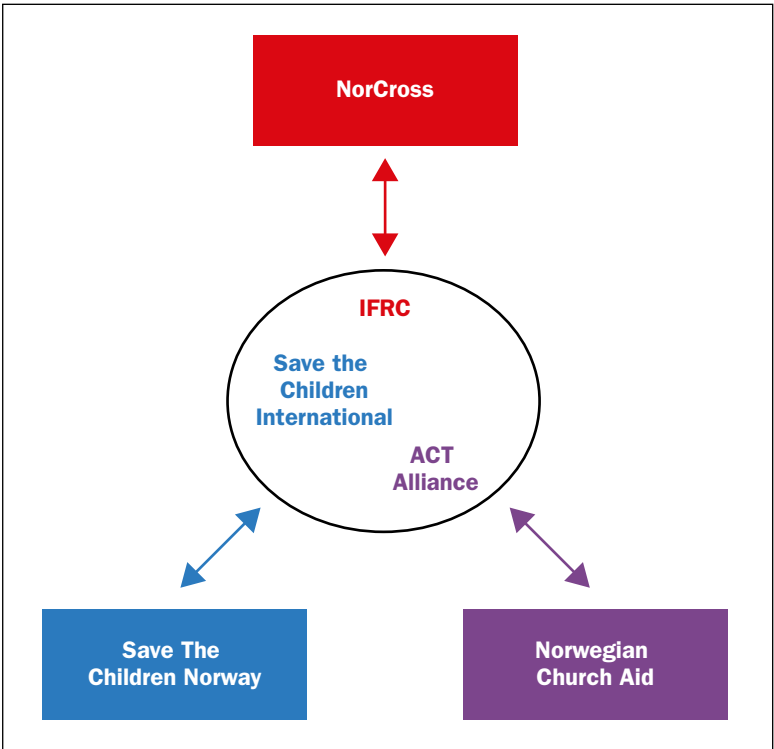


Figure 2.1 Superstructure of three CSOs with their international organisations



Membership, Governance and Management

Of the six organisations, only one is run as a private foundation (NRC): the other 5 are membership organisations. Membership can be individual or institutional. NorCross for example is essentially an individual membership organisation, with 133.000 to 170.000¹⁹ members in Norway. SCN is also grounded on individual membership, with 8.000 registered members and 100.000 regular supporters on

¹⁹ At the time of writing, www.rodekors.no reports both figures.

the national territory²⁰. NPA has over 10.000 members in Norway but also has important institutional members (the Trade Unions and labour organisations). Conversely, the faith-based organisations are founded solely on institutional membership: NCA (although legally registered as a private foundation) is owned by the churches and parishes of the Norwegian Church and other national Christian organisations (the Salvation Army, the Baptist Union, the Methodist Church, etc.). Similarly, Digni is an umbrella organisation representing 19 faith-based (Christian) associations supporting some of the same organisations as NCA.

The nature of the membership has implications for governance and management: while all have a managing board and staff, only the membership organisations (except Digni) are governed by a General Assembly of members, usually elected by local groups to represent local branches or activity groups at the General Assembly. The latter elects a smaller board that meets more often (usually called the Executive Board) and constitutes a middle link in the delegation chain between management (responsible for day-to-day decisions) and the Assembly, which owns the organisation.

The sizes of the governing bodies are not too dissimilar for the membership organisations: NPA's executive board is composed of 9 members plus representatives of the labour organisations; SCN's executive board has 10 members and is elected by another intermediate body of 15 members with advisory functions; NCA's executive board, elected by a 39-member Supervisory Board representing the Diocesan councils and other Christian organisations, has 9 members (including two members of staff); NorCross has a 26-member National Board plus two Advisory bodies of 19 district leaders and 6 regional leaders. NRC has no membership and self-nominates its board²¹; finally, Digni is governed by a General Assembly (consisting of one representative from each member organisation), which decides Digni's overall strategy and the statutes, plus an executive board of 9 members who are elected by the Assembly.

National territory and country presence

The membership typology is also related to the "downstream" aspects of the 6 organisations, particularly their local and regional presence in Norway. This section will also address their presence on the national territory and globally, in developing countries; but also to some extent other aspects like the presence of youth organisations, emergency response teams, and political or religious affiliation.

The only non-membership CSO (NRC) also has zero local presence in Norway while those based on individual membership have several district and local activity groups: NPA has 136 local branches, including 6 regional offices and 12 refugee reception centres; NorCross has 19 district offices and 400 local branches in Norway; SCN has 5 local offices and 84 local branches and activity

20 Save The Children Norway (2009) *Norad Multi-Year Application for Organisations with a Core Funding Agreement, 2010-2014*.

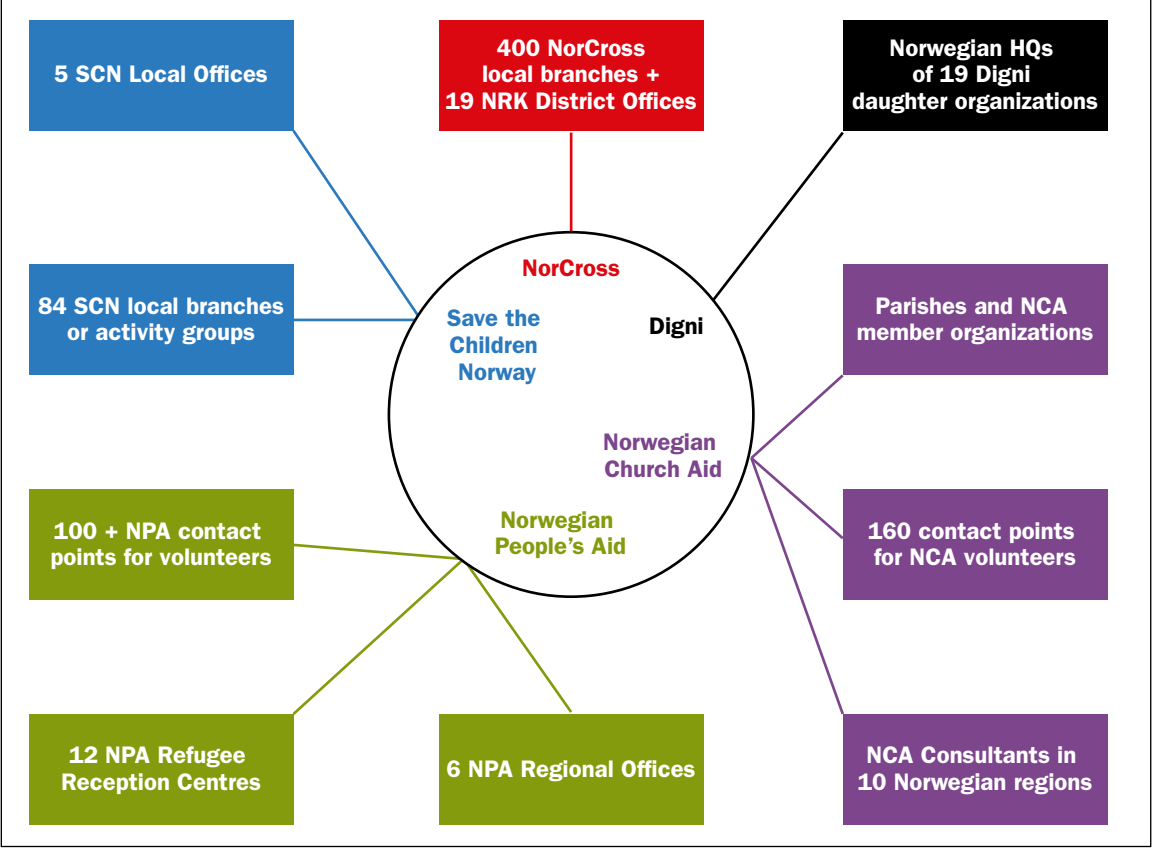
21 Information on the governing bodies of the 6 organisations has been found on their official websites.

groups in Norway; finally, NCA draws from churches and parishes all over the Norwegian territory and has regional consultants in 10 different regions plus 160 contact points for volunteers. Figure 2.2 illustrates the ramifications of the CSOs on the national territory.

Being development organisations, the substructure is not limited to local and regional branches in Norway but also extends to country offices and contact points in other countries, particularly developing countries. SCN – although it's independent from SCI – does not have country offices in the traditional sense. SCN fund and support country programmes in more than 20 countries and continues to be directly engaged in the development of those programmes, even if the administration of these offices are done jointly by involved SC members and through Save the Children International's administration. In addition, SCN supports SCI's global structure implementing programmes in more than 100 more countries. Digni is not directly represented in the field but its 19 organisations have a strong international presence and many of them have also field offices.

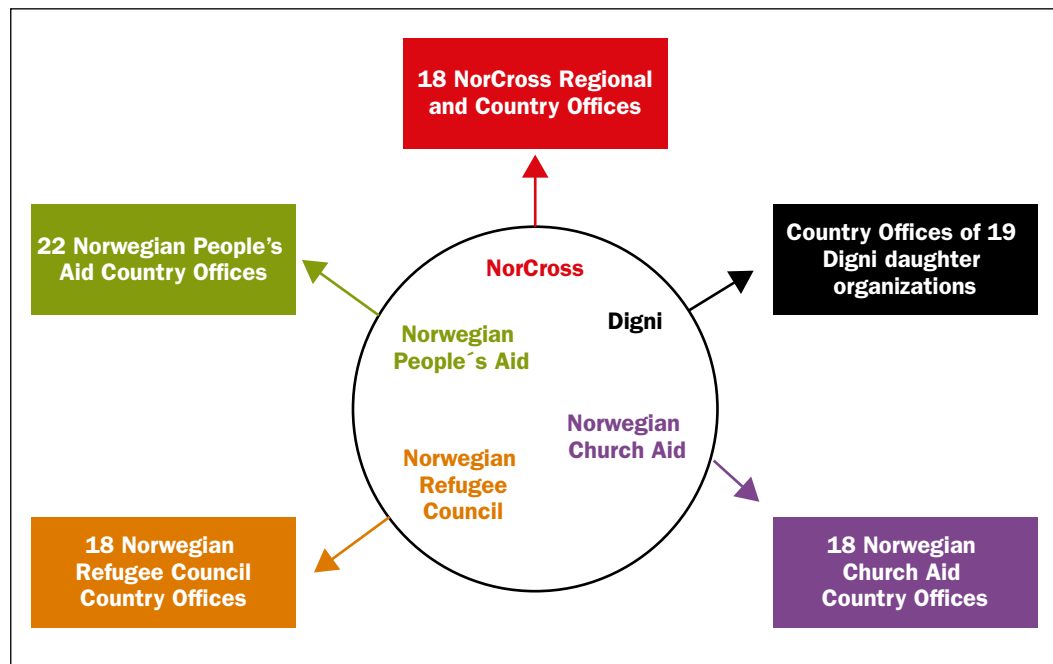
The other 4 CSOs, however, have 18 (NRC and NorCross), 20 (NCA) and 22 (NPA) full-fledged regional and country offices (see Figure 2.3). Half the CSOs under study have their own youth organisation (NorCross, SCN and NCA) and two manage an emergency response team (NCA and NRC, with NRC's being 10 times bigger than NCA's, even though NCA works mainly through partners, even in emergency situations).

Figure 2.2 Ramifications of the CSOs in Norway



All except two (Digni and NRC) rely on volunteer work²², particularly NorCross, which is able to mobilize 40.000 volunteers per year. NorCross's characteristic volunteer policy also accounts for their relatively low staff count. None of the six organisations have a political affiliation, although NPA declares itself “politically independent but not politically neutral”, standing for social justice and for the rights of the most vulnerable groups (they also stress that they do ‘cooperation’ and ‘solidarity’ as opposed to ‘charity’). Finally, four do not have any religious affiliation while two (Digni and NCA) are based on Christian values, focused on human dignity, and collaborate with other faith-based organisations.

Figure 2.3 Field presences of the organisations



Conclusion: the influence of context

The design for this study takes the influence of context into consideration. Background factors may influence the characteristics of the evaluation systems, for example, the open and democratically governed organisations with dense local and regional representation in Norway might carry out evaluation activities through participation and democratic election, without formally commissioning research and studies, particularly from external consultants and researchers. At the same time, more centralized organisations, might lack a number of inbuilt, “natural” feedback mechanisms and might find the feedback they need only in formal reviews and structured evaluation research. Another hypothesis might be that larger organisations would tend to be more formalised and centralised, whereas the smaller ones could be based on more informal patterns of management. Organisations that rely on funding from only one donor agency might become more dependent and hence more likely to adapt systems and procedures like that organisation. It is not likely that Norwegian membership

²² However, DIGNI's 19 members rely to a large extent on voluntary work.

have much influence on the development of M&E systems, national and international partnerships are much more relevant and influential and increasingly so. Whether the context actually has any significance remain to be seen, and it will be matter of discussion in the remaining chapters. A full treatment of subject of possible causal links between the context and the M&E systems is presented in annex 7.

3. Monitoring and Evaluation Systems

Introduction

This chapter provides a comparative description and analysis of how the monitoring and evaluation systems are designed and positioned within the six Norwegian CSOs – in other words the institutional setting within which evaluations take place, the importance given to evaluation within the organisations and how this is manifested in policies, organisational structures, investments in human capacity, financial resources and procedures. The role and functions of Norad will also be discussed – in particular to understand to what extent the dialogue with Norad is appropriate and supports internal and external learning. The chapter is divided into three main sections:

- Policy guidance: this section deals with the overall strategic choices, directives and/or guidance in respect of monitoring and evaluation.
- Organisational structures: this section discusses the independence of the evaluation function, its place in the organisational hierarchy and its relation to other units, human resources, budgets, and institutional memory.
- Process characteristics: this section discusses the processes through which evaluations are planned, implemented and processed onwards, quality assurance, professional development, training and networks.

Under each of these heading both descriptive as well as analytical elements are presented and discussed. The data provided build on the interviews and the document analysis of the evaluation team.

Policy Guidance

Extent of Formalised Policies

The first question to be addressed is to what extent the organisations have a policy for evaluation and, if so, what are its main elements? Our starting point is that a policy should be located in a document that bears the title 'policy' and three organisations (NRC, SCN and Digni) have explicit policy documents. However, as the literature on policy and governance makes clear, policy guidance can exist even if there is no such document²³. The policy intent could be found in other documents, in what are called manuals or rules and

²³ See for example Blakemore, Ken (1998). *Social Policy: an Introduction*, and Paquette, Laure (2002) *Analyzing National and International Policy*. Rowman Littlefield.

regulations, or in one extreme form, 'in the walls of the organisation'. The three other CSOs do have "policy like" documents in this sense, such as handbooks on evaluation, evaluation guidelines and frameworks – providing the basis for an organisational policy on evaluation.

The NRC policy document is entitled "Evaluation Policy. Learning from Experience". According to this, the purpose of evaluation is "to make a contribution towards organisational learning, performance review and accountability". There is also an Evaluation Handbook with a step-by-step guidance on how to prepare, implement and follow up evaluations. SCN has prepared a comprehensive Evaluation Handbook with Save the Children International partners (as part of the harmonisation processes within the Alliance) and there are other documents with requirements and principles for M&E. Digni has a policy document on evaluation from 1999 which was initially prepared in response to a need to increase the quality of evaluations and partly to demands from Norad. This document will be revised in 2013.

NorCross does not have an explicit policy document. There is a document entitled "IFRC Framework for Evaluation" which has been produced by the Federation in Geneva and disseminated amongst member organisations. This is an introduction to evaluation; it presents definition and purposes, and describes the process of evaluation. It does not address monitoring, but defines how monitoring as such is different from evaluation. Furthermore, NorCross has instructions for quality assurance for projects that also contain guidelines for how these are to be evaluated – in the meaning of ex ante evaluation.

NCA has a Global Strategic Plan that includes an evaluation plan for headquarters- initiated evaluations and there is a chapter in the Handbook on Project Management that is close to an evaluation policy. NPA has a book called "Observing Change" which is quite clear and provides detailed guidance on how and why to do M&E. It contains definitions, outlines potential purposes and contains some strategic guidance – in this case expounding on the importance of narratives, qualitative data, and the virtues of participation.

In conclusion, the organisations have a basis and some relevant policy support for their M&E activities. The emphasis clearly lies on the evaluation part of M&E, we have not seen any monitoring policies and the NPA handbook is the only one that deals explicitly with both M&E. Several policy documents and handbooks have been prepared in recent years – most of them of high professional quality and some excellent. The next question is what a really good policy should look like and against what benchmarks might it be judged?

Defining Monitoring and Evaluation

An important part of a policy is to define the subject area and explain why the activities need to be carried out. As we have mentioned, there are

several definitions of evaluation and related subjects; in our view, some are better than others. For example the Program Evaluation Standards²⁴ defines evaluation as ‘the systematic inquiry into the worth and merit of an object’. This is very short, precise and strategic, defining the core requirements while leaving open options for making different practical decisions depending upon particular circumstances, for example, to what extent evaluation should be independent, participatory, or on what precisely it should focus²⁵.

The CSOs mainly use the OECD/DAC definition, but the NRC has adopted the definition of evaluation used by the Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP)²⁶: “A systematic and impartial examination of humanitarian action intended to draw lessons to improve policy and practice and enhance accountability”. SCN uses the ALNAP definition for its humanitarian work, but has in its guidelines adopted the OECD definition of evaluation as “the systematic and objective assessment of an on-going or completed project, programme or policy and its design, implementation and results”. In our view, the Program Evaluation Standards definition has a better fit with what these organisations actually do and could possibly serve them better in a policy.

The OECD/DAC defines monitoring as ‘a continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds’. As the definition shows, this is much less precise than evaluation, which defines a very specific process and even a written report. Monitoring is wider, more complex, more diffuse and dispersed in the organisation, and involve far many more more actors. This is well understood in the organisations and expressed in guidelines and handbooks. But as monitoring can be so many different types of data collection is also less specified in the overall documents, unless it is guidance on a specific monitoring system, as for example the SCN outcome indicators.

Purposes of Monitoring and Evaluation

The purpose of evaluation is usually said to be a combination of learning, guidance and accountability. It could be expected that a policy either affirms all of these purposes or gives the evaluation function a focus, which tilts it towards one or the other of them. Monitoring is seen as different from, but complementary to evaluation as “a continuous collection and analysis of information to assess project progress”, and it is interesting that monitoring is usually not given a purpose in the same distinct way as is evaluation.

24 Joint Committee on Standards (1994) *The Program Evaluation Standards*. London: Sage.

25 One consequence of using this definition is that it does not exclude processes that are termed evaluation by the CSOs. A definition that defines evaluation as related to a specific purpose (e.g. audit) or with specific sets of value criteria, or processes, would do so, and that would mean that the study would not capture the full extent of ‘evaluate inquiry’ (see chapter 1, definitions).

26 see http://www.alnap.org/pool/files/eha_2006.pdf.

The NorCross policy document notes that the purpose of evaluation is to “improve our work, contribute to learning, uphold accountability and transparency, promote and celebrate our work”. We have not previously encountered the notion to ‘promote and celebrate our work’ as a purpose of evaluation. If the evaluation shows that the achievements and results are excellent, the organisations would of course have a reason to celebrate, but it could hardly be the purpose of evaluation to celebrate results. We would suggest the choice of purpose is changed to reflect this. The policy from Digni describes different types of evaluation: who the evaluators can be, how the TOR should be constructed, etc. The new policy emphasises local involvement more than the previous one.

These six policies (‘inferred’ and explicit) have some common features. There is a much stronger focus on learning for improvement than on accountability and control. Accountability is not absent, but learning is dominant. That is fine; it is a strategic choice and should be set out in a policy. Monitoring is prioritised and receives more attention than evaluation – to a large extent because it feeds into and is a requirement for reporting to donors. It has more direct and immediate instrumental value. That, too, is a strategic choice and needs to be set out in a policy.

There are few efforts to craft a more specific and tailored NGO approach to M&E – taking into consideration the idiosyncrasies of the voluntary sector. The Handbook from NPA is one exception – focusing on the measurement of social change. SCN’s ambition to involve children in the evaluation process is another example. The policies also seem to put a higher value on participation than evaluation policies among other actors in development cooperation do, and that may be special feature of CSO systems. These are also reasons why the definitions used in the Program Evaluation Standards might suit the organisations better.

Disseminating the Policy

For a policy to become effective it must be known. If there is no policy document that becomes a problem: how can employees and partners know a policy if there is no explicit policy document? New employees and partners would not know what such a policy is. In Digni, the policy is well known by the larger member organisations and compliance is good because the Digni Secretariat approves all TOR. “Observing Change” is widely disseminated in NPA and among partners. In the NRC, there is an increasing demand for evaluation from countries, and the policy is well known. The same is true for SCN. Awareness of the evaluation guidelines is low in NCA - partly because there is no single policy document and while a policy may be inferred, it is much more difficult to find. The same is true for NorCross, which is in the process of developing a policy, “as the organisation needs something which is more adapted to their own reality than the international framework is”.

Organisational Structures of the M&E Systems

Creating Evaluation Units

The first structural issue concerns whether there is a specific evaluation unit or not within the organisations, and if so, where that unit is located. In theory, an evaluation function could either be located within a separate and specialised unit, or it could be part of another organisational entity – and that could either be a staff function such as planning, audit, research or policy development, or it could be part of operations and implementation. In that case, monitoring and evaluation would be part of the job descriptions of programme coordinators, project managers and even of advisers. There are naturally pros and cons with all solutions.

All six organisations have chosen a structure whereby the evaluation function has its own identity and is either a unit of its own or organised together with other staff functions such as planning, administration, or audits. SCN has organised the evaluation function in a specific M&E unit, which is located in the Strategy and Management Department. There is one M&E Senior Advisor assigned to the Unit reporting to the Director of the Department. The Unit is responsible for coordinating the large thematic evaluations; it takes part in the global harmonisation process of M&E within the Save the Children International, develops and maintains SCN's M&E system and provides technical support and training to country programmes. An M&E network is being established with one M&E focal point from each thematic area (e.g. education, child rights governance), from different functions (e.g. grant managers, fundraising) and across departments, coordinated by the Senior M&E Adviser. The technical advisers participating in the network have M&E responsibilities as part of their job descriptions.

In Digni, there is one person who is responsible for the thematic area of evaluation while all the five advisers work with evaluations and approve TORs related to their project portfolio. In NPA, there is an Advisor for Monitoring and Evaluation²⁷ who from 2008 used to be full time, but whose time devoted to evaluation was proposed to be reduced to 50% in mid-2012, but the intentions were never implemented. These drastic decisions on the priority and emphasis given to evaluation were primarily taken for budgetary reasons and do not seem to reflect any strategic thinking on the role of evaluation in the organisation, though there are obviously such consequences.

In the NRC, the evaluation function is linked to one Evaluation Adviser in the Strategic Management Support Division – a division that also includes strategic planning – reporting to the Director of Management. There is an intention to move the Evaluation Unit higher up in the hierarchy so it will report directly to the General Secretary. Before the restructuring that took place in 2010, the central evaluation function in NCA was part of the Quality Assurance Division, a division with weak links to the International Department. The new Planning, Monitoring and Evaluation Unit is located within the International Department and reports to

²⁷ While the desk officers coordinate the evaluations for their project portfolio.

the Head of the Division. The unit follows up country plans, providing technical support to country offices in the formulation of objectives, the selection of indicators and establishing monitoring systems, etc. The unit carries out evaluation “sometimes, but very rarely”. In NorCross, there is a Coordinator for Quality Assurance and Evaluation that also works on the development and implementation of NorCross M&E system in the Administrative Section of the International Department. There is also a Section for Quality Assurance and Internal Auditing, which reports directly to the Secretary General, but this unit only does ex ante evaluations of project proposals.

Competence and Experience

The six organisations have all managed the human resource requirements for the monitoring and evaluation function well. The evaluation units have qualified and competent staff – often new and young with relevant educational background and professional experience – but rarely with expertise solely in monitoring and evaluation. While there are no PhD or MA degrees with an evaluation focus, those working in these functions do have relevant social science backgrounds that give basic familiarity with M&E theories, concepts and practice, they have had exposure to methods through training workshops and seminars, and they have practical experiences of the function.

The trend is not that senior programme staff moves into evaluation, but that a new group of staff has been recruited during the last five years. The evaluation advisers are all committed to developing an evaluation practice and culture in their organisation. Most of the M&E staff have a broad and ambitious mandate covering a broad range of tasks from training, preparing terms of references, supporting country programmes, commission evaluation, developing monitoring systems, etc.

Independence

The question of independence is often hotly debated in the evaluation literature²⁸. There are two approaches to the independence issue: those who see independence as an end in itself and those who see independence as an instrumental objective. We take independence as having instrumental value, and hence the next question to ask is what is the purpose or aim of such independence? Only on that basis can we then analyse what risks the six CSO are running with the organisational structures for independence they have put in place. Table 3.1 presents the link between practices and the aims of independence. Those aims are basically four in number:

1. To counter deliberate efforts at biasing evaluation findings.
2. To counter unintentional biasing of evaluation findings.
3. To bring an outside perspective to bear on the programme, its design and mandate.
4. To ensure the appearance of independence.

28 John Mayne provides an excellent summary of the issues and the debate in his article '*Independence in Evaluation and the Role of Culture*'. In Barbier and Hawkins (2012).

The table outlines nine types of evaluation practice, and each of these is set against the aims of independence. The arguments for how a particular practice impacts on the objective of independence is taken from Mayne (2012). The orange colour in the cells reflects the practice of the six CSOs and what that achieves in terms of the aims of independence.

Table 3.1 Evaluation Practices and Connection to Aims of Independence

Type of Evaluation Practice	Aim of Independence			
	Countering Deliberate Project Bias	Countering Unintentional Evaluator Bias	Bringing in Outside Perspectives	Providing Appearance of Independence
<i>Using External Evaluators</i>				
• External Consultants	High	Low	High	High
• Organisational Separation	Medium	Low	Medium	Low
• Evaluator Authority	Medium	Low	Medium	Low
<i>Using External Reviewers</i>				
• Audit assurance	Low	Low	N/A	High
• Quality reviewer	Medium	Medium	Low	Low
• External reviewer	Medium	High	Medium	Medium
<i>Involving stakeholders</i>				
• Advisory Committee	High	High	High	Medium
<i>Governance and oversight</i>				
• Reference Group	Medium	Low	Medium	Low
• Governance Committee	High	Low	Medium	Medium

Source: Variables and columns from Mayne (2012), estimates of cells based on interviews and documentation.

A summary analysis of their practices suggests that the six CSOs are quite effective in countering deliberate project bias and in bringing in outsider perspectives. However, they do not convincingly provide an appearance of independence, but such rhetorical face-saving devices would presumably not be of much interest anyway. Additionally, current practice does not address the risk of unintentional evaluator bias. Looking at the cells where their evaluation practices relate to this aim, there are two practices in particular that are identified; (1) using external reviewers, and (2) involving stakeholders through advisory committees. These would be examples of where the aims of independence could easily be reached through changing practices.

The discussion of independence focuses on evaluation and it seems that independence is not an issue in monitoring. However, there are probably times and occasions when it is relevant to ask who collects monitoring data and if that has any consequences for reliability and validity. When an independent evaluation becomes reliant on monitoring data that it has not had any control over, the nature of that independence also needs to be discussed.

Position of M&E Function and Reporting

The evaluation function is organised internally in all the organisations. It is located either in the International Department or a Strategic Planning/ Management Department. Evaluation is not a line function in any of the organisations. The staff members are all advisers. Hierarchically, the units are mostly located two levels below the Secretary General.

None of the evaluation units report directly to the Board. The Boards seldom get evaluation reports and they rarely request evaluations to be undertaken. The Boards' function is primarily to ascertain that the organisations have a functioning M&E system that meet requirements for accountability and learning. As one Board member expressed it 'it is important for us to know that the organisation has a well-functioning evaluation system, but we do not see any reason to get reports on individual evaluations'. There is neither a demand nor an active involvement of the Boards in the planning, discussion and utilisation of evaluations. The Chairman of the Board in NorCross also emphasised that the Board makes sure that evaluation happens and added that the Board had also initiated at least one evaluation.

The Board in NPA deals only marginally with evaluations and does not receive or review any reports. The Board members sometime travel to visit programmes, and then they are briefed on evaluations, if there are any. In the case of the NRC, selected evaluations are presented in Board meetings. There is no demand from the Board – in the sense that they ask for evaluations and take decisions based on evaluation findings and recommendations. Some evaluations are presented to and discussed at the Board of Digni. The same is true for SCN. They receive a presentation on how results have been documented and reported in the annual reports to Norad every year. Moreover, SCN has a Control Committee selected from the Board, which receives an update on how the M&E system develops. The Evaluation Advisor sometimes

presents evaluation summaries to the Board. Evaluations are part of the general reporting to the board in NCA. In later years, the board has been more active in requesting that evaluations should be presented at board meetings.

Another structural issue concerns how the evaluation function is linked to other functions in the organisation. In SCN, the evaluation function is linked with strategic planning and analysis in the sense that both are parts of the same Department. It is also responsible for the monitoring function. This includes also the design, maintenance and management of global indicators. Evaluation is also connected to strategic planning in the NRC as part of the Strategic Management Support Section, but in practice is not so clear how the substantive link between strategic planning and evaluations is operationalized. In NorCross, evaluation is not formally related to other functions, but when an evaluation is started a reference group is set up, whose membership includes the Project Coordinator and the Thematic Adviser. The same is true for NPA. In NCA, the Planning Monitoring and Evaluation (PME) Unit works closely with the Finance Department to produce joint progress reports and with the Area Team Leaders – supporting clusters of country offices.

Budgets and Financial Information on M&E

It is surprisingly difficult to find financial information pertaining to monitoring and evaluation. The overall costs associated with monitoring and evaluations are listed in table 3.2. We have estimated the different cost elements, but as the precise numbers are unknown we have based the estimates on average personnel costs in Norway and on average costs of evaluations. The table outlines the various direct and indirect costs associated with monitoring and evaluation. The real costs in any one of the six organisations might of course be quite different depending on how much evaluation is done in a given year, but as a first estimate we believe this figure of more than NOK 5 million is a realistic 'guess'.

Neither NCA nor NPA have an overall target or budget for evaluations at headquarters. The costs of evaluations at country level are usually included in programme and project budgets. In some cases, it is also possible to apply directly to Norad and MFA for funding of evaluations. Digni, SCN and NorCross estimate that 3-4% of total turnover is utilised for M&E. In NRC, there is no M&E budget at headquarters, but possibly around 1 to 2% were said to be spent on evaluation. The organisations appear reluctant to use general funds to pay for evaluations, and it is common that donors pay for project evaluation at country level. In SCN, the central evaluation budget for 2010-2014 was approximately 9 million NOK. Headquarters spends around 1 million NOK annually, covering the cost of one thematic/strategic evaluation. None of the organisations have clear rules for how a budget for an individual evaluation should be decided. It is said to depend on multiple factors, such as the size of the budget available, donor priorities, complexity of the programme, etc.

While the figures in the table are rough estimates, some estimates are more uncertain than others. The monitoring function is much more dispersed and

often involves small amounts of time from large numbers of people. There is some form of monitoring on each intervention and hence the amount of monitoring data is high, and the costs for retrieving, storing and using it very difficult to estimate.

Table 3.2 The composite costs of monitoring and evaluation (annual, in thousand NOK)

Cost Item	Estimated amount	Comment
Evaluation unit; costs for office and personnel	1.500	Based on estimated annual salary including and benefits, social costs, office and overhead administration
Management of evaluation unit	250	Time allocated to evaluation processes of the 'next in line' supervisor, as above
Costs for policy development and training, guidelines and handbooks	250	Production of monitoring and evaluation guidelines, checklists, and other support costs
Costs for system of outcome indicators; design (discounted over 5 years)	100	Mainly staff and consultants, and reference groups, the project indicators developed in the appraisal phase
Costs for system of outcome indicators; data collection	250	Covering staff time, consultancy support, systems software, etc.
Collection of monitoring data	100	Share of time of personnel in field offices and on projects
Costs for web maintenance	100	Based on average costs for consulting services
Costs for reference groups	50	Estimate of working weeks for stakeholders
Costs for conducting evaluations commissioned from headquarters	800	Estimated costs for one large strategic evaluation, a mixed international team, and with several field visits.
Costs for evaluations in field offices/partners	2.000	Estimated 10 evaluations/reviews at an average cost of 200.000 NOK
Costs for disseminating reports, seminars etc.	10	Printing and distribution of reports and seminars, based on 1 major report/year

Source: The evaluation team's hypothetical estimate of budget components and amounts, presented as an example of an analysis that needs to be done in each organisation.

Institutional Memory

The concept of an institutional memory is important for the use of evaluation findings in the longer run. However, few of the organisations have anything but the most basic and ad hoc structures in place. As such, evaluation reports are difficult to find and only a few are publicly available. SCN has no central database where all evaluations are publicly available and retrieved. Only some of its evaluations can be found in the Norad database. However, SCN sends a complete list of all evaluations to Norad every year, and from 2013 all reports will be made available in SCN's electronic library. In the NRC, all evaluations initiated by headquarters can be found on the English version of the NRC homepage. The large numbers of country level evaluations are not easily accessible. Digni has its own database (PETRUS) where members can upload evaluations, but it is so far not mandatory. However, all evaluations should be uploaded on the Norad database. NorCross has no internal database. Evaluations are in the archives and as such difficult to find. The same is true for NPA. NCA doesn't have a database either, but the reports are on the Norad site – at least all the external commissioned evaluations. Internal evaluations are not archived in a systematic manner.

Where and how monitoring data are stored is another question, as such data flows are not easily identifiable in the way that evaluation reports are. When the data is part of a system of outcome indicators, the information feeds into this system and it is the indicators in the system that are of interest. When the monitoring data is part of the follow-up of interventions, it is usually part of the filing system in the organisation. The retrieval of monitoring data is thus dependent on management responsibilities, and it is normally project/ programme officers who are in charge. When the monitoring data feeds into performance management systems, they become the responsibilities of the units in charge of such systems.

Conclusion on Structures

In sum, the formal structures of the M&E systems are to some extent developed across the six organisations. While the formal structures are in place, a number of more ad hoc structures and processes have been developed. Thus:

- All the organisations have an M&E Unit with designated staff.
- The Unit is typically located within the International or Strategic Planning Departments.
- Such Units comprise one full-time staff member or a part-time person often combining responsibilities for monitoring and evaluation with other tasks.
- The staff are advisers, many newly recruited, well trained and with relevant experience.
- The evaluation function is located internally and as such is not independent. That the evaluation process and reporting lines are not independent is not perceived as being a problem. It is also emphasised that the main purpose of evaluation is learning.

Table 3.3 Summary of policy and structures

Process	SCN	NRC	NorCross	NPA	NCA	Digni
Does the organisation have a formal, written policy on M&E?	Yes	Yes	No	No	No	Yes
Does the organisation have manuals and guidelines on M&E?	Yes	Yes	Yes	Yes	Yes	Yes
Is there an explicitly stated purpose of M&E	Yes	Yes	Yes	Yes	Yes	Yes
Is there widespread awareness of the policy	Yes	Yes	No	Yes	No	Yes
Is there a separate Unit for M&E	YES	Yes	No	No	Yes	No
Are there one or several persons who are responsible for M&E	Yes	Yes	Yes	Yes	Yes	Yes
What is the main role of that unit/ person *	C,A,S,T	C,A,S,T	C,A,S	C,A,T	C,A,S	A
Does the evaluation function report to the Board	No	No	No	No	No	No
Is there a transparent and comprehensive M&E budget	No	No	No	No	No	No
Is there an evaluation database	No	No	No	No	No	No

Code: * Coordination, Advice, Systems Development, Training

Source: Interviews in the six organisations

- The Boards play mostly a passive role. They don't demand and only very exceptionally initiate evaluations and only occasionally read and review evaluations. As such, evaluations are not regularly used in or only contribute rarely to Board decision-making.

- There are formal links between the M&E Units and other Units/Departments in the organisations – mostly the International/Programme and Strategic Management Departments, but the substantive links are weak – making it difficult to assess the actual synergies between evaluation and planning.
- There are no accurate data on the resources (human and financial) that the organisations devote to evaluations. Small amounts of money are used at headquarters level on thematic and strategic evaluations, but there is no overall budget line for such activities. The costs of evaluations at country level are almost exclusively covered by programme and project budgets.
- None of the organisations has a complete overview of all evaluations in a central database. Headquarters-initiated evaluations are mostly available on web-sites, but not programme and project evaluations. The Norad data-base is not systematically used by all the organisations. Only a selection of evaluations are uploaded and made publicly available there.

Table 3.3 presents a comparison of the six organisations. As the table shows there are differences and similarities. It is rather puzzling that none of the differences can be related to contextual factors. For example, while the largest and the smallest organisations both have written policies, the other relatively large and relatively small do not. The nature of membership and governance does not seem to have any significance for whether evaluation functions report to the boards. Nor does the role of the evaluation unit relate to the type of geographical presence in Norway or abroad. The size of funding or the personnel numbers cannot be correlated to any of the contextual factors either. We will return to this at the end of the chapter after we have also reviewed the process characteristics of M&E systems.

Process Characteristics of M&E Systems

The evaluation process can be analysed in terms of how evaluations are initiated (planning and decision-making), how evaluation teams are contracted and supervised, how quality is assured, and how the evaluating findings are received. In the following paragraphs, we will look at these aspects one at a time. Several of the process factors are closely linked to the structural issues as well. It is worth noting that monitoring is not subject to the same discrete decisions; there's no separate starting point when monitoring starts, and no reference groups or the like. There are no special decisions on who will do the monitoring, as that will follow from job descriptions of project staff. Hence, when we discuss the processes of M&E the emphasis is very much on the distinct evaluation process, while monitoring is part of daily management in the organisations.

Decisions to Evaluate

The first question is where evaluations are initiated: low in the hierarchy at country offices and/or at project management level; high at the senior management level and/or in discussions with funding agencies; or formally, according to set rules and regulations, for example determining that there will

be mid-term and end-of project evaluations of all projects or programmes over a certain size – no matter what.

In SCN, the headquarter is responsible for global thematic/strategic evaluations, while country offices are responsible for the programme and project evaluations. The thematic evaluations are suggested in an Evaluation Plan for the strategic period and forms part of the framework agreement with Norad. There is usually one thematic/strategic evaluation carried out every year. In a country programme, it is the thematic programme manager who is responsible for commissioning the evaluation and for the final report. A donor may also require and initiate an evaluation. The decision to evaluate a project, a programme, or a thematic area should be in line with the thresholds set out in the Management Operating Standards. In NCA, project evaluations are initiated locally, while it is less clear who takes responsibility for country and/or thematic evaluations. Headquarters only commissions few evaluations.

In NPA, evaluations are initiated primarily on the basis of an evaluation plan, which is part of the agreement with Norad. In NorCross, suggestions come from advisers and coordinators, and then the Evaluation Coordinator formulates a plan. This plan becomes part of the Agreement with Norad. In Digni, the thematic evaluations are centralised, while project evaluations are decentralised and delegated to member organisations. Every project proposal should include an external evaluation – so the evaluation becomes mandatory. NRC has a central M&E unit, but they have also similar units within some of the country offices. A minimum of one evaluation in each NRC country programme should be conducted every second year. Country strategies and plans are accompanied by M&E plans. Headquarters can request an evaluation of a particular project/ programme, but the general rule is for country offices to initiate and commission country level evaluations.

Table 3.4 How Evaluations are initiated

Process	SCN	NRC	NorCross	NPA	NCA	Digni
At project/ programme and country levels, according to needs	Yes	Yes	Yes	Yes	Yes	To some extent
At headquarters levels, according to perceived needs	Yes	Yes	To some extent	Yes	To some extent	To some extent
According to set rules and regulations	Yes	Yes	No	No	No	Yes

Source: Interviews in the six organisations

There are several overlapping layers to the process of determining how evaluations are initiated. Table 3.4 provides a summary across the six agencies. The planning and initiation of evaluation is a bottom-up and top-down process, but increasingly country offices and partners initiate country level evaluations and are also responsible for the majority of evaluations. Headquarters limit their evaluation work to thematic and strategic evaluations. There is not always a good link between the two, nor an overview of the entire evaluation portfolio, or an effective synthesis of lessons learned.

Decisions on Evaluation Parameters

The question of who actually undertakes the evaluations, whether they are carried out by external experts or by internal personnel has been partly covered when we talked about structures and independence. It is of course also a question of process, and practice varies across the organisations. NRC makes a distinction between internal and external evaluations, but most are said to be external evaluations “since they are more likely to be impartial”. However, teams consisting of both internal and external experts also carry out many evaluations. Major evaluations are advertised and competitive tendering procedures are followed (more than USD 90,000).

In Digni, independent consultants carry out the external evaluations. There has been a gradual change from using external evaluators from Europe to using local consultants “people understanding the local context much better”. For thematic evaluations, there would normally be an external consultant, one from the respective country and the partner organisation involved. In NorCross, most evaluations are done by external experts, but often with an internal person as a member of the team. Invitations are sent to people, institutions and firms that could be interested in tendering. The same is true for NPA. Their guidelines state that a staff member can accompany the evaluation.

In NCA, all the evaluations included in the Global Strategic Plan are external. Normally, someone from headquarters will be part of the evaluation team to ensure internal learning. There are multiple methods for commissioning consultants – direct invitation, a limited invitation to bid and sometimes a public tender. In the case of SCN, Save the Children International recommends that an external, independent team leader is contracted to lead the evaluation, though the final decision will usually be determined by the funds available and donor requirements. It is their experience that this will add to the credibility of the evaluation. It is their practice that proposals from three individuals or firms should be obtained. Large evaluations follow a competitive tender process.

In conclusion, external independent experts recruited through a public competitive tendering process will mostly carry out the larger evaluations. This is a major change from only a few years ago, when only a few evaluations were advertised publicly and often carried out by a small group of consultants well-known to the organisations. However, there is still significant variation between the organisations in terms of how open and transparent they are in the selection of consultants. Here there seems to be a difference between the organisations that are part of international networks (NorCross and SCN) as these networks

facilitate links to external evaluation teams. It should also be added that the available “expert market” in Norway is limited. Chapter 5 examines a sample of 30 evaluation reports. The authors of these reports come from a variety of backgrounds: there are teams from partner countries, Norwegian teams, and mixed teams. The most common is mixed teams. In a few cases, the evaluations are done by one person only, but the most common practice is to have evaluation teams of three to four persons.

Developing Terms of Reference

The quality of an evaluation will often be related to the quality of the terms of reference. Are there norms and standards for the terms of reference? Is the evaluation process defined and will it proceed according to a pre-determined set of rules or procedures? The Evaluation Handbook in the NRC explains the purpose of the TOR, including norms, standards and formats. A Steering Group with representatives from headquarters and the relevant country office(s) is established for each evaluation project to agree on the TOR, select external consultants, review preliminary findings and establish a dissemination and utilisation strategy. An Evaluation Manager is also appointed for each evaluation. Digni has guidelines for preparing TORs. The evaluation process to be followed is also laid out: The member organisation initiates and drafts the TOR, Digni approves these and the team members, and in collaboration with members and partners commissions the consultants and follows the evaluation process through to the end.

NorCross follows guidelines from Norad as well as the Framework for Evaluations from IFRC. This document outlines the evaluation process. The Handbook from NPA has a checklist of questions to ask and outlines the role of evaluators, but there are no standard norms and formats. In NCA, the “Routines and Guidelines” provides guidance and lays out formats for the TOR and the evaluation process. The same is true for SCN – guidelines and formats are available and well formulated.

In conclusion, there are written norms and standards for preparing TORs as well as rules and regulations for guiding the evaluation process. This suggests that the evaluation function has “matured” and been professionalised. It is more difficult to determine to what extent it has stifled and made the evaluation process too bureaucratic and less innovative, but the survey data presented in next chapter indicates that the risk that this might happen is not high.

Management Response Systems

There has been extensive debate on the use of evaluations, and one formal organisational reform that seeks to improve use has been the initiation of management response systems. Such a system makes it mandatory for those in management in an organisation touched or effected by the recommendations of an evaluation to respond to them. They could disagree with the findings but then they would have to explain why; or they could agree with some or all of the conclusions and findings, in which case they should explain what they will

accept and what they won't – and why. It thus becomes transparent how evaluation findings and recommendations are to be used ²⁹.

Again, practice varies. SCN has a formal Management Response System. The response is a public document and gives the programme or project team the opportunity to respond to the recommendations made by the evaluators. The evaluation manager prepares the response within one month of approval of the final report. An Action Plan is developed (more for internal purposes), which lays out how the findings and recommendations from the evaluation will be followed up. The NRC has introduced and enforced a similar management response system.

What happens when there is no formal management response system? The organisation can of course still respond to the conclusions and recommendations in exactly the same way, but the response is then not documented in a transparent manner. In order to know whether the organisation has acted on the recommendations one would have to scrutinize the decisions on the project³⁰. NCA does not have such a system: it is up to the individual Adviser or Programme Officer to review the recommendations and suggest follow up. There is no formal pressure or mechanisms for using the findings and recommendations of evaluations. NPA does not have any formal system either, but according to our interviews management or the reference group would normally write a note in response to evaluation findings. There is a similar practice in NorCross. Digni requests members and partners to respond to all evaluations and discuss the results with the respective organisations. The members have also a responsibility to control the quality of evaluations. They recognise the need to formalise the process of receiving and responding to evaluations better “so that all advisers will do the same”.

In conclusion, two of the organisations have a formal management response system and an established organisational practice for receiving, responding to and following up evaluation findings and recommendations. The existence of a more formal system seems to have increased the number of people reading evaluations and it might also have increased the readership (since a written response has to be prepared). Additionally, it might well also have strengthened the utilisation of the findings and recommendations – though we have not been able to assess the quality and results of the process.

Quality Assurance

Chapter 5 presents our analysis of the quality of evaluation reports. However, it does not examine whether the organisations carry out any quality control themselves. We consider this issue here.

Digni claims that it does apply quality control to all evaluations, but more to the process than the product. One member organisation established a reference group to do the quality control, but this is not standard practice. NorCross has no

²⁹ See SADEV Report 2008:7 *'Reaping the Fruits of Evaluation'* for a full discussion of the topic.

³⁰ Note, one would have to verify whether the management response has actually led to the intended changes. It cannot be taken for granted that actions have been taken even though it is stated in a management response protocol that they should be. The management response system is a formal instrument and a tool to try to ensure a response takes place, but it is not to be confused with the actual decision-making.

established quality control, but the Framework summarises the Program Evaluation Standards (Joint Committee on Standards, 1994). NPA does not have any formal quality control, but its Handbook contains instructions on how to write a good report, make recommendations, etc. In NCA, quality control is a standard procedure in evaluations, but there is no formal checklist for such assessments. The questions asked are: Is this credible? Is this correct? Do we agree with the conclusions? In the NRC, the Evaluation Adviser acts as the focal point for all evaluations and is responsible for quality control. SCN check draft evaluation reports against the original TOR and relevant design documents. In principle, data should also be reviewed for its accuracy.

The evaluation literature contains many books, papers and articles on quality, but these refer to the quality of evaluation reports. We have not come across any discussion of the quality of monitoring – neither of structures or processes, or the reliability and validity of the information contained in them. This could be an area to study more, in particular as monitoring is increasingly used in evaluation and as more resources are devoted to the monitoring side of M&E.

Competence Development

There has been an increasing focus on monitoring and evaluation within the six organisations in recent times, and this seems to be continuing. The question is where the knowledge to act professionally originates? To some extent, it comes from recruiting people with the necessary skills (see above), but it also comes from training. In several of the organisations, the investments in learning have been considerable. In SCN, training is provided for country staff and for head office staff. During 2012 most head office staff had basic training in SCN's new M&E system, approach and global indicators. In the NRC, there is a session on M&E in the induction course for all new staff. The Evaluation Adviser also provides technical support and training to selected country offices. There are no internal training opportunities in NCA, but staff members are encouraged to attend external training (e.g. through the Norwegian Development Network) and a system is under development. An e-learning module on M&E is available for staff globally. The M&E Adviser in NPA routinely has a session in the introduction course for all new country directors on principles in planning, reporting and monitoring. In addition, training is conducted in country programmes with programme staff and partners.

According to the survey (discussed more fully in Chapter 4), more than 80% in total among 207 respondents have had some training in monitoring and evaluation, but many ask for more. Almost 90% think that new staff members should be exposed to training in monitoring and evaluation as part of their introduction programmes. The assessment of evaluation quality presented in chapter 5 also suggests that there is a need to share best practices, to jointly analyse the opportunities for methodological development, and to raise the overall level of quality.

Professional Networks

A very important part of learning and professional development occurs in networks. There are several such networks for monitoring and evaluation. In

Norway there is Bistandstorget (Norwegian Development Network) and the Norwegian Evaluation Society; there are evaluation societies in all countries, and also joint activities at the Nordic level. Organisations such as SCN and NorCross have their own networks for evaluation staff. At the European level there is the European Evaluation Society, and beyond Europe the African Evaluation Society. The latter, in particular, would be a very important network to be part of for all who work in the area of development evaluation.

However, our analysis indicates that for these six CSOs networking is very weak: few of the six organisations take part in any network beyond the Norwegian level or outside their own organisational context. There are exceptions. For instance, the Evaluation Adviser in SCN is a member of the Norwegian Evaluation Society and takes an active part in the international Save the Children M&E network, while the NRC participates in the Norwegian Development Network and ALNAP. NCA has its own national network and is linked to like-minded international networks. M&E issues are discussed within the ACT Alliance. NCA is also a member of INTRAC and the Norwegian Development Network. NPA, NorCross and Digni are also members of the Norwegian Development Network. But there is a large potential for more to be done. Through more thorough networking, the organisations could be stimulated by new visions and ideas and methods; they could develop contacts with more persons and organisations skilled in evaluations, and learn more by having a more active presence in various professional networks.

The organisations with international connections tend to have broader networks than those with only a Norwegian focus. Internal demands, needs and requirements seem to absorb most energy and resources. Networking between M&E staff across the Norwegian organisations is practically absent, and there also seem to be few regular links to Norad's evaluation work.

The Role of Norad

This study is specifically asked to analyse to what extent the organisations find the established system and the contacts they have with Norad appropriate, particularly in relation to internal and external learning, and to what extent the organisations have followed up recommendations from Norad's evaluations³¹? The evaluation team covered the subject through a group interview with staff in Norad's Civil Society Department, and through discussions with each of the organisations.

There is no doubt that the dialogue with Norad is important both for the organisations and for Norad. Most important are the organisational reviews regularly initiated by Norad. However, most of the reviews have commented more on organisational structures than M&E, but there is evidence that they have played a significant role in initiating and supporting reforms within some of the organisations. There are also regular annual meetings between Norad and the organisations where issues of M&E are covered even if Norad has not been very prescriptive. Norad has also to some extent been instrumental in forming

³¹ See Annex 1, and table 1.1.

the M&E agenda through requirements and a professional discussion (SCN). Norad has pushed the RBM agenda and communicated requirements that SCN should develop global indicators, establish a global baseline and stressed the importance of reflecting in annual reports how evaluation findings have been used.

On the other hand, we found little evidence that the substance of M&E carried out by the organisations played a major role in their interaction with Norad. There has also been little direct support and guidance on how the requirements should be met. What we found for example were the following:

- Interest is focused on the Evaluation Plan for the duration of the framework agreement.
- Priorities have been conflicting between various parts of Norad.
- With one exception, there was no indication that M&E policies have been presented to Norad or discussed with them, nor that the organisations have been given feedback on the existence or non-existence of formal evaluation policy documents.
- No indication that the organisational structures have been presented to Norad or discussed with them, or that the organisations have received feedback on this area.
- No indication of discussions with Norad of the evaluation processes, competence development, or networking.
- No presentations to or discussions with Norad of specific evaluation reports, their findings and conclusions.
- No discussions with Norad of quality of M&E or quality assurance processes.

The process of applying for funds and arriving at a framework agreement marginally touches M&E practice. M&E systems/processes are expected to be in place, and brief presentations are contained in the applications that are made. The Evaluation Plan is presented and analysed, and forms part of the agreement and subsequent reporting. However, even though looked at by Norad, there seems to be little discussion and we heard of no occasions when Norad had made any comments or recommendations on the original documentation.

Digni refers to the fact that Norad initially required an evaluation policy. It was first perceived as an imposition, but was gradually recognised as important to the organisation itself. However, Norad is not found to be proactive in the area of evaluation (except for the evaluation database). There is generally no response from Norad on evaluations. As one interview respondent said; "My impression is that Norad doesn't really know what they want. It is like someone going to buy shoes who doesn't know what they want and ends up finding faults in every shoe and not buying any at all".

During the interviews at NorCross, we concluded that the changes in respect of M&E in that organisation were initiated internally. The dialogue with Norad did not appear to have gone into the details of M&E development. The response from NPA was similar. However in the latter organisation, the interview

respondents³² made reference to the Norad-commissioned organisational review in 2007, which was said to have triggered several changes. That review pointed to many practical reforms that could sharpen monitoring and evaluation, and the organisation decided to follow-up on the recommendations. There was no pressure, nor indeed any specific dialogue with Norad on this, so it is not likely that the organisation took these initiatives for reasons other than because its own management had decided that it would be useful to develop the evaluation function.

NCA recognised that significant changes followed the Norad-MFA review in 2006. Most respondents agree that Norad could improve the use it makes of evaluations it receives from Norwegian CSOs in a number of ways, such as in facilitating learning between organisations. For the NRC, the most significant contribution from Norad followed the Organisational Review in 2009. There is no dialogue with Norad in relation to the evaluation function as such. SCN's experience is that professional dialogue with Norad has been shifting. Norad has influenced the M&E work by insisting on more and better reporting of results. However, it has not been prescriptive (nor helpful) in clarifying what that means in practice. The current dialogue tends to focus on administrative and financial matters. On the other hand, the formal requirements for the framework applications emphasise the need for systematic evaluation plans so there is an important indirect effect originating from Norad. In short, the Civil Society Department does not appear to have the resources or the time to read evaluation reports, or to review or enter into a dialogue on evaluation policies, and it does not provide any feedback when the organisations inform about changes in their monitoring and evaluation systems, for example, in the annual reports.

Differences and similarities

This chapter has shown that there are many differences between the organisations; some have written policies and others not. The location of an evaluation unit, or the position of a person responsible for the M&E function in the organogram varies. The roles of the person/s with assigned responsibilities for evaluation vary. However, there does not appear to be any systematic pattern to the differences or the similarities. There is no clear evidence suggesting that one of the six organisations is more centralised than any of the others, or more formalised in its approach to M&E. There are differences, but they seem to have developed more or less ad hoc and according to the ideas people have had when addressing the M&E function and when trying to address particular organisational needs.

This conclusion is based on the findings presented in this chapter. But it does not tell the whole story. We have also applied a different, and more rigorous approach to investigating the differences between the six organisations and connecting them to the contextual setting, using the Qualitative Comparative Analysis (QCA) method. However, this did not point to many clear links or

³² An organisational review in 2011 followed up on the recommendations of the 2007 review and concluded that NPA had implemented all the recommendations pertaining to monitoring and evaluation.

differences either, as the analysis presented in Annex 7 shows. The overall picture suggests that CSOs that are less dependent on Norwegian public funding seem to be more ready than others to devise a management response system and to connect evaluations more closely to decision-making. Table 3.5 presents some of the potential links between context and organisation of the M&E function.

Table 3.5 Implications of Organisational Behaviour

Organisational behaviour and characteristics*	Implications*
More private funding than other CSOs	Decision-making closely connected to evaluations
Less private funding than other CSOs	Absence of a management response system
Less international public funding than other CSOs	Evaluations closely connected to decision-making
	Absence of a management response system
More Norwegian public funding than other CSOs	Absence of a management response system
Less Norwegian public funding than other CSOs	Evaluations closely connected to decision-making
Use of evaluation for accountability rather than learning	Absence of a management response system
M&E loosely linked to planning and reporting	Open and less systematic use of evaluations
	Use of evaluation for learning rather than accountability
	Policy like instruments and documents rather than an explicit evaluation policy
	Absence of a management response system
Instrumental use of evaluation	Evaluation function dispersed rather than concentrated
	Absence of a management response system
Open and less systematic use of evaluations	Use of evaluation for learning rather than accountability
Policy like instruments and documents rather than an explicit evaluation policy	Absence of a management response system
Use of evaluation for learning rather than accountability	Open and less systematic use of evaluations

**The statements of organisational behaviour and characteristics, and implications, are presented in chapter 7 and are based on the requirements of the QCA method to provide brief and qualitative descriptions in both of these categories.*

Conclusion

The policies, structures and processes established for monitoring and evaluation have developed fast in recent years. There are systems in place in all six organisations. They share some common characteristics, but there are also differences among the organisations. At this point in time, the six organisations could learn from their different experiences.

SCN states that the single most important strength of the evaluation system is its ability to produce a large thematic evaluation such as “Rewrite the Future” with a very high technical quality and which has had broad internal and external use. The ability to jointly agree and develop a common M&E system with standards and guidelines, and global indicators at outcome level are also important achievements. The organisation recognises the need to improve its ability to report on qualitative results. There is also a concern pertaining to the challenge of effectively using the increasing number of evaluation reports within the organisation.

The NRC believes that a major strength it has is that evaluation is now embedded within the organisation. The NRC has been professionalised and internationalised during the last five to ten years. No one questions the value of evaluation – even if they might question the value of individual evaluations. It is also recognised that there is a need for a stronger evaluation system in which all the disparate pieces in M&E come together. There is still no central evaluation budget and it is acknowledged that more human resource inputs on M&E are required.

NCA claims that the most important strength of its evaluation system lies in the ability to involve partners at country level and make evaluations useful locally. It is also significant that doing evaluations is now seen as a regular/normal activity and evaluation has become increasingly professionalised. The major weakness for NCA is the inability to share and systematise learning which takes place at country level more widely across the entire organisation and within its broader partner network. It is still working with an informal system recognising that it needs better standards and guidelines.

In NPA, the major strength is perceived to be the quality of a number of evaluations while the main weaknesses are that many reports are found to be too theoretical and that staff don't have time to read and make use of the evaluations. In NorCross, it is recognised that a large number of evaluations are completed at headquarters and in countries. A weakness is the high number of reports produced and the onerous reporting systems: bureaucracy stands in the way of learning. The openness existing with member organisation and the willingness to learn from even critical evaluations are important in Digni – as is also the high level of participation. One weakness is that evaluations are not sufficiently shared, another that all the organisations within the Digni umbrella do not embrace the principles of and approaches to evaluation and critical reflection to the same extent.

Table 3.6 sums up the mapping of the process characteristics of the M&E function in the six organisations, and this table can be read in parallel to table 3.3, which summarizes policy and structural aspects of the systems. The two indicates differences in degree rather than differences in kind, but the two tables also indicate that as the organisations are more alike than different, they have much to learn from each other in respect of how different organisational solutions work. Whether the systems function well also depends on their coherence with the evaluation culture. Are these policies, structures and processes broadly supported by and do they resonate well with broader norms, values and attitudes in the different organisations? This is the subject of the next chapter.

Table 3.6 Summary of process characteristic

Process	SCN	NRC	NorCross	NPA	NCA	Digni
Does the organisation have guidelines for the evaluation process?	Yes	Yes	Yes	Yes	Yes	Yes
Is there a system to assign responsibilities to an evaluation manager?	Yes	Yes	No	No	No	No
Are there guidelines on how to write TOR?	Yes	Yes	Yes	Yes	Yes	Yes
Is there mandatory consultation with partner organisations and local stakeholders?	Yes	No	No	Yes	Yes	No
Is there a formalised management response system?	Yes	Yes	No	No	No	No
Is there a quality assurance system for evaluations	No	No	No	No	No	No
Does the organisation have its own staff training?	Yes	Yes	Yes	Yes	Yes	No
Does the organisation take part in external M&E networks?	Yes	Yes	Yes	Yes	Yes	Yes

Source: Summary of interviews and document analysis.

4. Analysis of Evaluation Cultures

Introduction

As the previous chapters have demonstrated, organisations go about evaluation in different ways. As evaluation is one of many organisational processes it is likely that differences in organisational culture will influence evaluation culture. Organisational research demonstrates the existence of different organisational cultures; there are norms and values that interact with structures and processes in organisations. There are even significant differences between organisations that work in the same environment. There are thus good reasons to believe that there are also differences in evaluation culture. Norms, values and attitudes, for example around leadership, sources of expertise and legitimacy, approaches to uncertainty and risk-taking, the style and nature of debate and controversy, etc. will all affect evaluation.

That being said, there is an argument that the evaluation profession has its own culture. Evaluation culture would be identified by shared norms, values and attitudes on independence, transparency, systematic approaches, and not the least, the value of professional critical thinking. The cultures of evaluation, their national foundations and other levels of integration have recently been brought together in an anthology edited by Barbier and Hawkins (2012). Our analysis builds on the thinking presented in that volume, in particular the notion of organisationally rooted evaluation cultures. We would argue that it is helpful to classify the array of different evaluation cultures into four categories or groups. These four are called the “traditional hierarchy”, the “professional bureaucracy”, the “ad hoc network” and “the trustful group”. They are briefly described in Figure 4.1.

How is an evaluation culture related to monitoring? In fact, they are closely related and one way to describe the differences that follow is that in some cultures the emphasis tends to be on the monitoring side, and in others on the evaluation side. It would be more appropriate to speak of M&E cultures than of evaluation cultures, but here we choose to follow the terms generally used in research on the subject. There are three points to bear in mind when examining and discussing these evaluation cultures. **The first** is that they are all seen as useful in their contexts and have served managers, staff and stakeholders well in drawing out the information they ask for and expect of the system. Why have the systems worked well? Because they fit with ‘how things are done’ in each of the organisational environments. The evaluation cultures and the systems they have shaped mirror the organisational cultures³³. **The second** point is that the evaluation cultures so described are not perfect, even though they function well in these four organisations. Reforms would make them function even better, but

33 See for example; Denison, Daniel R. (1990) *Corporate culture and organizational effectiveness*, Wiley; Hofstede, Geert (1991), *Cultures and Organizations: Software of the Mind*, McGraw-Hill Professional Parker, M. (2000) *Organizational Culture and Identity*, London: Sage.

the process of reform would need to be tailor-made to each organisation; for example, the reform trajectory that would make the 'ad hoc network' function better would look quite different from what would make the 'professional bureaucracy' function better. There is no standard solution that fits all, and the reason each has to be developed according to its own logic is that the overall fit with the context of an organisational culture should not be broken. If an organisation that looks like a traditional hierarchy would try to develop an evaluation culture such as the trustful group this would in all likelihood fail.

Figure 4.1 Four Evaluation Cultures.

The Traditional Hierarchy	The Professional Bureaucracy
<p>Characteristics: There is a clear and consistent hierarchical structure in respect of decisions pertaining to evaluation. There are similarly clear hierarchical structures related to reporting, though not necessarily producing a completely transparent system. As there are no routines to be followed each new evaluation process pursues a course of its own, depending on the actors involved. Evaluation is primarily undertaken for decision-making Support. It is an activity that is feared and respected, and mainly done by external experts. There are seldom any doubts as to whether evaluations are used.</p> <p>Typical Organisations: Public agencies with a history, such as tax administrations, defence, regional and provincial agencies. Large private firms in sectors such as engineering, forestry related products.</p>	<p>Characteristics: Evaluation is strongly formalised and institutionalised; it is governed by policies, rules and routines and anchored in a rational and purposeful system. There's nothing personal or coincidental in the exercise of the evaluation. The evaluation function is professionally conducted and adheres to internationally recognized standards of work. Evaluation is undertaken for control, accountability and learning. It is an activity that is respected and well understood, but there is also an extent of cynicism around evaluation "rituals" and often a fear that evaluations are not put to use.</p> <p>Typical Organisations: Organisations that pioneered evaluation in the 1960s and 1970s, often in the education sector, higher education and research, environmental agencies and development cooperation.</p>
The Ad Hoc Network	The Trustful Group
<p>Characteristics: Decisions to evaluate are taken in a network and the decisions are ad hoc – and so are the organisational solutions. There are no policies and no fixed organisational structures upon which responsibility for evaluation is based. Over the years, it is possible to discern a pattern, and there's also a systematic approach to the methodological development and the gathering of monitoring data. Evaluations are seen as processes and not necessarily encapsulated by the accumulation of written reports. Instead, evaluative information is made available from the websites and through discussions. Evaluation is undertaken to support learning and decision-making, and it is highly regarded, interesting and useful.</p> <p>Typical Organisations: Private sector organisations, particularly small and medium sized, technically oriented, mixed private-public partnerships, civil society organisations.</p>	<p>Characteristics: Few formal mechanisms. An evaluation policy serves to explain how decentralised decisions are made and how participatory and open activities should function – and thus they are also very divergent and context dependent. Managers devolve responsibility to staff, who work closely with and involve partners, which again shows a high degree of trust between partners. Such organisations illustrate the strength of loosely coupled networks and hence the two key words of 'trust' and 'group' characterise the evaluation culture well. Evaluation is mainly connected to learning, and is viewed with enthusiasm, interest, and with a sense of enjoyment.</p> <p>Typical Organisations: Civil society organisations, and generally speaking young, innovative and socially oriented organisations, be they public or private.</p>

Source: Forss (2012) *Four Organisations – Four Evaluation Cultures*. In Barbier J-C. and Hawkins, P. (Ed) (2012).

The third point finally is that these four cultures, though based on empirical evidence, are constructs and many organisations will probably recognize themselves in one or the other, in a mixture of them, or in between. It is not to be expected that an evaluation culture in a real-life situation will be a complete mirror of any one of these evaluation culture constructs. In the following paragraphs we present an analysis of the evaluation cultures of the six CSO organisations. To begin, we look at the aggregate picture because the organisations are actually quite similar: there is a broad consensus in relation to many of the key norms, values and attitudes around evaluation. When there's a difference between the organisations we point that out, but by and large the differences are such that one of the organisations might differ from the others in the response to one question, but when looked at together, the differences are not that significant.

As noted in Chapter 1, a key component of our study involved a survey. A questionnaire survey was constructed and distributed to 284 persons in the six organisations - to what we defined as programme staff, that is, project and programme coordinators, sector and regional advisers at headquarters in Oslo and in field offices³⁴. By the time-deadline set, we had received responses from 207 persons. Some e-mails bounced back because the addresses were wrong and in some cases we received an e-mail explaining why no response was given. In all, there was a response rate of 76%, which should make the analysis reliable.

Table 4.1 Response rates from the 6 CSOs

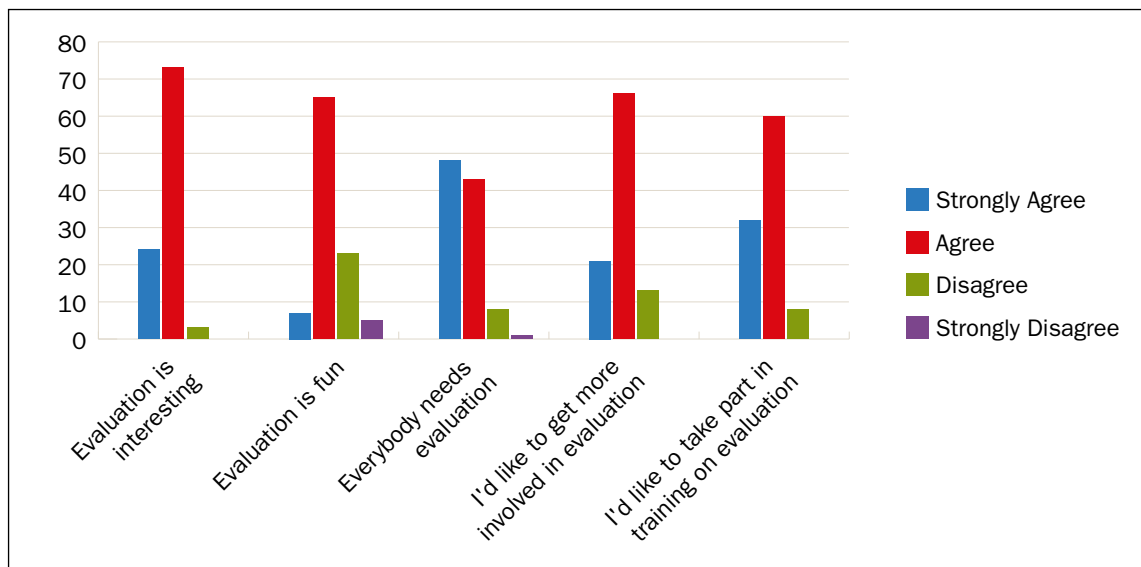
Organisation	Number of surveys sent out	Number of responses	Response rate in %
Digni	57	51	89
NCA	40	32	80
NorCross	52	39	75
NPA	45	18	40
NRC	33	29	88
SCN	47	36	77
Not disclosed in response		2	
	274	207	76

Support for Evaluations

The survey results shows that in general evaluation appears to be an activity that is strongly supported and there are positive associations with it. It is sometimes said that there can be an 'evaluation fatigue', but we found exceptionally little evidence of this here. Diagram 4.1 shows the responses to a number of questions that reflect attitudes towards evaluation.

³⁴ The responses were anonymous and we did not ask for much background information as the organisations are small and the anonymity might have been questioned if we asked for information on, for example, male/female, age, education, position in the organisation, etc. Unfortunately that means the study lacks a clear picture of the respondents.

Diagram 4. 1 Basic attitudes towards evaluation.



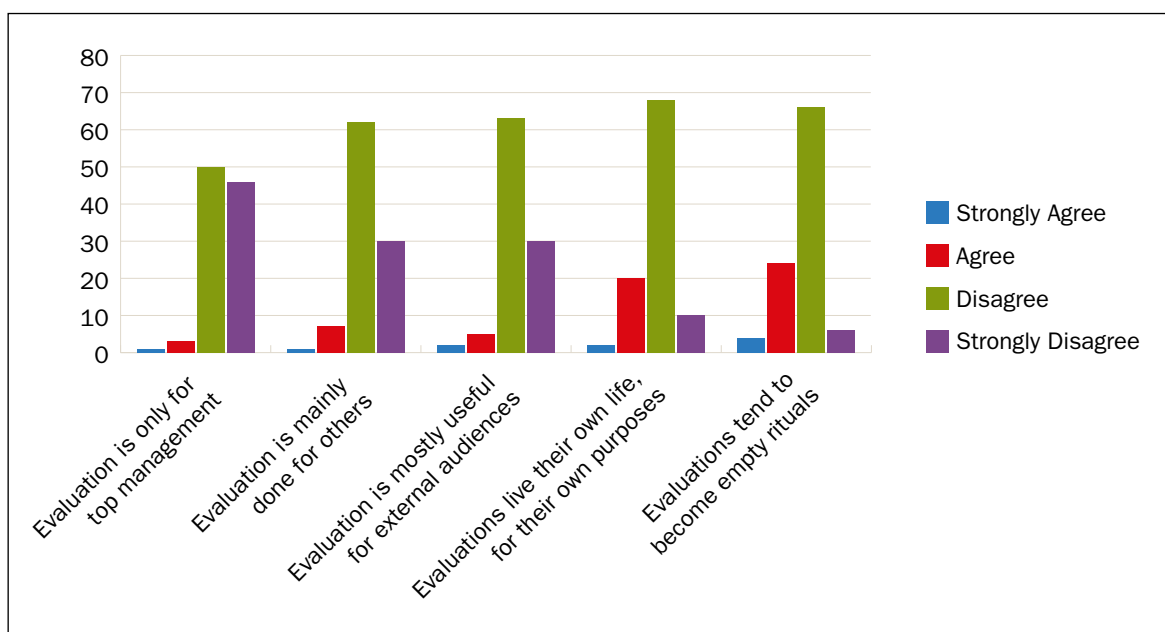
The respondents have a positive and supportive view of evaluation. 97% agree (including 24% who strongly agree) that evaluation is interesting, somewhat fewer agree that it is fun, 48% and 43% respectively strongly agree and agree that everybody needs evaluation. This is remarkably strong support for the function of evaluation.

It is tempting to think that the respondents are so positive because they don't really know much about evaluation. That, however, would be a mistake. 85% of respondents have had some form of training in evaluation (17% through a university programme, 58% from some external workshop, and 79% from some form of internal programme/workshop or the like), 71% have been engaged in evaluation (this could be in many roles, such as commissioning, part of team, part of reference group). Only one respondent stated s/he never read evaluations. Not only are the respondents very positive, they often have considerable experience with the subject. Furthermore, it has obviously been experiences that encourages people to be exposed to evaluation again. As the diagram shows, 87% would like to get more involved in evaluation activities, and 92% would like to be exposed to training on evaluation.

Who Benefits from Evaluation

The perception of who benefits from evaluation may be a reason for the positive attitudes to evaluation. Diagram 4.2 shows the degree of agreement/disagreement to a number of provocative statements about the purpose of evaluation. It is clearly seen that the respondents primarily see evaluations as a tool for themselves and their colleagues. They disagree strongly with the idea that evaluation would mainly be for top management and/or for others, including for external audiences. They also disagree with the proposition that evaluations tend to become empty rituals – organisational processes that live their own lives, for their own purposes.

Diagram 4.2 Stakeholders in Evaluation



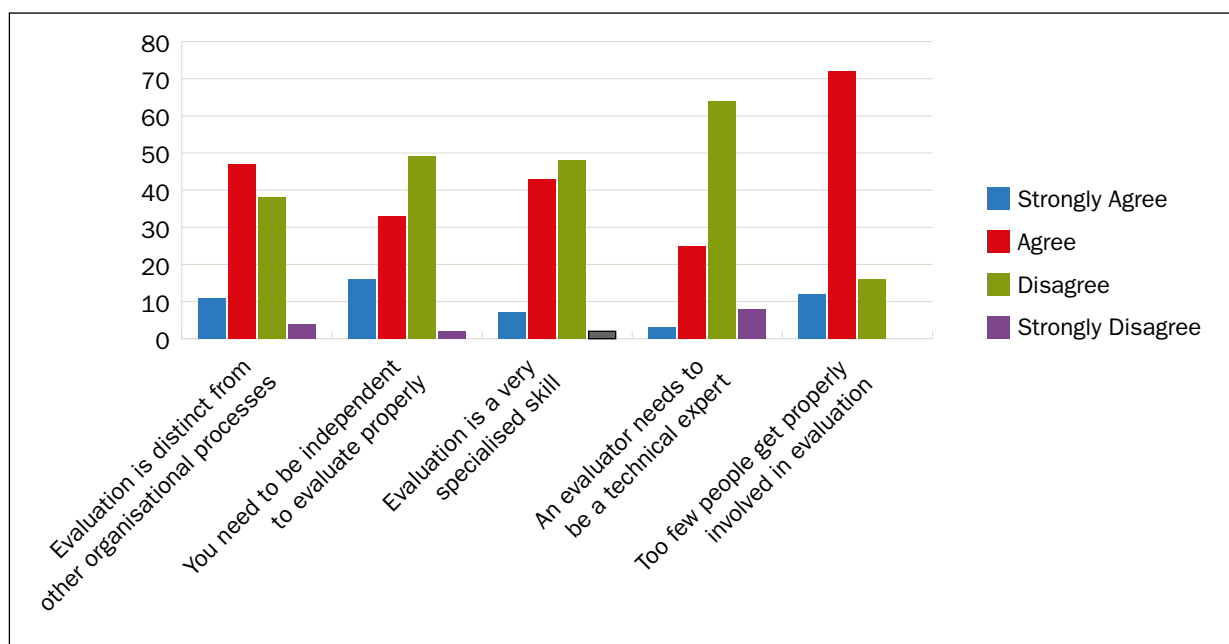
These results are also surprising. The evaluation literature is rich with examples of how evaluations become rituals: when Vedung (2000) talks about use of evaluations he mentions ritual use as a category of its own³⁵. Some 20 – 25% of the survey respondents agree with the statement, and they probably have some concrete experience in mind. The respondents to the survey are primarily coordinators and advisers, and staff in country offices. They do not represent top management. It is also interesting that they see themselves as the persons who benefit from evaluations. This should have consequences for the location of the evaluation function. Chapter 3 discussed where an evaluation unit might best be placed. The six organisations have placed the evaluation unit within an international department. It is not an external function or a function reporting separately to the board or top management. The organisational structure supports the view of who is to be served and probably also reflects such views. It is also clear that evaluations are not expected to serve other agencies; they are not done for Norad for example, or for other external audiences.

Beliefs about Evaluation

What kind of an activity do the respondents take evaluation to be? The majority have training and practical experiences in evaluation, but still evaluation can be many things (which, by the way, is supported by 70% who agree with the statement that evaluation is seen to be pluralistic). The survey responses in Diagram 4.3 do show results that are difficult to interpret. About equal numbers agree or disagree with the notion that evaluation is distinct from other processes in the organisation. Those who agreed with that statement also tend to agree with related statements, such as that evaluation is a very specialised skill, that

35 Vedung, E. (2000) *Public Policy and Programme Evaluation*. Transaction: New Brunswick

Diagram 4.3 Requirements for the Evaluation Process



you need to be independent to evaluate properly, and that the evaluator needs to be a technical expert. There is a high degree of correlation between the answers to these questions, but it does not provide a complete fit. Many who agree that evaluation is a distinct process do agree that too few get properly involved.

These opinions tend to influence attitudes to participation. Thus the replies suggest that the more specialised are the skills required, the less integrated the evaluation process is, and the more external the less scope there is for participatory approaches. These responses indicate that while there is a high degree of consensus in favour of participatory approaches to evaluation, there are also many who see a dilemma in that the more the necessary technical skills, competences, and structural features of evaluation (such as independence³⁶) are pushed the more difficult it is to adopt genuine participatory approaches. The same split is revealed in responses to another statement related to training. Thus 49% agreed and 51% disagreed with the statement 'Only those who are specifically trained for the purpose can evaluate properly'. This explains the very strong support for evaluation training shown in Diagram 4.1.

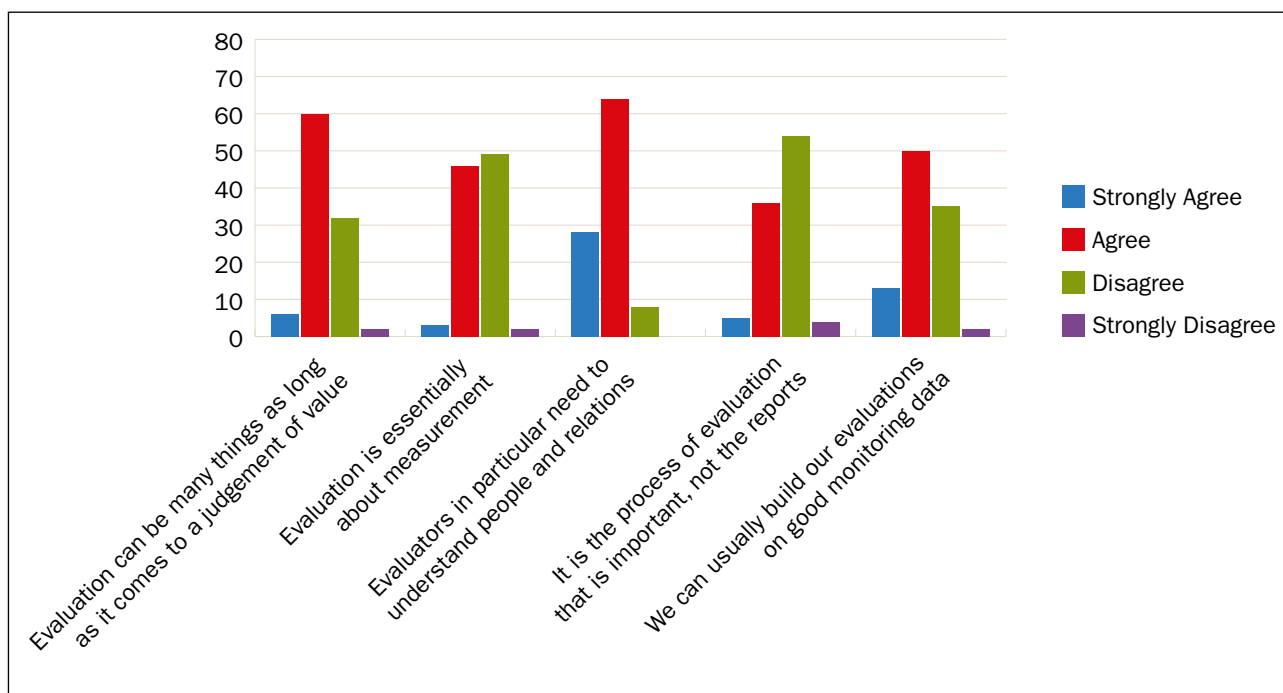
In the discussion of the pros and cons of participatory approaches, it is sometimes said that you need to be close to activities to really understand them and to be able to evaluate properly. That view was presented as a statement to agree or disagree with, and a majority of 75% disagreed. Such widespread sentiments underpin a belief in the virtue of independence, the credibility and rigour of scientific methods and due process. They are also connected to a strong belief that these processes serve the organisation well. There does not

³⁶ The survey did not define 'independence' (nor other words) and hence the response is based on the recipients own understanding of the concept, and that may vary from one person to the other.

appear to be any perceived contradiction between independence and specialised skills and the utility of evaluation for the organisation.

Research on evaluation can identify different approaches to the evaluation process. There are those who emphasise the measurement of objectives and there are those who emphasise the process of evaluation, the space created for critical reflection, and the deliberation on worth and merit – rather than measurement. The evaluation culture is a short-hand for the orientation in this discussion; norms, attitudes and values on these issues differ sharply from one context to the other. Diagram 4.4 shows the value orientation of the six CSOs.

Diagram 4.4 The Nature of the Evaluation Process

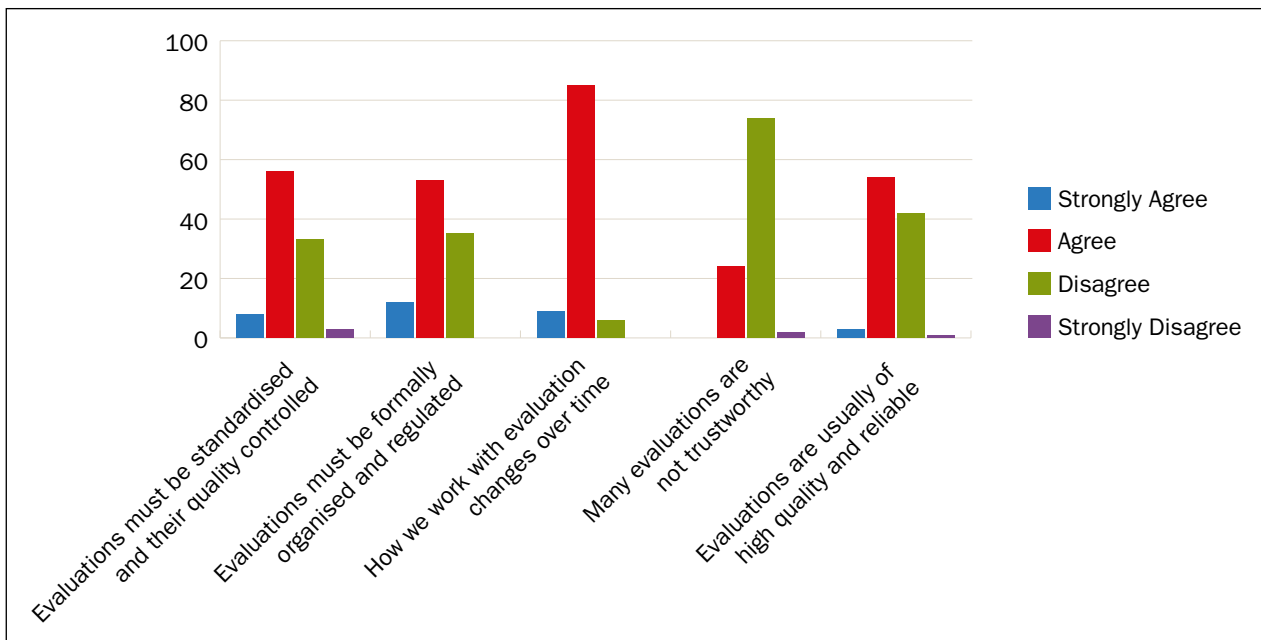


First, the results show that there is a relatively strong agreement that evaluation, at its core, is about judgement. The Program Evaluation Standards define evaluation as ‘the systematic inquiry into the worth or merit of an object’ and that definition places the judgement at the centre far more than the OECD/DAC definition does. The latter would lead the thinking more in the direction of measurement, and that thinking is also reflected in the responses, but there is still a slight majority who see evaluation as broader than merely an exercise in measurement. The third question in Diagram 4.4 makes this even clearer as more than 90% agree that evaluators must understand people and relations.

It is worth noting that 63% of the respondents to the survey think that evaluations can build on monitoring data. As next chapter shows, most of the evaluations do not build on monitoring data, and the evaluation teams often comment on the low quality/difficulty of finding and using, monitoring data. There is thus a mismatch between the real situation facing evaluation teams and the beliefs about availability of data. It is also interesting to note that the majority disagree

with the statement that the process of evaluation is more important than the report produced. In the wider evaluation debate, it is sometimes said that the process of evaluation is as important as the report, and hence it is argued that if you want to maximize use of the evaluation process, people need to get involved in the process³⁷. The attitudes reflected here seem to be that while people should certainly get involved in the process, the report is also seen as important. As there are limits to how many can actually get involved in the evaluation process, this reflects a realistic view of the use of time. The report is also important for the sake of transparency, documentation, and in the long run for the institutional memory of the organisation. With these different approaches, and in recognition of the dilemma around participation and involvement, it is not surprising that the survey responses also show a need for firm structures and processes around evaluation. Diagram 4.5 presents the different views on these issues.

Diagram 4.5 Quality and quality assurance



The diagram shows that the majority of respondents are of the opinion that evaluations are trustworthy, and they are often of high quality and reliable. But there is still a share of some 25% who disagree, and who also qualify their agreement when the statement is sharpened to say that ‘... evaluations are usually of high quality ...’. Increasing the attention to quality control and to more formalised approaches to organisation appear to be logical responses. Chapter 5 will turn to an external analysis of quality, which in many ways confirms the opinions expressed here.

It is interesting that there is widespread support for changes in the way the organisations work with evaluation. Chapter 3 showed that structures and processes have changed significantly in the past 5 years. The opinions expressed through the survey call for further changes. Table 4.2 presents

37 Michael Q. Patton (2004) *Utilization-focused Evaluation*. Sage: London; Forss, K., Carlsson, J. and Rebien, C. (2002) *Process Use of Evaluation – Types of Use that Precede Lessons Learned and Feedback*, Evaluation, Vol 8 (1) 29 – 45.

responses to three statements on potential changes in the organisation of evaluation. There is a very strong support for a formalised and structured approach to the evaluation process. This is also connected to a belief that evaluation findings must be used for future projects and to communicate with stakeholders. A belief in the benefits of evaluation and a commitment to the effort to make the most of it could hardly be expressed in stronger terms.

Table 4.2 Responses to statements on formalising structures.

Statement	Strongly Agree	Agree	Disagree	Strongly Disagree
The findings from evaluations need to inform, shape and influence future projects	58%	40%	2%	-
The recommendations should be either formally accepted or rejected and if accepted, there should be a system to ensure they are followed through	35%	60%	5%	-
The findings from evaluations need to help shape the messages that we use to communicate what we do to our members and the general public	29%	65%	6%	-

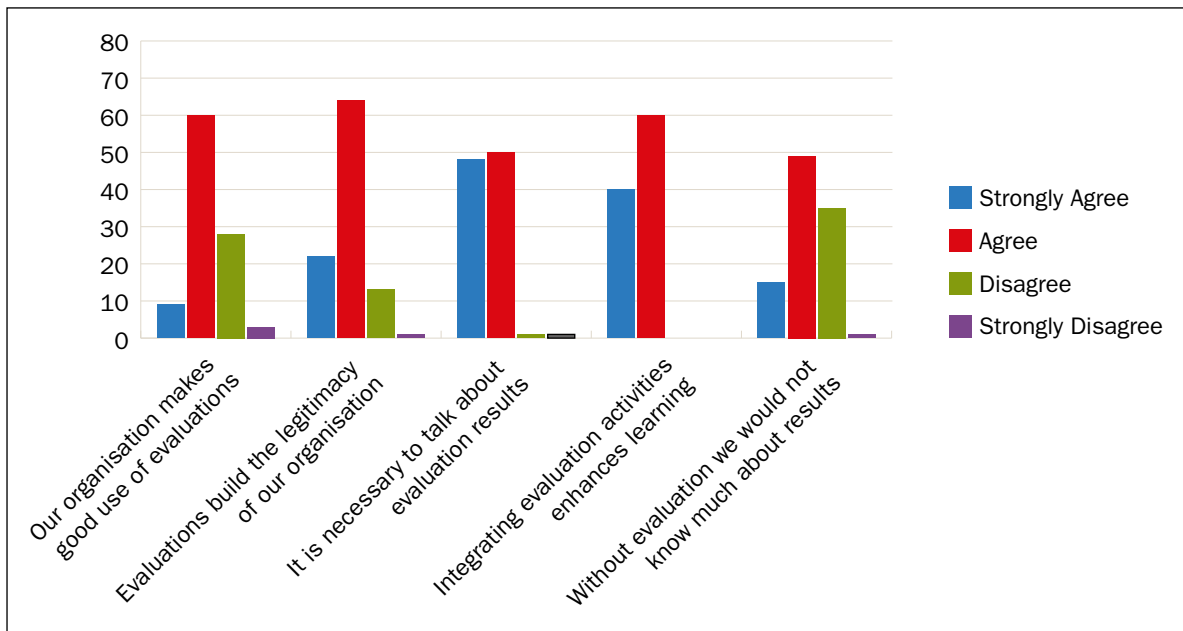
Use of Evaluation

But if that's the way people would like evaluations to be used in the future, what's their opinion on how evaluations are used at present? Again, the opinions expressed in the survey affirm strong beliefs in the utility of the evaluation process. Diagram 4.6 shows responses to a set of statements on the use of evaluations. First, almost 70% find that their organisation makes good use of evaluations. Almost 90% agree that evaluation builds the legitimacy of the organisation, and everyone agrees that it is necessary to talk about evaluation results. There is also unanimous support for integrating evaluation activities to other processes within the organisation in order to enhance learning.

That brings us back to the earlier discussion of how the survey gives expression to a wish to combine things that are usually seen as contradictory, namely;

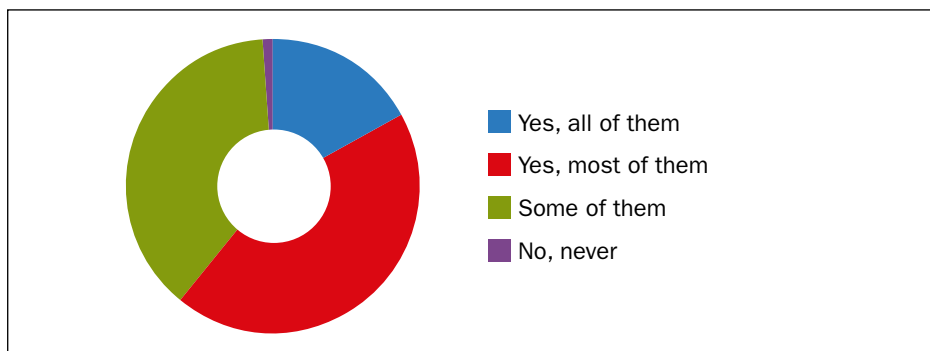
- To work with independent resource persons **and** also for partners and internal stakeholders to be involved.
- To deploy specific evaluation expertise **and** to also use technical expertise from the organisation and partners.
- That evaluation needs specific structures and processes **and** also that it needs to be integrated with other functions.
- That evaluation is primarily done for internal use **and** that it is also important for external ends, for example to communicate with members.

Diagram 4.6 Opinions on the Use of Evaluation



The responses endorse a view of the instrumental use of evaluations. 74% of the respondents agree (including 12% who strongly agree) that doing more evaluations would help convince partners to implement necessary changes. Close to that, there is also a view of the evaluation process as essentially serving specific decision situations – 92% agree that evaluations must arrive in time in order to be useful when making decisions. But in contrast, 98% agree that evaluations must be allowed the space and provide the time for reflection. It is also interesting to note how the respondents engage with evaluation reports, or rather what the reading habits are in regard to evaluations. 67% agree that they read evaluation reports as often as they can. Diagram 4.7 shows that the very high number of the respondents who actually read quite a lot of the evaluations being produced in their organisations. As chapters 3 and 5 show, the organisations produce a small number (1-4) of headquarters initiated reports every year, and several initiated at field levels (usually above 30).

Diagram 4.7 Response to the question ‘Do you normally read the evaluations commissioned by your organisation?’



The next question in the survey asked the respondents to what extent they read the reports, if it was the full report, or if they glanced through the report and read the executive summary, or if it is usually only the summary. The response revealed that 84% usually read the full report, and just 4 persons who said they usually only read the summary!

Reading a report must of course not be confused with using evaluative information, but it is usually a necessary precondition. If you have not been involved in the process it is through reading that you see the force of the arguments being made and can take part in the lessons learned. Reading and time for reflection are key ingredients in making use of evaluations, and the responses are encouraging. However 38% state of respondents say that they rarely read evaluations from other agencies such as Norad, Sida, Danida, and other CSOs. Still, the most common response to the question is that 51% say they sometimes do read evaluations from other agencies, and 8% do so frequently.

Changing Evaluation Cultures

So what do all these survey responses tell us about evaluation culture? The first question to reflect on is if we are faced with one evaluation culture, several, or even six different ones? None of the responses summarised in the tables above or discussed in the text presented data grouped together by organisations, but in some cases they are significant. For example,

- 91% (29 persons) of the respondents from NCA agree that ‘a penny spent on evaluation is a penny invested in learning’, but only 61% of the respondents (11) from NPA (average 86%).
- 72% (36) of the respondents from Digni agree that evaluation is a good learning tool, but only 44% (8) from NPA (average 64%).
- 71% (36) of the respondents from Digni agree that evaluations are of high quality and reliable, but only 35% (10) from the NRC (average 57%).
- 94% (47) of the respondents from Digni agree that the organisation makes good use of evaluations, but only 48% (18) from NorCross (average 69%).

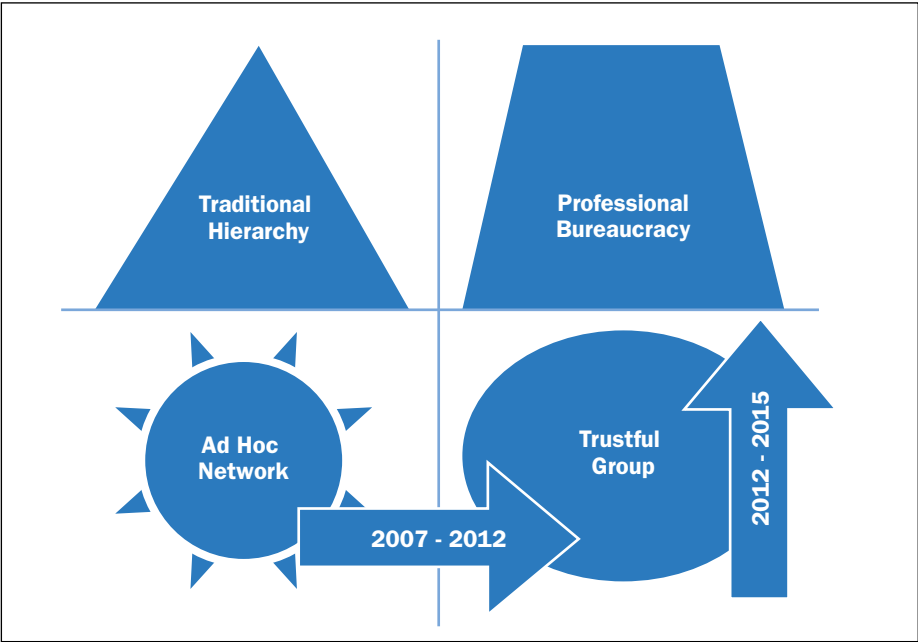
Thus there are differences, but they do not appear to be systematic. The response rate varies between the organisations, and we have a much lower response rate from NPA than from any of the other organisations and hence the few responses from NPA can hardly be taken to be representative for that organisation.

The second question is to ask if the responses given indicate a close fit with any of the evaluation cultures presented in Figure 4.1. At one level, and not surprisingly, the organisations when taken together do appear to come quite close to the characteristics of the Trustful Group. The evaluation function tends to be decentralised and not much subjected to formalisation and standardisation. Participation and engagement are lead words, and people have strong beliefs that evaluation is for themselves and for learning more than for top management and the outside world: learning is more important than control. There is an openness and transparency around the evaluation processes, and a willingness to engage with expertise. People emphasise the need for time and reflection.

Most importantly, the attitudes towards evaluation are positive: it's perceived to be interesting and even fun; the respondents take part in evaluation processes; they have had and want more training; and they see investment in evaluation as necessary and 'profitable'.

While most have the experience that evaluation processes yield trustworthy information, there are, however, those who are sceptical of the reports and many who recognize that there could be too much of a good thing, and that there's a risk of creating rituals. But evaluation cultures are not fixed, and there's also a widespread recognition that evaluation practices change. In order to address the shortcoming of the present systems, a large majority favour more guidance, stricter quality controls and standards around evaluation, and formalised management responses to evaluation recommendations. That would take the evaluation culture in the direction of the Professional Bureaucracy culture presented in Figure 4.1, though there are too many values around participation, learning and the nature of evaluative thinking at present to take the organisations right into that evaluation culture, even if it is perhaps possible to hypothesise a trajectory of change such as that illustrated in Figure 4.2.

Figure 4.2 Recent and possible projected changes in evaluation culture among the six CSOs



The arrows in the figure suggests that there has been a change in evaluation culture over the last five years, moving the organisations into more formalised, though not more centralised approaches to monitoring and evaluation. While doing so, it seems that core values around participation and transparency have been reinforced, including in respect of the evaluation function. While the function has received more attention and been more subjected to explicit design, it has also come to be more aligned to core values within and across the organisations. We have also suggested a trajectory of change for the next 5 years. While individual organisations do not necessarily follow exactly the same

trajectory of change, the evidence presented here suggest that they do follow the same general path of moving from a more ad hoc, informal approach towards a more structured, formalised approach to evaluation. The overall future direction of change is of course unknown, but if the recommendations made in this study are implemented then the direction of change shown in Figure 4.2 seems likely to occur. It certainly echoes the concerns and expectations expressed in the survey as well as in the review of structures and processes in chapter 3.

Conclusion

In sum, these opinions shed light on a multi-faceted and complex evaluation culture that strives to achieve many ends and reconcile what can be competing objectives. It is not a naïve approach. The survey responses also show well-informed criticism of, for example, the way decisions are made. A full 65% agree with the statement that many decisions are taken without any evidence base from evaluations.

How reliable are the survey results? It was noted above that 73% of surveys sent out were completed and returned. While 73% is a fairly good response rate, we need to consider not so much the percentage of non-respondents but whether they represent a particular group of staff contacted. It is often assumed that those who do not respond to a survey are those who care less about the subject, who have less knowledge, less interest, or, in this case, less experience of evaluation. To the extent that this sort of bias is reflected in the current survey responses, it is likely that all the responses discussed in this chapter probably reflect norms, values and attitudes about evaluation drawn from amongst a more 'enthusiastic' part of the staff of the organisations likely to be more positive and enthusiastic about evaluation. Additionally, as noted above, the comparatively low response rate from NPA suggests that one needs to be particularly cautious in assuming the general responses are an accurate reflection of views right across this particular organisation. All this needs to be born in mind when reading the results presented here.

Values, norms and attitudes interact with structures and processes. They relate to each other in complex ways. Structures and processes reinforce values, but they are also shaped by values. There are many and frequent feedback loops between them. Over time, they evolve together – at least in well-functioning organisations. If they do not reflect each other and if they evolve in different directions, there will be problems. Comparing the analysis in chapter 3 with the analysis in chapter 4, there is an alignment between the norms, values and attitudes around evaluation and the organisational praxis. The three aspects of structures, processes and culture do seem to form a coherent whole. The culture gives strong support for the evaluation function. People have high expectations around evaluation and evaluation is associated with positive attributes (interesting, fun, useful). Many have both training and practical experiences, but also want to learn more and to become more involved in evaluation. It is broadly recognized that evaluation requires specific skills that

involve knowing about people and relationships, as well as measurement and results. The doubts that people have and the scepticism that some give voice to often relate to the quality of evaluation reports and to the use of evaluation. In the next chapter we turn to an assessment of the quality of reports, and following that, in chapter 6, to use, and in particular learning from evaluations.

5. The Quality of Evaluation Reports

Introduction

The concern with quality in evaluation is often taking as its starting point the year 1974 when a number of stakeholders in the American evaluation community formed the Joint Committee of Standards. To cut a long story short, the Committee published a standards proposal 20 years later in 1994 and that, in itself, is revealing as it indicates that it is not easy to define and agree on what constitutes quality in evaluation.

The Joint Committee suggested 30 standards, grouped into four categories: utility, propriety, feasibility, and accuracy. These standards were path-breaking. They have been widely disseminated and applied in a number of different contexts to assess evaluation quality. They were recently updated³⁸, and have served to inspire most other approaches to quality, such as within the European Commission, OECD/DAC, and many national evaluation societies³⁹. Many organisations have used the Program Evaluation Standards to develop their own quality criteria.

Quality is created through a process that goes from the origins in terms of reference and evaluation questions and based on that view of the process we developed the format for assessment presented in annex 4⁴⁰. The format was then used to assess five evaluations from each of the six organisations. Our study focuses on the last 5 years and the organisations gave us lists of all evaluations completed during these years. We selected evaluations according to the following criteria: that (1) we would select those evaluations that cover the largest programmes in monetary terms, and (2) activities that were particularly innovative. We also wanted our sample to reflect the diversity of operations. We focused on evaluations commissioned from headquarters, but the sample of 30 reports includes evaluations commissioned from field offices. These 30 reports represent around 25% of evaluations commissioned during the last 5 years (from headquarters, much less from field offices), so they could be expected to give an indication of quality. The evaluations selected are listed in annex 5.

38 http://www.sagepub.com/booksProdDesc.nav?prodId=Book230597&_requestid=255617.

39 OECD-DAC Quality Standards for Development Evaluation, DFID Quality Assurance Template, Europé Aid Quality Control Checklist. Spencer and colleagues (2003) reviewed 29 frameworks geared to qualitative research (Quality in Qualitative Evaluation: A framework for assessing research evidence. Government Chief Social Researcher's Office. London: Cabinet Office.

40 A slightly different version of the same format was used in the study of quality in Sida's evaluations; Sida Studies in Evaluation 2008:1; 'Are Sida's Evaluations Good Enough?'

These 30 evaluations provide a picture of issues related to quality, but 25% is still a small sample. When we describe the quality in the text below we refer to these 30 reports, not to any others. The reader should be careful of generalising to all reports – it is likely that there are evaluations that are both better and worse. In addition, we do not compare quality between the six organisations. With a sample of five reports it is not realistic to detect systematic differences between the six organisations – if there are any. Based on the 30 reports, they do seem quite similar. The full model for quality assessment is presented in table 5.1 and it was also elaborated in the Inception report of the study.

Box 5.1 Format to assess quality – an operationalisation of the Programme Evaluation Standards

Descriptive category	Main issues assessed/described
Description of 'systems aspects' of the evaluation	<ul style="list-style-type: none"> • Cost of the evaluation • Sector, nature of evaluated object • Region • Evaluators/evaluation team • Host country participation
Description of methodological choices	<ul style="list-style-type: none"> • Basic evaluation question(s) • Evaluation design • Evaluation methods • Use of data collection instruments
Assessment of methodological choices	<ul style="list-style-type: none"> • Terms of reference and basic question(s) • Clarification of design and methods • Discussion of validity and reliability • Appropriateness of methodological choices • Design of data collection instruments
Assessment of evaluative findings	<ul style="list-style-type: none"> • Trustworthiness of assessment of management and implementation • Trustworthiness of assessment of outputs, outcomes, and impacts
Assessment of style and presentation	<ul style="list-style-type: none"> • Structure and presentation • Style of writing • Creativity in developing the report
Assessment of concluding matters	<ul style="list-style-type: none"> • Conclusions that are based upon evidence • Recommendations that follow from value premises, data analysis and conclusions • Lessons learned that are clear, succinct, and follow from empirical observations

Source: The model was presented in the Inception report and is developed from the format used in the Sida study of evaluation quality (Sida, 2008).

A Quick Overview

Before we go into the details of evaluation quality, we present a rapid overview. The model contains a holistic assessment, which should remind us see the forest and not only the trees. Quality is summed up in relation to five broad issues⁴¹:

- The **methodological aspects**; that is if the evaluation design, methods and instruments provide answers to the key questions and if, in general, choices are well argued and performed with self-critical reflection.
- The **presentation** of the evaluation; that is, everything related to structure and style, all weighed together and analysed as a whole.
- The **substantive understanding**; that is, the comprehension of the substance of the interventions, demonstrated in descriptions, analyses and assessments of issues at stake, for example in respect of organisational development, peace-building and conflict resolution, and other key subjects.
- **Creativity and innovation** in the evaluation process. Evaluation is a form of applied research, often conducted on a limited budget and with many other constraints. But it is conceptually tied to learning, and learning always benefits from creativity and innovation. To what extent have the evaluations been creative in their formulation of questions, in their search for answers, and in their communication of findings?
- **Depth and quality of information**; that is, whether the main information demands are clearly stated in the report as a whole, and particularly in its conclusions, lessons learned etc. Are the messages – whatever their focus – trustworthy?

Table 5.1 presents the results of our assessment⁴². The maximum score for each evaluation along these five variables was 30, as the maximum score on each was 6. The variables in the model each have the same weight. We have used a rating scale from 1 to 6 throughout the assessment. Consequently the minimum score that a report could get was 5. The distance between 5 and 30 was split into five categories, and thus we can see how many evaluations are in each category. This is presented in the table. The table shows that there are no really bad reports and there are also few excellent reports (two). Most are found around the centre, but there is no central tendency, the distribution is skewed towards the “upper end”. Indeed, it is notable that as many as 76% of evaluation reports are found to be rated as good. This means that they score highly in relation to a few quality indicators and more poorly in others. They would not achieve a total score of more than 20 if they were assessed as achieving scores of less than 3 in one or more categories.

41 Based on the model in the Sida study from 2008. The five issues presented and defined in the Sida study are also derived from the Program Evaluation Standards, but the issue of ‘depth and understanding’ is an elaboration of the feasibility aspect of the evaluation.

42 The assessments were made by Stein-Erik Kruse and Kim Forss, who first calibrated the use of the scales and then rated 15 evaluations each. A sample of 5 reports were selected to verify that we rated the qualities in the same way. Ratings have also been shared with Barbara Befani and with the Quality Assurance Team.

Table 5.1. Overview of quality rating

Rating interval	Description	Number of evaluations	Percentage of total
26 – 30	Excellent in respect of several quality aspects and very good in the others (Very good)	2	6%
21 – 25	Good in respect of many quality aspects, and satisfactory in respect of others (Good)	12	40%
16 - 20	Adequate, but not quite satisfactory, some shortcomings (neither good nor bad)	9	30%
11 – 15	Significant shortcomings in some respects and some shortcomings in most of the other (Poor)	8	24%
6 – 10	Significant shortcomings in all aspects of quality (Very poor)	0	0%

Source: The authors' assessment of 30 evaluation reports.

How 'good' is this outcome? Are there quality problems among the CSO evaluations? The answer to the first question is that it is quite good. Although there are few quality studies of aid evaluation, the few that exist present a similar picture, or one where the overall quality is lower⁴³. The rating suggests that the evaluation reports commissioned by these six CSOs are as good as the evaluations of many other actors in development cooperation; be they bilateral actors such as Sida, Norad, DFID; multilateral actors such as UNESCO, UNHabitat, UNDP and financial institutions; and of course, other CSOs. We did not find any evaluations that were really poor, and though some were below the middle rank, these were not necessarily totally wasted. There could be reasons for the low score, such as a very tight budget and a very narrow focus. Table 5.2 presents a closer look at where the problems lie – and of course also where the most significant strengths are. However, the response to the questions is based on the evidence from these 30 evaluations. While we have made a careful judgemental selection of reports to reflect different themes, activities and organisational origins, we cannot know for sure whether they are representative of the total number of evaluations.

The greatest strengths found in the evaluation reports lie in the evaluation teams' competences and understanding of the subjects they are evaluating. They do understand rural development issues, humanitarian assistance, organisational development, and all the other more specific development or humanitarian problems that the projects/programmes are trying to address. They know what they are looking at, they are professionals in the same fields – and that shows clearly in the reports. The evaluations are strongest in respect of

43 See for example Sida Studies in Evaluation 2008:1; Stern et al (2012) Broadening the Range of Methods for Impact Evaluation. DFID Working Paper 38, 2012.

Table 5.2 Assessment of the five dimensions of the holistic assessment

Quality	Score 1	Score 2	Score 3	Score 4	Score 5	Score 6	Sum
Methodology		4	11	10	4	1	107
Presentation		5	7	8	7	2	114
Understanding			2	8	13	6	144
Innovation		10	10	8	2		93
Depth and quality		2	8	5	10	5	128

Source: The authors' assessment of 30 evaluation reports.

their description of the interventions, the substance of activities, and their insights into the nature of work done.

While 'understanding' is judged to be very good, Table 5.2 also shows that the 'depth and quality of information' is also assessed as good – but not quite as good. Perhaps that is because the evaluation teams have demonstrably high competence in their different fields of investigation, but due to lower skills in 'methodology' and 'presentation', the findings do not quite have the accuracy that might be expected. The evaluations are weakest in respect of 'creativity and innovation' in the evaluation process, and they are also weaker in respect of methodology. Particular methodological weaknesses identified include the structure of the reports and the way they use data and present their findings. Nevertheless, it is no more than 15% that fall below what one might call satisfactory standards of presentation.

What are then the most important strengths of the evaluations? One clear strength is the 'depth and quality of information', another 'the understanding of the interventions' within the core activity field. 100% and 94% respectively were judged as having satisfactory, or more than satisfactory, scores in relation to these two aspects of quality. When evaluation teams are chosen, the organisations primarily look for experts in the fields being evaluated. The authors of the reports are generally people with that kind of expertise. Seldom does one see experts in evaluation methods and methodology contracted to undertake evaluations (never, in respect of these 30 evaluations). Generally speaking, although innovation and methodological expertise are appreciated when it comes to deciding how the pounds and pennies in an evaluation budget will be spent, preference goes to tried and trusted competence and knowledge. Why are innovative methods and approaches not found in this sample of reports? One answer might be that because the evaluations cover significant fields of cooperation and because they were awarded to evaluation teams after competitive bidding (or some similar but less formal selection process), there were high expectations on useful findings for decision-support. This combination of attributes probably does not lend itself to risk-taking and innovation in respect of evaluation methodology, neither from those who commission the evaluations nor for those who do them.

Assessment of Utility

Utility is the first mentioned of the quality standards and that's not by chance. The Joint Committee carefully explains that utility is the single most important quality. In this particular exercise where we have read the reports, we cannot say for certain whether the evaluations were used or not⁴⁴. We limit ourselves to discuss how potentially useful they, that is, to what extent they lend themselves to being used. To do that, an evaluation ought to respond to the demand for information such as that contained in the terms of reference, its conclusions should be clear and consistent, etc. We have translated these qualities into questions and rated them. The results are presented in Table 5.3.

The vast majority of the evaluations address well the questions listed in their terms of reference and they provide clear and consistent conclusions and recommendations that follow from the analysis. There are practical conclusions, directed to specific actors/decision-makers. There are some few exceptions to this overall assessment; 10% were given scores of 2 or 3 on these ratings, which is on the lower half of the scale. In contrast, around 60% of the evaluation scored 4 and 5 on the scale, and some 5 – 15% were found to be excellent in these respects. None were found in the lowest category.

The last three questions in Table 5.3 shift the focus away from recommendations to learning. The scores given here suggest that the evaluations contain less food for thought than they do for action and decision-making. More than half of the reports do not have any clearly marked specific section on lessons learned; two don't have any sections on lessons learned at all, the other 14 have one, but these are not easily found and contain little substance. Even though a patient, well intentioned and informed reader may, through digging, discover some useful lessons learnt, it would be hard work to find these and the things found would often not be couched in concrete terms and thus be difficult to operationalise to provide a better general understanding of development cooperation and improved effectiveness. The different colours of the cells in the table show where the distribution of answers is centred, and it is evident how this shifts from higher to a more even spread as the focus shifts from recommendations to learning.

Evaluation is in fact a process, and it does not necessarily end with the production of a report. Still, evaluations are often taken as synonymous to reports and much attention goes into reporting. The reason is that the written report both communicates easily, across time and space, and it is in theory also a transparent instrument for providing the evaluative judgement. The quality of the written report means much for the quality of the process as a whole, and a well-written report makes a big difference for utility. Table 5.4 sums up the criteria used to assess the utility of the reports and presents the scores given for the 30 evaluations.

44 However, understanding the extent to which and precisely how evaluations have been used (for instance the discussion of whether there is formal management response to evaluation recommendations and the relationship of evaluation to the Boards) is discussed in other parts of the report and was discussed in the interviews etc.

Table 5.3. The usefulness of evaluation reports

Indicators of utility	1. No, very poorly done	2	3	4	5	6. Yes, very well done	Not done	Total
Does the evaluation respond to the questions in the terms of reference?	-	1	3	8	13	5	-	30
Are the conclusions in the evaluation clear and consistent?	-		6	7	11	6	-	30
Do the recommendations follow from the analysis and conclusions?	-	2	3	8	12	5	-	30
Are the recommendations practical, can they be translated into decisions?	-	1	4	12	8	5	-	30
Are there recommendations for clearly specified groups of actors?	-	2	3	14	10	1	-	30
Are they relevant and for an informed audience interesting lessons learned in a specific section?	-	5	9	10	3	1	2	30
Can an informed reader identify and make sense of lessons learned through the intervention?	-	2	7	18	2	1	-	30
Has the evaluation added to a general understanding of development cooperation?	-	10	9	7	4	-	-	30

Source: The authors' assessment of 30 evaluation reports.

The table shows that the evaluations are generally found to be in the upper half of the rating scale. There are in most cases (around 60%) both adequate executive summaries and a clear logic to the reports, well-structured annexes, good language standards, frankness in addressing issues and evidence of an impartial style of writing. Only 15% of the reports did not appear to be frank, and only 12% gave the impression of not being impartial. The weaknesses in presentation primarily relate to the use of information in the reports.

Table 5.4. Overall Assessment of Structure and Style

Indicators of structure and style	1. No, very poorly done	2	3	4	5	6. Yes, very well done	Not done	Total
Is there a clear and adequate executive summary?		1	4	8	13	2	2	30
Is there a clear and logical structure to the chapters and to the report as a whole?		3	5	7	13	2		30
Are illustrations and figures used to facilitate reading and understanding?	1	2	10	5	3	-	9	30
Are tables, boxes and models well designed, clear and accurate?		4	4	7	7	1	7	30
Does the report make use of references and is it appropriately referenced?	1	8	5	2	4	-	10	30
Are annexes well-structured and readable?	-	-	7	12	7	-	4	30
Is the report free from grammatical and spelling errors?	-	2	4	9	13	2	-	30
Is the language of the report precise, varied and interesting, free from jargon?	-	4	7	8	10	1	-	30
Is the report frank, does it address issues squarely and straight on?	-	1	4	9	10	6	-	30
Is the report written impartially and does it apply different perspectives to issues treated?	-	2	2	11	13	2	-	30

Source: The authors' assessment of 30 evaluation reports.

Most of them rely solely on written text to present data, convey the analysis and present recommendations; no more than 20 – 30% make use of tables, boxes, diagrams, illustrations and figures to assist understanding, make reading easier and assist in communicating the core messages. The pedagogical skills evident in the evaluations in building a narrative and communicating effectively based on a written report could be better developed and would probably make the reports significantly more useful. It is also rare for the reports to make reference to their use of sources of information beyond the project/programme being examined – assuming that if the authors had made use of other studies, evaluations, research findings, feasibility studies, etc. they would have referred to them.

Assessment of Propriety

The word 'propriety' is used in this context to capture and encompass those aspects of the evaluation process that deal with the ethical issues of evaluation. Has the evaluation been conducted with due regard to the agreement, wellbeing and basic rights of those involved, as well as those affected by its results?

The first issue to assess is whether evaluations build on agreements with all parties and stakeholders. Those who are the subjects of the evaluation and who come to be involved in the process should have been consulted beforehand and should have had the opportunity to decline to participate. We have not found examples of any formal agreements being made – apart from those between the evaluation teams and those commissioning the evaluations – and hence it is difficult to say, simply from reading these reports, whether any of these ethical norms were either ignored or violated. The evaluation reports do not recount any stories of where they were denied access, encountered difficulties, or met with hostility or suspicion. This could signify that the organisations have such close contacts with communities and target beneficiaries that such interactions were managed by the organisations rather than by the evaluation teams. Or it could signify that the issues were not addressed. Or – actually not unlikely in practice – that if evaluators only spent a short time with the beneficiaries they were “too polite” to tell them if anything was amiss!

Another issue is whether the evaluations were designed and conducted with due respect and protected the rights and well-being of the direct project beneficiaries. The evaluations were often participatory⁴⁵ and they appear to have been conducted in an appropriate manner. No evidence was found in the reports to suggest that the process could have harmed stakeholders, interviewees, or others. Whether the evaluators actually respect human Dignity and interact in such a way that people don't feel their interests or Dignity has been threatened or harmed is a question that it is difficult to uncover from the reports. NorCross and NPA have policies that are quite clear and explicit on these issues, and those should go some way towards ensuring the evaluation teams to act in an appropriate (ethical) manner. The other organisations might benefit from making ethical concerns more explicit when their policies are developed.

45 While the evaluation reports claim to be conducted with participatory methods, the actual meaning of this varies – from a minimum of consultations through participatory workshops, to a full conduct of the evaluation together with partner organisations.

Conflicts of interest should be dealt with openly, and as we have seen, the evaluation reports do address a number of critical issues frankly and many times these are dealt with in an impartial manner. This is particularly visible in the evaluations of partnerships where the organisational processes are analysed. It is also seen in evaluations of goal achievement, where stakeholders have had different opinions on the extent of achievements and explanatory factors. Impartial assessment in such situations means to show that there are differences of the opinion – but also to conclude based on the evaluation findings. Again, to the extent issues relating to conflict of interest have arisen, they are not visible in the reports and did not usually come up during the interviews. Some interviewees mentioned that there were evaluation reports that were disappointing, or not up to the expected standards. That may also reflect a partial position in the face of criticism raised in the evaluations. It shows that a particular assessment of, for example, goal achievement, was contested. But these were some few rare exceptions among the interviews the evaluation team undertook.

Table 5.5 shows the assessments of consultations in the evaluation process. Other studies have identified the lack of consultations with groups that ought to have been involved as probably the most important ethical issue in evaluations of development cooperation⁴⁶. The number of interventions to be studied, the number of site visits and observation cases, and the number of respondents to speak to – all tend to be much lower than required to avoid bias and to ensure a fair representation for ethical reasons. Another issue is whether the evaluators sought information from those other than the targeted beneficiaries of the intervention. Such groups could have a very different view of the intervention and it is an ethical issue whether their voices are heard. As the reports all failed to mention such discussions, it seems that none of these 30 evaluations appeared to have consulted potential rather than actual beneficiaries of the interventions.

The patterns of consultations have to a large extent been appropriate for the task at hand, although there seem to have been a relatively larger number of occasions when we have noted that consultations were not carried out. That may well reflect an appropriate choice by the evaluation teams rather than a quality problem. A few of the studies were internal, organisational assessments, and so in these cases the question of whether host country authorities were consulted would most likely to have been irrelevant. There was one desk study, and given the chosen method used it was obviously not possible to interact with stakeholders or beneficiaries. Comparing these 30 evaluations to broader patterns of consultations in development cooperation, it appears that consultative practices are more widely applied and better implemented than in other contexts.

46 Forss, K. (2012) Utvärdering av utvärderingar en analys av kvaliteten hos ett urval rapporter från SADEV. Commissioned by the Ministry for Foreign Affairs. Stockholm.

Table 5.5. Overall Assessment of Consultations

Indicators of consultation	1. No, very poorly done	2	3	4	5	6. Yes, very well done	Not done	Total
Is the report frank, does it address issues squarely and straight on?	-	1	4	9	10	6	-	30
Did the evaluators consult relevant authorities in the host country?	-	2	2	8	13	-	5	30
Did the evaluators gather data from end-users or beneficiaries of the project?	-	-	2	5	16	-	7	30
Did the evaluators gather data from the implementing organisations	-	-	-	6	20	3	1	30
Did the evaluators involve any stakeholder groups in the evaluation process?	-	1	1	4	18	2	4	30
Did the evaluators interact with a reference group	-	-	-	-	9	5	16	30

Source: The authors' assessment of 30 evaluation reports.

Assessment of Feasibility

The feasibility standards are intended to ensure that an evaluation will be realistic, feasible and frugal. Starting with the latter, we have tried to assess whether the evaluation resources have been used efficiently and effectively. The point is that the evaluation should use its resources in the best possible way: not spend more time on field visits, surveys, interviews etc. than necessary, nor waste money on consultations that do not benefit the process. Although this is hard to estimate, the overall assessment suggests that the evaluations have done well in this area. There were no instances where we could see that too much data was being gathered, or where the analysis took longer than could have been expected. Most of the reports have a time schedule in annex form, and based on this one can conclude that the processes were efficient.

A proper assessment of whether evaluation resources were well used ought to be based on examining the complete evaluation budgets. Unfortunately for these 30 evaluations there was no detailed financial information easily available, and, going forward, that is a problem that needs to be addressed. The assessment is thus an estimate of whether time has been used efficiently. As the budget of an evaluation is primarily used to buy time (fees being the largest share of costs), it is possible to have an approximation of the efficient use of money as well.

Effectiveness is an assessment of whether the evaluation objectives were achieved, that is, whether the evaluations responded to the terms of reference. In 4 out of the total of 30 reports, we judged that the responses to the terms of references were not as good as might be expected (see table 5.6). The large majority responded well to the questions in the terms of reference. However, before judging conclusively that the four evaluations did indeed score poorly in terms of effectiveness it would be necessary to analyse not just whether the evaluation objectives given in the TOR were achieved, but also whether the objectives given were realistic; they might have been misconceived, or have been unachievable in relation to the evaluation budgets. In such cases, the problem lies with those who commissioned the evaluation rather than with the evaluation team.

Terms of reference and the transformation of these into practical questions that can guide the evaluation process are assessed in table 5.6. There were two instances when the terms of reference could not be assessed, as they were neither reproduced as annexes in the report nor explained or summarised by the evaluation team in the body of their report. That's obviously something that should never happen – not primarily for the sake of an assessment of the study, but for ethical reasons, for the transparency of the whole evaluation process. The vast majority do have terms of reference that are clear and focused. The fact that the majority are rated 4 on the 6-graded scale is because there is tendency to ask too many questions and not to focus the evaluation sufficiently. It is not a big problem, but it is there, and it would be better if all terms of references could be rated at the top-end of the scale.

Table 5.6. Assessment of the feasibility of terms of reference and questions

Indicators of feasibility	1. No, very poorly done	2	3	4	5	6. Yes, very well done	Not done	Total
Are the terms of reference clear and focused?	-	-	1	18	8	1	2	30
Does the evaluation interpret and focus the task as defined in the terms of reference?	-	-	3	14	11	1	1	30
Is the basic question clearly stated in a specific section?	-	1	7	11	10	1	-	30

Source: The authors' assessment of 30 evaluation reports.

The scores given for the bottom two indicators in table 5.6 show that in some instances terms of reference that tend not to be so clear are sharpened by the evaluation teams. Transforming terms of reference to questions that can guide a practical process of inquiry is often a necessary part of an evaluation assignment and, as we can see here, 75% of the reports do that well and 60% demonstrate that that process has happened in an easily identifiable section of the report.

Assessment of Accuracy

The standards connected to accuracy are of two kinds. First there are the process standards that specify whether the conduct of the evaluation and the contents of the report are such that they follow good practice to the extent that they are able to guarantee that the findings will be accurate. These consist of three sub-questions: (a) Is there a separate section describing methodological choices, (b) is that section reasonably exhaustive, and (c) were the reasons for opting for the methods chosen well argued? In the discussion of these issues, it is particularly informative if the evaluation team 'warns' the readers of any potential and important information-gaps are likely to arise as a result of the approach they decide to adopt (nobody would know that better than the evaluation team!). Second, we raise the question of whether the assessments as such make sense. Irrespective of whether the reports contain a description and discussion of methods, do we as readers find the presentation robust and trustworthy? Do the data gathered and presented in the reports necessarily point us towards the conclusions drawn by the evaluators? Might other conclusions been drawn, or do they seem odd when the evidence is considered?

Table 5.7 presents the assessment of methodological choices. What sort of benchmark informs our judgement here? In our view, we do not think that evaluation reports should be written in the style of academic theses. We are not looking for methodological dissertations. However, this does not mean the bar should be set too low either. The standards we apply are based on the widespread consensus emerging out of the Joint Committee on Standards and other quality discussions⁴⁷. As do others, we also think that evaluation teams should document what they do and that evaluators should explain their methods. It is an essential feature of good practice.

As the ratings in the table show, most scores are clustered in the lower-to-middle range which in turn leads on to a discussion of whether the glass is half full or half empty. Thus, 50% of the evaluation reports do not have a satisfactory section that describes methods; 78% lack a discussion of threats to validity and reliability; 30% do not discuss limitations to the assignment. Consequently the transparency of the assessment process is weak and the reader will often have difficulties understanding how conclusions were reached and thus the logic that leads to the recommendations given. We could conclude that the instruments for data collection (such as interview guidelines, surveys, observation protocols,

⁴⁷ See for example the Program Evaluation Standards (1994), sections A1 – A11, pp. 127 and following.

Table 5.7. Assessment of methodological choices

Indicators of methodology, choice of methods and instruments for data collection and analysis	1. No very poorly done	2	3	4	5	6. Yes very well done	Not done	Total
<i>Description of Methodological Choices</i>								
Is there a section that describes the methodological choices fully?	-	2	9	8	7	-	4	30
Is there a discussion of threats to reliability and validity?	3	6	6	2	4	-	9	30
Is there a clear statement of limitations to the evaluation?	1	3	5	14	5	1	1	30
<i>Data Collection Methods</i>								
Are the data collection methods chosen appropriate to answer the evaluation questions?	-	2	4	10	13	1	-	30
Is there a relevant and adequate selection of sources of data?	-	1	8	10	10	1	-	30
Does the choice of methods suggest that the evaluation will get reliable and valid data?	-	2	7	9	11	1	-	30
<i>Instruments for Data Collection and Analysis</i>								
Are the instruments for data collection well constructed?	-	2	4	6	3	1	14	30
Are indicators appropriate?	-	1	1	12	8	1	7	30
Are benchmarks fair and relevant?	-	-	-	5	4	1	20	30
Are rating scales well designed?	-	-	-	-	4	-	26	30

Source: The authors' assessment of 30 evaluation reports

focus group agendas, etc.) were satisfactory or better in 30% of the reports, but in 47% of them the instruments could not be assessed because they were nowhere to be found.

In our view the core problem lies in the area of presentation that could easily be addressed. With a trained eye it is still possible to analyse the methodological choices and the methods (if not the instruments) used. In more than 60% of the cases where the methods were satisfactory or well chosen, there were at least focus group agendas, etc.) were satisfactory or better in 30% of the reports, but in 47% of them the instruments could not be assessed because they were nowhere to be found. Adequate sources of data collection, and in 63% of the cases the choice of methods seemed to be robust enough to ensure that the evaluation would be able to obtain or gather valid data.

But it should be noted that the methods and the instruments are often qualitative and build on participatory approaches. Individual interviews, focus groups, and observation dominate as sources of data, supplemented by document analysis – which is taken to mean analysis of plans, progress reports, feasibility studies, and sometimes monitoring data. While it is common that the methods are structured through the use of indicators (63%), it is not so common to use rating scales or benchmarks as the basis for the evaluative judgment made.

The analysis of implementation suggests that when the evaluation teams address an issue, they mostly do this well and present a trustworthy analysis (a mean rate of between 4 and 5 on the 6-graded scale). However, there are also some aspects of implementation that are frequently not analysed, in particular financial management and to a lesser extent leadership and governance. This should not be seen as a quality problem in any specific evaluation – the chances are that the terms of reference did not ask for an analysis of financial management or leadership. However, when we look at as many as 30 evaluations that make up a substantial share of the evaluations commissioned by these 6 CSOs, it becomes obvious that there is a serious lack of information in their evaluation systems. The reports do not generate the learning inputs or the decision support that they could be expected to do.

Other aspects of implementation are taken care of much better. The findings in respect of planning, organisational structures, networks and linkages and coordination are generally good and in some cases excellent. The evaluation teams present credible analysis, supported by data. Several of the evaluations that are rated highly have a clear focus on implementation issues. There are examples of evaluations of NorCross partner organisations, partnerships with churches, or with politically affiliated organisations.

Table 5.8. Assessment of implementation

Indicators of implementation	1. No very poorly done	2	3	4	5	6. Yes, very well done	NA	Total
Is there a trustworthy analysis of leadership and governance?	-	1	1	9	7	2	10	30
Is there a trustworthy analysis of planning?	-	1	-	11	9	3	6	30
Is there a trustworthy analysis of financial management?	-	-	1	3	6	1	19	30
Is there a trustworthy analysis of coordination?	-	1	1	13	9	2	4	30
Is there a trustworthy analysis of networks and linkages?	-	1	2	12	12	1	2	30
Is there a trustworthy analysis of organisational structures?	-	2	1	13	9	2	3	30

Source: The authors' assessment of 30 evaluation reports

The assessment of results is more comprehensive, but here too there is one aspect that is seldom analysed: efficiency. Efficiency is not analysed at all in 36% of the evaluations and it is only done in a satisfactory manner in 39% of the reports. Again, there are probably good reasons why the evaluation teams did not examine the issue of efficiency in particular cases, but it becomes a problem at the systems level when one key aspect of results seems to be left out so frequently. Another important issue that received comparatively less attention and where there are fewer trustworthy results is impact. It is addressed satisfactorily in less than 50% of the reports.

There is a cluster of problems around impact assessment: the evaluations do not consider design issues. If they discuss methods, they usually focus on the choice of methods: participation, case studies, and the balance between instruments for data collection. They do not consider design for impact assessment and they do not discuss and comment on the different possible approaches one could take to analyse impact – be they experimental approaches, theory-based, etc. The evaluations do not have any approach to study causal linkages, and when that's lacking it becomes quite difficult to address impact in a trustworthy manner. They don't discuss how they judge the causal links (between inputs, outputs and outcomes); and thus the reports do not have the data and information with which to derive firm conclusions; they

don't discuss the fact that this is a (big) problem for coming to conclusions about impact. The CSO evaluations are not unique, this is a well-known problem and one would hardly expect the CSOs to lead the way in this rather difficult field of methodological development.

Nonetheless, it must be considered quite satisfactory that there are trustworthy conclusions on effectiveness and relevance in 72% and 78% of the evaluations respectively. Comparisons with other studies of evaluation systems show the same pattern⁴⁸; these evaluations too focus on relevance and effectiveness, there is less information on impact and sustainability, and least of all on efficiency. It is rather surprising, because efficiency is in many ways the easiest aspect to assess. Efficiency is essentially a ratio between inputs and outputs and it does not require any causal explanations or any estimates of future events or extensive contextual analysis (as do sustainability and relevance). It does require a benchmark though, and that may cause difficulties. The methods used to analyse and derive conclusions on efficiency are fairly rudimentary and not at all as difficult and complex to assess as the other four issues.

Table 5.9 Assessment of results

Indicators of results*	1. No very poorly done	2	3	4	5	6. Yes, very well done	Not done	Total
Is there a trustworthy assessment of efficiency?	-	3	2	9	4	-	12	30
Is there a trustworthy assessment of effectiveness?	-	1	2	12	12	1	2	30
Is there a trustworthy assessment of impact?	1	3	8	10	4	-	4	30
Is there a trustworthy assessment of sustainability?	-	3	3	13	5	1	5	30
Is there a trustworthy assessment of relevance?	-	1	-	10	12	4	3	30

* for definitions of efficiency, effectiveness, impact, sustainability and relevance, see the OECD/DAC (2010) Evaluating Development Co-operation: Summary of key norms and standards
Source: The authors' assessment of 30 evaluation reports

48 Apart from the comparative studies mentioned above, see also Forss, K. (2012) *Aggregation and Analysis of Results, Recommendations, and Lessons Learned from Sida's Strategic Evaluations*. Commissioned by Sida, Stockholm.

Conclusions

This chapter has focused on the evaluation part of M&E for two reasons. There are quality standards for evaluations that can readily be applied. There are easily defined reports that can be picked up, read and analysed. To comment on the quality of monitoring data would be an entirely different process; a process that we haven't developed for this assignment. However, as more money is spent on producing monitoring data, the quality assurance may become an issue in the future. A future study of quality might need to focus on monitoring data rather than on evaluations. In many ways, the results of this assessment mirror the views of those in the six CSOs about the quality of evaluations. As chapter 4 shows, the responses there indicate that they have a good sense for the quality of evaluation reports. The results of the survey indicate that 74% disagree with the statement 'evaluations are not trustworthy'; 56% agree that 'evaluations are of high quality and reliable'; and 56% agree that there is 'a need for quality control'. The review indicates that the majority of evaluation are of high standards and are indeed trustworthy, but there are exceptions to the rule. In particular, the methodological aspects need attention.

These six CSOs' experiences with evaluation are not very different from most other agencies working in development cooperation, though, if anything, they appear to do slightly better⁴⁹. They experience widespread difficulties in assessing impact and pay less attention to efficiency – as do others. Likewise, when it comes to assessing implementation the same strengths and weaknesses in relation to coordination, networks and structures prevail. However, other evaluation systems would appear to have a stronger focus on the explicit choices of methodology, methods and instruments. The problem here is not that the evaluation teams make uninformed or poor choices, but rather that they don't document what these choices were or why and how they have been made.

Quality can be summarised in relation to the four criteria of utility, propriety, feasibility and accuracy. While the evaluations as a whole do score well in all four areas, there are of course problems with some evaluations and there are differences between the four criteria. The ethical aspects of the evaluation process need more systematic attention. The evaluations have extensive and relevant consultations with stakeholders, beneficiary groups and partner organisations. The feasibility aspects are mostly well taken care off. Still the absence of basic financial information on the evaluation process remains a persistent problem, even if this lack of data is not the fault of the evaluation teams. What can be derived from the reports themselves suggests that the evaluation teams work efficiently and effectively. The terms of reference tend to be focused, even though there is scope for improvement. The accuracy criteria appear to be the weakest of the four dimensions and where systems development needs to focus. Evaluation utility is the foremost quality criterion, and the large majority of evaluations respond well to this component of the terms

49 Other studies of evaluation quality referred to above; Stern, E. et al (2012); Schwartz, R. And Mayne, J. (Ed) (2005), Forss, K. et al (2008), Indevelop (2012) indicate that the quality of evaluation reports are about the same, while some point to other and larger problems. The studies can of course not be compared as such, as the samples of reports, the criteria for assessment, etc. differ.

of reference and have clear recommendations to specific actors; what's more, the recommendations are for the most part derived from the evaluation process. People in the organisation also have a high opinion on the utility of evaluations, so the information coming from the survey and from the interviews support the findings in respect of this sample of evaluation reports. In the next chapter we turn to a closer discussion of what it means to use evaluations, in particular for purposes of learning.

6. Evaluation; Use and Learning

Introduction

The question of whether evaluations are used or not, and if so, for what reason, has been discussed with vigour in the evaluation community for many years. The appetite for self-reflection and criticism is high. Christie (2007, p. 8) notes, "Evaluation utilization is arguably the most researched area of evaluation and it also receives substantial attention in the theoretical literature."

In this section we seek to answer the questions whether monitoring and evaluation systems are used and connect this to the different types of use that are mentioned in the literature on evaluation. It is important to link use with the previous chapter on evaluation quality. Not all evaluations lend themselves to use, they may not be useful and they may not have the characteristics of accuracy and propriety that would be required. The review in chapter 5 indicates that as much as one evaluation in four might fall below the required quality standards. We should qualify this, and also make the analysis more rigorous. While many evaluations are useful, accurate, etc. from beginning to end, some are partly useful and accurate, but there may be sections and parts where the evidence is meagre or the methods have not been adequate.

As a starting point, the previous chapters have shown:

- 100% of the respondents to the survey agree or strongly agree that evaluation is a good learning tool.
- The same 100% agree or strongly agree that integrating evaluation activities enhances learning.
- 80% of survey respondents agree that their organisation makes good use of evaluations and the answers to other survey questions confirm that initially high view on use.
- The quality assessment found that around 70 – 75% of evaluations rated above what would be considered satisfactory on the different aspects of utility.
- That still means 25 – 30% fall below that benchmark, and hence any utilization of these evaluations must be considered very carefully, and it would be expected that the use of such information in decision-making would be supplemented by other sources of information.
- The interviews gave a slightly less positive view of the use of evaluations.
- 74% of the respondents to the survey could cite and provide examples of an evaluation that they found useful. The examples from the survey show almost equally many evaluations although there are some few that are mentioned several times (Rewriting the Future, from Save the children, for example).

These indications of use, patterns of use and connection to learning show that there are many questions to answer and it is not an easy subject to delve into as it concerns so many other organisational processes. We would like to remind the reader that we are only studying the six CSOs, and even with them we have a very clear focus on headquarters. Many if not most decisions are taken at the field level and in negotiations and consensus-building processes with partner organisations.

While both monitoring and evaluation contain 'evaluative information', the scope for learning differs. The process of evaluation incorporates, by definition, the processing of information to arrive at conclusions and recommendations. An evaluation contains processed information, ready to be used and also ready to be debated, reflected upon, and to learn from. Monitoring, on the other hand, is essentially a process of collecting raw data. The analysis, the process of arriving at conclusions, is basically not part of the activity of monitoring itself. That being said, it is of course obvious that a stream of information on performance will trigger reactions and action. The study of use and learning from monitoring would need to be broader and fully integrated with a study of management and decision-making, that is, a more comprehensive approach than this study allows.

Use of Evaluations

As indicated above, there have been many studies of the phenomenon of use in evaluation. The synthesizing work of Kirkhart (2000) and Weiss (1998) draws attention to the diversity of use, its unpredictability and its political nature. Box 6.1 presents a typology of use, drawn from Saunders (2012). The aim for all who write on the subject is to increase use, but many are equivocal about factors that could do so. Kirkhart and Weiss both emphasise the diverse ways in which evaluations are used rather than single out the conditions under which positive use seems to occur. Patton (1997) goes further in championing use. He asserts that the potential for use is largely determined a long time before a study is completed, thus pointing to the importance of design for use – forcefully stated as 'intended use by intended users'. Consequently his work, centred on 'Utilization-focused evaluation' takes a strong stand on the possibility to explicitly design systems that produce that intended use.

Apart from studies of the use of evaluation there is a much broader area of research on the use of knowledge generally, and that includes research, evaluations, public inquiries, journalistic research, and expert knowledge wherever that is to be found (Boswell, 2009). The implications of adopting such an approach is that it is important to look for actual practices and to identify patterns and context. Lederman (2011: 160) concludes her research on use '... that it is time to abandon the ambition of finding 'the important' characteristics for use and to adopt a focus on context-bound mechanisms of use instead'. There are good reasons to be rather careful when concluding on the use of M&E systems in the six organisations studied here, which the reader should bear in mind when the conclusions in respect of use are summed up in Table 6.1.

Box 6.1 Typologies of Use

1. **Instrumental:** when decision makers use the evaluation findings to modify the object of the evaluation in some way.
2. **Conceptual:** when the evaluation findings help the programme or policy makers understand the programme in a new way.
3. **Enlightenment:** when the evaluation findings add knowledge to the field and thus may be used by anyone, not just those involved with the programme or evaluation of the programme.
4. **Process use:** cognitive, behavioural, programme and organisational change resulting from engagement in the evaluation process and learning to think evaluatively.
5. **Persuasive or symbolic (justificatory):** when important stakeholders are persuaded that the programme or organisation values accountability or when an evaluator is hired to evaluate a programme to legitimize a decision that has already been made prior to the commissioning of the evaluation.

Source: Saunders, M. (2012) The use and usability of evaluation outputs. Evaluation Vol 18, No 4.

The use of evaluation findings is thus rather high, with the emphasis being on instrumental use and process use. This is supported by the structures and processes in the organisation, such as those that were described in chapter 3 and by the evaluation culture portrayed in chapter 4. Interestingly, but perhaps not surprisingly, these kinds of use are congruent with the qualities of the evaluation reports that we looked at in chapter 5. These evaluations – with their strengths and weaknesses – lend themselves well to instrumental use and the implementation of the evaluations have many of the characteristics that enable and encourage process use.

The evaluation policies however, do not seem to reflect this nexus of mutually reinforcing culture and practice. When it comes to describing the purpose of evaluation, that is, 'intended use', the policies emphasise learning. Amongst the three usually articulated purposes of evaluation – accountability, decision support and learning – the rhetoric prioritises learning. Some policies explicitly identify learning as the main purpose of evaluation, others mention or give prominence to these three purposes (or some close variant) and thus implicitly give equal weight to all three purposes – while in practice the whole system is heavily geared towards instrumental use and for practical decision support in the management of interventions.

Table 6.1 The extent of different types of use of evaluations

Type of use	Qualitative indications and motivation
Instrumental	High , as evaluations are often undertaken with the explicit aim to feed into mid-term reviews or otherwise on-going partnerships and projects. Evaluations contain concrete and practical recommendations and the evaluation process itself has legitimacy and credibility. There's a strong culture of expecting and preparing to use evaluation findings. The evaluations are initiated and implemented close to the intended users.
Conceptual	Medium , as evaluations focus on providing inputs for decision-making on current interventions, the emphasis is rather on understanding the intervention as it is, rather than on understanding it in a new way. The choice of evaluation teams and the design of evaluation questions do not generally work in favour of conceptual use. That being said, there are examples of evaluations that have conceptual use, and these evaluations tend to be well-known in the organisations, and are often read and quoted.
Enlightenment	Medium – Low , primarily because evaluations are not widely disseminated, the reports are not made public and the institutional memory is almost non-existent. Few evaluations have the general references and the broad perspectives to provide fundamental insights that go beyond specific interventions.
Process use	Medium – High , as programme coordinators and advisers – together with partner organisations – are closely involved in all aspects of the evaluation process. They are also managing the monitoring information and are the persons who integrate the use of monitoring information in evaluation.
Persuasive or symbolic	Low , the evaluation culture such as that emerges out of the survey emphasises the concrete and practical use of findings to support decisions. There is also a strong emphasis on the value of learning, and also on the independence and integrity of evaluation teams. Furthermore, external stakeholders do not seem to be much concerned with what the evaluations say and they do not take part of the evaluative information coming from the organisations – hence there is no such symbolic use of findings.

Source: Synthesis of the interviews in the six CSOs.

Learning as Changes in Knowledge Structures

Part of the problem may reflect how one defines learning. As usual, there's a common sense understanding of the word and different theoretical constructs. As much of the literature on organisational as well as individual learning says, the key to learning is to affect knowledge structures⁵⁰. That raises the question of how knowledge structures change. For one thing, that depends on how well developed they are (meaning rich, diversified, advanced). Piaget (1976) pointed out that well-developed knowledge structures facilitate change, and then change occurs by assimilating new knowledge and further refining the already existing knowledge structures.

Less well-developed knowledge structures may have difficulties absorbing new knowledge without drastic reorientation. It is then common to speak of accommodation. In plain terms, the more you know the easier is it to learn more – and to do so without painfully drastic changes of knowledge structures. The key terms here are seen in the distinction between the two kinds of learning;

- **assimilation**, processes that **gradually and incrementally** change existing knowledge structures; and
- **accommodation**, processes that **drastically and fundamentally** change existing knowledge structures, for example through paradigm shifts.

Returning then to the question of how well-developed knowledge structures are, it would be presumptuous for us to attempt to form overall judgements about knowledge levels in the six CSOs that are the subject of our study. Still we can dare to make some substantive comments based on the survey information that provides some additional data and information even though we need to be cautious how we use this as we don't know how representative it is.

The respondents have been in the organisation on the average 7 years. Most have taken part in monitoring and evaluation activities and most have had training in both. A large majority worked with other CSO organisations before joining the one they work with now. Thus our respondents comprise a group of persons who could be expected to have quite well-developed knowledge structures on the subjects that the evaluations deal with – cooperation for long-term development, institutional change, as well as humanitarian action. This would suggest that learning occurs mainly through assimilation. Learning takes place, but the new information is incorporated into the old knowledge structures and leads to their further development and refinement.

The **depth and quality of information** in the majority of evaluations was found to be high and the conclusions were trustworthy and at the same time the **substantive understanding** was also high. These were the two outstanding qualities in the overall summary of the evaluation reports. If these characteristics accurately reflect reality, it is clear that the information contained in evaluations

⁵⁰ It would probably take this study too far to go into the theory of learning and expound on knowledge structures, but valuable sources of information are found in Kelly (1955), Piaget, (1976), Perris, (1986), Argyris and Schön, (1978), and Schein (1985).

lends itself to an instrumental use of the evaluation findings and to learning through assimilation. The evaluations were not characterised by innovation and creativity, and it is also clear that the focus of the evaluation findings tend to concern only some aspects of implementation and some aspects of results (mainly relevance and effectiveness). Hence, the evaluations do not seem to lend themselves to conceptual use or for enlightenment (see table 6.1), and they do not contain the kind of perspectives and insights that could trigger learning through accommodation.

Unfortunately there is no policy guidance on the type of learning expected from evaluations and one cannot judge whether the frequent reference to the need for learning might conceal a more fundamental deeper question as to whether a more rapid change of knowledge structures would be required. One theoretical perspective could throw light on the question. Thus, Argyris and Schön (1978) distinguish between single-loop and double-loop learning: the difference between the two is that single-loop learning is supposed to take place within existing knowledge structures, developing and refining them; in contrast, double-loop learning implies that old knowledge structures have become redundant and that they no longer help us perceive accurately and interact in the world "in a meaningful way" (to put it drastically). Argyris and Schön also identify a third level of learning, which they call deuterio-learning. This means the capacity of learning how to learn, that is, how to systematically generate and make use of single- and double-loop learning.

In terms of learning, the strength of the present monitoring and evaluation systems lie in their capacity to generate information that feed into the decision-making situations, often by incrementally adding to what people need to know. A high degree of coherence between a perceived need for information, translated into a practical demand in the evaluation process, and finally supplied through the process and not least in the evaluation reports, builds a strong belief in the systems. Legitimacy and support follow suit, as the survey data shows.

Also in terms of learning, the weaknesses of the monitoring and evaluation systems are to be found in the partial assessment of issues, methodological weaknesses that become obstacles to evidence-based information on those aspects of implementation and results where knowledge structures are probably not as well developed. The information in the systems does not feed into double-loop learning to the same extent, even though there are some examples of such processes. The third level, of learning to learn, is not at all addressed through the monitoring and evaluation systems of the organisations, and we have not come across any debates, research processes, or organisational reviews that address such forms of learning.

Incentives and Obstacles

Debates on organisational learning face one fundamental difficulty and that is whether organisations can be said to possess knowledge structures in any form of empirically verifiable way. Much as that could be an interesting subject to discuss, it is probably just as well to leave it in the short term. Instead, one can

make some progress based on the assumption that it is actually individuals who learn, who possess knowledge structures that can be identified, changed, and subject to discussion. So, rather than focus on organisational learning one can focus on what organisations certainly can do – namely facilitate and encourage – or possibly hinder and discourage – learning by people in the organisation (and their partners, consultants, members and other stakeholders).

There are not that many formal incentives for the use of evaluations generally, or for learning. None of the interviews led us to highlight any particular examples so there does not appear to be any difference between the six organisations. Instead, the incentives must be constructed from the work situation in the organisations and it would be possible to infer that the following factors would be viewed as incentives;

- **sense of satisfaction** of a job well done when an evaluation process has been completed;
- **recognition** from partner organisations and those implementing an intervention when the evaluation process has come to an end and results in the kind of instrumental use that everybody would be comfortable with;
- **compliance** with rules and regulations and the monitoring and evaluation policy, which presumably feeds in to career development;
- **the intellectual rewards** that come out of a deeper understanding of the interventions, which is a personal, internalised reward system.

It would be possible to list a number of explicit and formal incentives that could encourage the use of evaluations generally and for learning purposes particularly. That could be provision of resources, assistance in commissioning and managing such evaluation tasks, policies that clarify the role of evaluation and how it is expected to lead to use and learning, connections to career development, recognition of outstanding examples as well as a high degree of tolerance, even encouragement, for risk taking in this field. Unfortunately it is much easier to point to features of the larger organisational system that in theory would constrain use and learning. Chief among these are;

- **lack of resources** in general and uncertainty about how to budget an evaluation process, where in most cases resources for evaluation would have to be obtained from different budget lines, where there are opportunities to use funds that might often be viewed as more immediately useful;
- **uncertainty** about what specific roles and responsibilities are needed in the evaluation process; the ultimate decisions on the wording of the Terms of Reference, choice of teams, interaction through reference groups and otherwise during the process, and finally the acceptance of a report, where the practice varies and depends much on the type of evaluation undertaken, the personalities involved, the partners, and the expected outcome of the evaluation;

- **the strength and capacity of technical support** for the evaluation process, which is the task of an evaluation unit in some of the six organisations, but not in all of them;
- **non-existent institutional memory**, that is, no person or office keeping track of the organisation's experiences with evaluation, the monitoring data available, the previous evaluations that can inspire or serve as warning examples, the processes to engage stakeholders and prepare for use, etc. None of the organisations have any structure or function that can be pointed at, which provides or facilitates their institutional memory.

There are a number of other factors that could also constrain the monitoring and evaluation function that we have no information on, but common examples from other studies would suggest these would include the following that could be investigated further: the impact of staff turnover, recruitment patterns and career development; policies on training and further education; and the degree of attention given to evaluation from the higher levels of the organisation (where, for example, the lack of interest often shown by the Boards would be, if not an obstacle, certainly a lack of incentives).

In sum, monitoring and evaluation are functions that are supported by an evaluation culture that has emerged over time and that identifies a number of positive attributes with evaluation. Based on that, there are also highly personalised and implicit incentives to undertake these tasks, but there are few explicit incentives. Clearly there are obstacles in the organisation to progress being made, but the question is how significant they are. The list above names a few, but while they are certainly real they can also be overcome – and have been overcome. We have emphasised that they are obstacles in theory, but even if resources have to be negotiated, roles and responsibilities clarified, these are work processes that are also accomplished and that don't seem to generate much anxiety. Hence, the somewhat startling conclusion would be that the organisations neither provide strong incentives for monitoring and evaluation, nor do they set up any particular obstacles.

Process Use of Monitoring and Evaluation

We now turn to what is called the process use of evaluation. This concept was introduced by Patton (1997) and is identified as the kind of usage that:

is indicated by individual changes in thinking and behaviour, and program or organisational changes in procedures and culture, that occur among those involved in evaluation as a result of learning that occurs during the evaluation process (1997:90)

In an effort to try to isolate this form of learning, Forss, Carlsson and Rebien (2002) tested the empirical support for the notion of process use, and tried to pinpoint what exactly it means by a close analysis of evaluation processes. This generated a new typology, which could in fact be a subset of the fourth cell in table 6.1. The starting point for that study was that when respondents were asked about the utility of an evaluation, they would often comment that the

results could have been anticipated, but it was useful to go through the process. The question is; just what do they mean by that? As the study found out, it could mean one or a combination of five different things and these are described in Box 6.2.

Whether process use becomes an important element of use in an organisational context depends on the design features. It is probably those who are active on evaluation teams that have the largest chances of realizing process use, in particular the forms of use that were called 'developing networks', 'learning to learn', and 'creating shared understandings'. That does not preclude that others could benefit. If, for example, an organisation appoints a team member, or a reference group is actively involved, or even when the level of interaction with those who commission the evaluation is strong, these forms of process use can extend beyond the evaluators. The process use called 'strengthening the intervention' benefits the management of the intervention, and in particular the beneficiaries – the intended target groups. The process use called 'Boosting moral' is also realised by those engaged in the evaluation process, but depending on the nature of the findings could also be quite important for the organisation being evaluated and/or those who commission the evaluation.

Box 6.2 Categories of Process Use

1. **Creating shared understandings;** the evaluation process creates and organises knowledge and the different actors who take part in the process share in this. The value of such shared understandings lies in its ability to foster consensus and facilitate joint decisions.
2. **Strengthening the intervention;** the evaluation process can contribute to the objective of the intervention in the interaction with managers, staff and potential beneficiaries, for example by reminding them of courses of action to pursue, focusing on a particular objective, etc.
3. **Developing professional networks;** those who take part in an evaluation meet large numbers of people and while that may not be of much use to an evaluation team, for those in the organisational context it could be quite useful.
4. **Boosting moral;** which probably depends on what the findings are, but even if the evaluation does not give the audience reason to celebrate, the fact of finding out and uncovering things you did not know may at the same time demystify problems and fortify the ambition to resolve them.
5. **Learning to learn;** to the extent that the evaluation process generates new information it is also a practical exercise in reasoning, applying logic and using data sources, and as such it can also develop those particular skills.

Source: Forss, K., Carlsson, J. and Rebien, C. (2002)

Process use does seem to be facilitated by the six organisations, although there are some differences between them. They all commission evaluations close to the management of interventions, there are usually reference groups tailor-made for the task, and it does happen that people in the organisation take part in the evaluative work. A very large number of respondents to the survey had themselves practical experiences of both monitoring work and evaluation (71% and 90% respectively). There are not any strong inhibitions to take part in evaluation; the survey shows that people believe they can and should get involved, and many think they should engage more in evaluation. It is also clear from the survey that people prefer monitoring and evaluation to be integrated with other organisational processes and that doing so connects to learning and use.

The first and the final categories of process use do imply learning. It would be difficult to see how shared and understandings and learning to learn would not happen through changes in knowledge structures. The other three kinds of use do not necessarily imply learning, though it has to be said that, for example, changes in morale might come from learning. It is a grey zone, and in practice it may well be that process use in many ways becomes synonymous to learning – but carefully specified learning processes.

It is possible to explicitly design the evaluation processes to increase the chances of process use, and in particular the choice of methods will have consequences for the outcomes hoped for or achieved. The evaluations that were studied in chapter 5 relied to a large extent on qualitative data, and interviews were the most common form of data collection. Interviewing is undertaken in the form of data collection that exposes most people to the process of inquiry, and hence it maximizes the likelihood of process use. On the other hand, synthesis studies undertaken by reading documents would offer the least engagement with people, while conducting a survey would not really be a form of data collection that provide high chances of process use, so if it is possible to choose interviews instead that would maximize process use.

Within the broad category of interviews, the particular choice of form would also have implications for process use. The evaluations encountered in chapter 5 often built on focus groups and the deliberative process fostered through a methodologically skilled interviewer would also increase the process use. Focus groups are different from the group interviews, and the latter would be less likely to generate process use. The individual interview format can be designed in different ways, and the more open format increases the chances of process use, while the more standardized and structured format decreases the likelihood of process use. In sum, the evaluations primarily choose open individual interviews and, following that, focus groups, as their preferred instruments for data collection and that would support process use.

Conclusion

The evidence presented in this and preceding chapters indicates that monitoring and evaluation processes are used and they generate or help to facilitate learning. The focus here lies on the use of evaluation, as a process and as reports. Monitoring leads to evaluative information, but neither this study nor any of the literature we have used, focuses on use and learning from monitoring data. As monitoring data are often aligned to evaluation processes, we can tentatively speak of use and learning from both, but the reader should be aware that there are more issues to discuss around use and learning from monitoring than this study brings up. In sum, the response to the question of use and learning is very much a question of whether the glass is half full or half empty. The monitoring and evaluation systems in the six organisations have in common a tendency to generate strong instrumental use and process use, but less of conceptual use and enlightenment, and the learning processes occur through assimilation of knowledge. Learning takes place in zones of comfort and lead to a gradual increase of knowledge. It is also strongly supported through the evaluation culture, and there are many practices in the organisation that support it. Still, evaluation use is a field that exists in relative isolation: there are few incentives but also relatively few obstacles.

There are small differences between the organisations, so for example does NPA have a stronger focus on participatory approaches to evaluation and these, in turn, generate higher levels of process use. There were evaluations in the SCN that generated the form of use that has been termed enlightenment. Other evaluations were geared to instrumental use. But the instances of such use cannot be related to contextual factors. There are no contextual features that could become obstacle if any of the other organisations wanted to initiate more participatory evaluation or the kind of studies that could lead to conceptual use or enlightenment.

7. Conclusions and Recommendations

This chapter provides a summary of the preceding discussion and analysis in relation to the three objectives of the study.

First Objective: Map the evaluation function in selected Norwegian CSOs

This was essentially a descriptive task and Chapters 3 and 4 describe the evaluation function notably in relation to policies, structure, processes and evaluation culture. Depending on the level of analysis, that description could be seen as one map or six different maps. At a high and aggregate level of analysis it is one map, particularly as far as evaluation culture is concerned. At a lower level of analysis different practices are followed within the different organisations, but these are differences of degree rather than differences in kind.

At that aggregate level, the map exhibits several common features. The policies tend to be descriptive and non-prescriptive. All six organisations have established units for M&E, usually within the international department, or a similarly named unit, two to three levels down in the hierarchy. These units often structurally combine planning and results with monitoring and evaluation; in some cases M&E is linked to strategic planning, and in one case ex ante evaluation is linked to internal audit.

Evaluations are usually initiated according to bottom-up processes, but occasionally according to requests from management or initiatives at central level, and in the case of SCN and NorCross from their partners in the international organisation. The budgets are decided in an ad hoc manner, based on estimates of the time needed. Larger evaluations are announced and tender proposals invited; for small evaluations, potential evaluators are usually approached directly to discuss the assignment. Evaluation teams are mostly made up of external consultants and experts, often with teams comprising some members who are Norwegian or international consultants and some who are from the host country. There are no formal systems to follow-up the evaluation recommendations, but the informal systems exist and often seem to work well.

The 'map' also points to differences. Three of the six organisations have explicit policy documents and the other have policy-like rules, regulations, guidelines and handbooks. Two organisations have formalised management response systems. Three have extensive internal personnel training in monitoring and evaluation. One has an extensive system to monitor outcome indicators. Two organisations have clear budgetary commitments to monitoring and evaluation.

Second Objective: Assessment

This objective was to assess the evaluation function of the organisations, regarding both the relevance and quality of reports and the system for transferring results, ensuring use of evaluations and learning.

Each of the six organisations has some form of results based programme management (RBM). The organisations have plans for evaluation and such plans form part of their agreements with Norad. The plans are implemented, evaluations commissioned and delivered according to the plans.

The organisations all have monitoring systems. There are two quite different forms of monitoring; two organisations have systems based on outcome indicators, the others have project-based indicators and formats for reporting. The systems focus on internal needs for information. External stakeholders get the same information and there are no systems set up purely for the purpose of external requests for information. Preliminary studies and needs assessments are carried out before an intervention is approved. Baselines are usually not established either for beneficiaries/target groups or for institutional development initiatives. The nature and quality of ex ante analysis is often commented on and criticised in ex post evaluations.

Most evaluations are formative and most are commissioned by the organisation itself, not jointly with others. The majority of evaluations analyse one project or programme, or one partnership. A few cover country programmes and more complex objectives. The focus in the evaluations is on relevance and effectiveness – when the implementation process is analysed, the focus is on planning, coordination, and networks. Financial management is seldom covered and the sample of 30 evaluations did not contain any analysis of corruption. Theories of change are seldom made explicit and rigorous methods of causal analysis are generally not used and thus issues of attribution not addressed in any depth. The evaluations do not make use of counterfactual analysis, and impact assessment is not prioritised. The contextual information is often rich and thorough and used to assess relevance. The most common method used for data collection is interviews, including focus groups. The methods tend to be participatory. Beneficiaries are relatively often included. Several of the evaluations focus on civil society capacity building. Quantitative indicators are rarely used and benchmarks are not made explicit. Unless a cross-cutting issue forms part of the TOR and is a particular focus of the intervention, it would not be addressed – except for gender equity issues, which are often covered.

The quality of the reports is generally good, but there are shortcomings, in particular in respect of explaining the methodology used but also in relation to adopting creative and innovative approaches in the evaluation process. Effectiveness and relevance are covered much more frequently than the other three aspects of results. The evaluations have sufficient funds to address the Terms of Reference, but additional funds would be needed if the evaluations were to extend their coverage to assess impact in greater depth and to address the issues of sustainability and efficiency.

There are no systems for quality assurance other than the reference groups that are set up, or the feed-back from the person who commissions and receives (and comments on) the report. The evaluations seldom undertake comparisons explicitly, either with baseline data or relevant benchmarks or through the use of counterfactuals. The evaluative judgements drawn are based on experience and on what the evaluators consider to be 'worth and merit'. The evaluations seldom deal explicitly with attribution and to the extent that they make causal claims (which is rare) these are based on rudimentary forms of contribution analysis. The internal validity could often be considered quite high but the external validity is low.

The organisations have not established any formal systems for learning. The results from the reports are not transmitted onwards in any systematic fashion. The evaluation reports are made public and most entered into the Norad database. Some are published on the organisation's web-page, and can be found in bookshelves in the organisation. There are no systems whereby evaluation reports are systematically collected and made available.

Learning and use take place in a variety of ways. The most common form of use is instrumental, but other forms also occur. The learning that takes place tends to be of the 'first-loop' type, that is, assimilation within existing knowledge structures. Local partners, in particular implementing partners, are involved in monitoring and evaluation. They are consulted on the initiative to evaluate and on the terms of reference, and they provide information to evaluation teams; but here practice varies between the headquarters and country level evaluations. Consultations on the former are less widespread. The organisations have capacities to initiate, conduct and learn from evaluations. They have functioning evaluation cultures. There are weaknesses and strengths to the systems, which can be summarised as follows:

Weaknesses:

- Policies tend to be descriptive and do not articulate purpose and other strategic choices explicitly.
- Several of the processes of the evaluation function are ad hoc and there is a need to formalise and systematise them more, in particular in relation to,
 - Choice of design and methods,
 - Budgeting the evaluation process,
 - Quality assurance,
 - Management response.
- There is no institutional memory of monitoring and evaluation.

Strengths:

- Strong commitment to evaluation and many positive attributes associated with it.
- Relatively skilled and dedicated human resources; many have been exposed to monitoring and evaluation.
- Good understanding of evaluation as a complex process, involving both measurement and assessment, personal and relational skills and technical competence.

- Cooperation with implementing partners and beneficiaries and a commitment to participation in the process.
- Significant investment in structures and processes and new skills added in recent years.
- Strong support for changes that would address the shortcomings listed above.

The discussions that take place between the organisations and Norad do not go into any details on monitoring and evaluation and Norad is not perceived as taking any keen interest in these functions. Norad is concerned that the organisations have systems for monitoring and evaluation, but until this study was commissioned it had been assumed that the organisational descriptions in the application for funds, and the subsequent annual reports, were satisfactory. The Norad staff members in the Civil Society Department rarely read evaluation reports from the organisations, nor has the dialogue given rise to any inquiry into the qualities of evaluation or monitoring information. The organisational reviews have sometimes provided insights on monitoring and evaluation, and when that has happened the organisations have been quick to implement suggested changes.

Third Objective: Recommendations

In light of the strengths and weaknesses identified above, the study identifies six areas where recommendations are suggested.

1. Policy guidance

Recommendation: The policies or the policy-like statements on evaluation and the evaluation function need to be sharpened to become more practically useful and to give strategic guidance to the evaluation function.

A policy statement should be short, no more than 5 pages of text, it should be developed in consultation with staff members, and it should be easily available and associated with a dissemination strategy. It might be helpful if it is seen to be established by the Board of the organisation. The policy should cover the following issues.

- i. It should define monitoring and evaluation in such a way that the key characteristics are clear and that this reflects how the organisations want to work with evaluation.
- ii. It should clarify the purpose of evaluation and set priorities among potentially conflicting purposes.
- iii. It should indicate the organisational resources to be allocated to monitoring and evaluation.
- iv. It should describe the division of roles and responsibilities of different parts of the organisation for monitoring and evaluation.
- v. The policy ought to specify clearly how the organisation will respond to evaluations, especially the management response to the recommendations and if they are accepted how/when they will be implemented.

2. Engaging Governing Boards

Recommendation: Boards and senior management need to ensure that M&E functions work well and they should provide strategic guidance.

Monitoring and evaluation are key functions for learning and for knowledge about results. A visible engagement from the Board and senior management provide an incentive to develop the M&E function professionally. The Boards do not necessarily need to be presented with evaluation findings, but they need to know how the organisation works with monitoring and evaluation, and they might need to have a summary of evaluation findings presented to a Board meeting once or twice a year.

3. Review the functions of M&E units.

Recommendation: The organisations should undertake a review of the present job descriptions of their M&E units and check these against the potential functions of such a unit, and then possibly revise tasks and functions.

At present the M&E units undertake several tasks relating to evaluation, but the M&E units in the six organisations do different things and some of the potential functions of such a unit are not carried out anywhere. The following is a list of functions that an M&E unit **might possibly** undertake:

- Developing the monitoring systems for outcome and impact indicators; designing the systems, allocating responsibilities for data collection, processing and analysis.
- Developing the monitoring systems for intervention level indicators; designing the systems, allocating responsibilities for data collection, processing and analysis.
- Undertaking all or some of the functions of the monitoring systems, such as data collection, processing and analysis, and disseminating/using the findings.
- Develop procedures for evaluation, such as by providing examples of Terms of Reference, examples of reporting, practices of participation, methodological guidance, etc.
- Undertake training in monitoring and evaluation for the organisation itself, for field offices, and for partner organisations.
- Be responsible for the long-term evaluation plan and, based on this, set up and monitor a budget for monitoring and evaluation.
- Provide a help-desk function for monitoring and evaluation and assist coordinators, advisers and project managers with advice and assistance.
- Participate in reference groups for evaluations, at country level or at headquarters
- Implement strategic evaluations, develop Terms of Reference, commission the evaluations and receive the final reports.
- Take part in external professional networks and develop the monitoring and evaluation function continuously.

- Manage the quality control of monitoring and evaluation.
- Supervise a management response system and ensure it functions properly.
- Establish an institutional memory in respect of monitoring and evaluation, with a database of monitoring and evaluation related information.
- Manage communications and public relations related to monitoring and evaluation to ensure; (1) that the key messages from evaluations are communicated with the agency's "public"; and (2) that the other communications material sent out (e.g. on impact) by the agency is consistent with the findings from evaluations.
- Follow up and evaluate the monitoring and evaluation function, ascertaining that the organisation's purpose for evaluation is achieved.

We are not recommending that an M&E unit should undertake all these function, but we recommend an explicit choice and formal decisions on where such monitoring and evaluation related responsibilities will be placed, and within an M&E unit is one option.

4. Budgets and budget follow up

Recommendation: Annual or multi-year budgets for evaluation need to be established and a linked cost control system or oversight function established.

Monitoring and evaluation can be costly and it is necessary to have a cost control of these functions as well as of other functions. The budget needs to take account of direct as well as indirect costs, and such costs need to be included in the construction of an annual budget and taken into account when the value of monitoring and evaluation is assessed. Evaluators often assess interventions against the question of whether they provide 'value for money'. The monitoring and evaluation function needs to be accountable in the same fashion, and hence one needs to know how much it costs.

5. Developing and standardising key processes

Recommendation: Key processes related to the evaluation function need to be developed and standardised.

There are three key aspects of the evaluation processes that need to be developed. These are:

- i. **Anchoring the evaluation process in the organisation** and among partners through the establishment of reference groups with clear mandates. There should be regulated procedures for how and when this is done, what responsibilities this implies, and how associated costs will be met. It is important to distinguish between different types of evaluations and not overburden relatively 'light' evaluations with bureaucratic procedures, but still make sure that they benefit from best practice in terms of organisational commitment.

- ii. **Ensuring quality control**, and clarifying responsibilities for ensuring that this occurs. In particular, this study has shown that there are systematic weaknesses in evaluation design, assessment of outcomes and impact, and in terms of causal analysis. While not all evaluations may need to cover these topics, some should. Quality assurance needs to be seen as an activity that is taking place before, during and at the end of the evaluation process.
- iii. **Developing management response systems.** Such systems help to make sure that the efforts invested in evaluation are not wasted and that they document how the organisation responds to evaluation findings. The systems need to include a higher level of transparency around evaluations which can, when properly managed, raise the level of debate and contribute to organisational learning.

6. Review the incentives and obstacles to learning from monitoring and evaluation

Recommendation: Current incentives for and obstacles to further learning need to be reviewed with a view to deepening and extending the evaluation learning process.

This study concludes that there are not any particularly strong formal and explicit incentives to use and learn from monitoring and evaluations. This needs to change. The motivation to learn is internal and implicit and comes from knowledge of a job well done. At the same time there are not any significant obstacles either, even though one can hypothesise that job pressure, mobility and turnover rates, financial pressure, and many other things at times have detrimental consequences for use and learning.

However, the impact of such processes lies beyond the scope of a study of monitoring and evaluation. Decisions on human resources and career development, financial management, etc. are related to many other factors besides whether they provide incentives or disincentives for evaluation, and hence they are not treated systematically in this report. In the long run, it would be useful for the organisations themselves to analyse what these incentives are and how they can be strengthened, and if there are any obstacles that can be removed. Such a review could be part of the functions of an M&E unit, and part of processes of 'learning to learn'.

References

- Alkin, M.C., & Taut, S.M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, 29, 1-12.
- Argyris, C. and Schön, D. (1978) *Organisational Learning*. San Fransisco: Jossey-Bass
- Barbier, J-C. and Hawkings, P. (2012) *Evaluation Cultures: Sense-making in Complex Times*. New Brunswick: Transaction
- Blakemore, K. (1998). *Social Policy: an Introduction*
- Bye, V. (2000), "Fra fattigdomsbekjempelse til bygging av det sivile samfunn. En kartlegging av erfaringer med norsk bistand gjennom frivillige organisasjoner (1987-1999)", NORLAT.
- Boswell, C. (2009) *The Political Uses of Expert Knowledge*. Cambridge: Cambridge University Press
- Christie, C. A. (2007). "Reported influence of evaluation data on decision makers' actions: An empirical examination." *American Journal of Evaluation*, 28, 8-25.
- Dension, D.R. (1990) *Corporate culture and organisational effectiveness*, New York: Wiley
- DFID Quality Assurance Template
- Edwards, M. (2004) *Civil Society*. Cambridge: Polity Press.
- Europe Aid Quality Control Checklist
- Forss, K., Cracknell, B. and Samset, K. (1994) Can Evaluation Help an Organisation Learn? *Evaluation Review* Vol 18 (4): 574 – 592
- Forss, K. (1997), "The Quest for Quality - Or Can Evaluation Findings be Trusted?" *Evaluation* Vol 3 (4) 1997.

Forss, K., Carlsson, J. and Rebien, C. (2002) "Process Use of Evaluations: Types of Use that Precede Lessons Learned and Feedback". *Evaluation* 8(1) 29 – 45

Forss, K. (2012) *Utvärdering av utvärderingar en analys av kvaliteten hos ett urval rapporter från SADEV*. Commissioned by the Ministry for Foreign Affairs. Stockholm.

Forss, K. (2012) "Four Organisations – Four Evaluation Cultures." In Barbier J-C. and Hawkins, P. (Ed) (2012)

Forss, K. (2012) *Aggregation and Analysis of Results, Recommendations, and Lessons Learned from Sida's Strategic Evaluations*. Commissioned by Sida, Stockholm.

Fowler, A. (1997), *Striking A Balance: A Guide to Enhancing the Effectiveness of NGOs in International Development*, Earthscan, London.

Hofstede, G. (1991), *Cultures and Organisations: Software of the Mind*, New York: McGraw-Hill,

Indevelop (2012) *Final Annual Report 2011 – 2012. Framework Agreement for Sida Reviews, Evaluations and Advisory Services on Results Frameworks*. Report to Sida. Stockholm.

Johnson, K. et al. (2009) "Research on Evaluation Use. A Review of the Empirical Literature From 1986 to 2005". *American Journal of Evaluation*. 30 (3): 377 - 409

Kelly, G. (1955) *The Psychology of Personal Constructs*. New York: Norton

Kirkhart, K. (2000) "Reconceptualising evaluation use: an integrated theory of influence". *New Directions for Evaluation* 88: 5-24.

Kruse, S.E. (1998), "Hvilke resultater skaper de frivillige organisasjonene? Erfaringer fra et utvalg evalueringer." NORAD.

Kruse, S.E. (1999), "Hvilke resultater skaper de frivillige organisasjonene? En gjennomgang av evalueringsrapporter fra norske frivillige organisasjoner." NORAD.

Kruse, Stein-Erik (2003), "Evaluating Development Effectiveness, Challenges and the Way Forward", The World Bank.

Lederman, S. (2011) "Exploring the necessary conditions for evaluation use in program change". *American Journal of Evaluation* 33(2): 159 – 78

MFA (2006), "Nye roller for frivillige organisasjoner i utviklingssamarbeidet". Utredning fra utvalg oppnevnt av Utenriksdepartementet.

Mayne, J. (2012) "Independence in Evaluation and the Role of Culture". In Barbier and Hawkins (2012)

Norad's Civil Society Panel (2012), "Tracking Impact, An exploratory study of the wider effects of Norwegian civil society support to Countries in the South".

Oakley, P., (Team Leader) S. Folke (Co-ordinator In-Depth Studies)(1999a) *The Danish NGO Impact Study: A review of Danish NGO activities in developing countries. Synthesis Report*. Copenhagen, Danida, Ministry of Foreign Affairs.

Oakley, Peter (1999), "The Danish NGO Impact Study. A Review of Danish NGO Activities in Development Countries", INTRAC.

OECD/DAC Glossary of Evaluation Terms.

OECD-DAC Quality Standards for Development Evaluation

Paquette, L. (2002) *Analyzing National and International Policy*. Rowman Littlefield.

Parker, M. (2000) *Organisational Culture and Identity*, London: Sage.

Patton, M.Q. (1997) *Utilization-focused Evaluation*. London: Sage

Perris, C. (1986) *Kognitiv Teori i Praktik och Teori*. Lund: Studentlitteratur.

Piaget, J. (1976) *The Essential Piaget*. Edited by Howard Gruber and Jacques Voneche. New York: Basic Books

Riddell, R. C. with L. Bategeka, K. Kotoglou and D. B. Ray (2008) *Assessing and measuring the impact of aid: evidence, challenges and ways forward*. Final Report of the research for the Advisory Board on Irish Aid. Oxford: Oxford Policy Management.

Riddell, R., S.E.Kruse (1997), "Searching for Impacts and Methods. NGO Evaluation Synthesis Study", OECD/DAC, Ministry of Foreign Affairs, Finland.

SADEV Report 2008:7 *Reaping the Fruits of Evaluation*. SADEV: Karlstad.

Saunders, M. (2012) "The use and usability of evaluation outputs: A social practice approach". *Evaluation* Vol 18 (4): 421 - 435

Schein, E. (1985) *Organisational Culture and Leadership: A Dynamic View*. London: Jossey-Bass

Schwartz, R. and Mayne, J. (2005) *Quality Matters. Seeking Confidence in Evaluating, Auditing, and Performance Reporting*. New Brunswick: Transaction.

Sida Studies in Evaluation, "2008:1; Are Sida's Evaluations Good Enough?"
Stockholm: Sida

Spencer et al (2003) *Quality in Qualitative Evaluation: A framework for assessing research evidence*. Government Chief Social Researcher's Office. London: Cabinet Office

Stern et al (2012) "Broadening the Range of Methods for Impact Evaluation". DFID Working Paper 38, 2012.

Ternstrom Consulting AB (2011) *Results of Development Cooperation through Norwegian NGOs in East Africa. Evaluation Department, Report 1/2011*. Oslo: Norad

Van Rooy, A. (Ed) (1998) *Civil Society and the Aid Industry*. London: Earthscan.

Vedung, E. (2000) *Public Policy and Programme Evaluation*. Transaction: New Brunswick

Weiss, C. (1998) "Have we learned anything new about the use of evaluation?" *American Journal of Evaluation* 19(1): 21 – 33

Woolcock, M. (2009) "Toward a plurality of methods in project evaluation: a contextualised Approach to understanding impact trajectories and efficacy" *Journal of Development Effectiveness*, Vol. 1, No. 1 (March 2009), pp. 1–14.

Annexes



Annex 1 – Terms of Reference

Terms of Reference for the Study of the evaluation function in Norwegian civil society organizations

Background

Norway is providing billions of Norwegian kroner (NOK) in support of development cooperation through civil society organizations (CSOs). A substantial part of this is channeled through assistance schemes administered from Norad to Norwegian CSOs. In 2010, with a total Norwegian support to CSOs of 3, 6 billion NOK, 1, 7 billion or 46 % were channeled through Norad. The largest budget line is managed by the Civil Society Department (SIVSA) in Norad. In 2010 SIVSA provided more than 1, 2 billion NOK to Norwegian CSOs and their national partners in 74 countries for long-term development projects. A large part of this support is channeled through relatively big CSOs covering important sectors or fields of interest: Norwegian Refugee Council (NRC), Norwegian Red Cross (NORCROSS), Norwegian Church Aid (NCA), Norwegian Peoples' Aid (NPA), Save the Children-Norway and Digni. For several reasons SIVSA now enters into multi-year agreements with these organizations. An important basis for such agreements is judgments of the organizations' ability to handle the resources allocated and implement the initiatives supported. In order to contribute to a quality assured, efficient and results based use of these resources - including reaching the overall goals for aid through the cooperation with CSOs – it is crucial that the organizations have a suitable system of quality assurance and evaluation in place, and that it is implemented effectively in the organization. The evaluation function in the organization is central in this perspective.

The evaluation function is referring both to the evaluations themselves within the organization and their use, and the monitoring and evaluation system (M&E) set up for policy development and management in the organization. On this basis a study will be conducted of the evaluation function in Norwegian CSOs with Norad support.

The purpose of evaluations is twofold, serving both learning and control function:

- a) To systematize knowledge of results and achievements and challenges with a view to improving similar measures in the future (learning function)
- b) To consider whether a measure has been implemented as agreed and/ or whether the anticipated results have been achieved (the control function)

Whether the main emphasis is on the learning or the control/documentation function will vary according to the evaluation in question. The use of evaluations as a source of knowledge and continuous learning is, however, at the core of all evaluation activities.

This study will pay particular attention to the question of learning: How should the monitoring and evaluation function in the organization be set up in order to contribute to a knowledge based organization, focusing on knowledge development and quality assurance of knowledge in relevant areas, knowledge sharing and use?

Purpose and objectives

The purpose of the study is to provide an overview and assessment of the role of the evaluation function in Norwegian CSOs, to ensure that the knowledge and evidence produced is of high quality, and that it is shared and leads to learning. It will furthermore give insight and recommendations contributing to knowledge oriented monitoring and evaluation system suited for quality assurance of the organization's work and results.

The study will have the following objectives:

- Map the evaluation function in selected Norwegian CSOs
- Assess the evaluation function of the organizations, regarding both the relevance and quality of reports and the system for transferring results, ensuring use of evaluations and learning
- Provide recommendations to promote an evaluation function adequate for quality assurance of the organizations' work and results, and for knowledge sharing and learning, internally and externally

Scope of work

The study will include a mapping and assessment of the evaluation function in select Norwegian CSOs. It will describe how the evaluation function is structured within the organization, looking at areas such as independence (line of reporting, organizational independence of the director, budget/core resources, selection of evaluation object, area of responsibility), accountability (evaluation policy specifying roles and functions, systems for follow-up, publishing of reports and related documents), learning/use (learning strategy, quality control of decentralized evaluations, training), and questions of transparency (publication of ToR) and who conducts the evaluations (independent consultants/researchers and/or staff). It will further assess the quality of the products produced by the evaluation function, focusing on methods used, whether the evaluations assess the results and impacts of the projects studied and discuss the question of being able to attribute the results to the particular activities supported. It will also consider which incentives and capacities are in place to initiate, carry out, learn from and use evaluations in the daily work.

The following organizations are preselected as cases for study: NRC, NORCROSS, NCA, NPA, Save the Children-Norway and Digni.

The timeframe of the study will be 5 years, looking at the present monitoring and evaluation function, including assessment of important changes over the period, and selecting evaluation reports finalized in the period 2007-2011, including if relevant also assessment of Terms of Reference of not yet finalized evaluations from the period.

In response to the purpose and objectives set for the study, the following issues and questions – not an exhaustive list - should be included:

- A) Does the organization have a results based program management? Is there an evaluation strategy? Is there an evaluation policy?
- B) Does the organization have a monitoring system? Does it cover internal and external needs for information? To what extent are there preliminary studies, needs assessments and baselines for programs and projects in the organization?
- C) How is M&E organized in the organization? How are evaluations decided and budgeted? How is the process: announcement etc? Who conduct the evaluations? What kind of system exists for follow-up?
- D) What kinds of evaluations are conducted in the organization? Are joint evaluations considered when relevant? Is analysis of outcome/impact included (what kind of method is used to identify the counterfactual)? Are assessments of theory of change, program theory included? Are alternative theories of change assessed? Is analysis of context included? What are the methods used? To what extent are the evaluations regarded as relevant, covering the work of the organization? To what extent are evaluations aiming wider than the project level? To what extent and how is contribution to capacity building in civil society - capacity building of the organization and the partners - covered? To what extent is anti-corruption and financial management covered? To what extent are cross cutting issues, i.e. women's perspectives, vulnerable groups, human rights, MDGs covered in the evaluations?
- E) What is the quality of the reports? What criteria are covered? To what extent do they look at relevance, effectiveness, efficiency, impact and sustainability? Do they consider whether the right things are done, and in the right way? Are the evaluations sufficiently budgeted in order to enable responding to ToR? What kind of quality assurance system is used for the evaluations (reports)? Since evaluation is to compare with something – either a counterfactual or a desired benchmark – against what do they measure their findings to draw conclusions and recommendations? How do they deal with the attribution issue and how does this compare to established best practice? What are the external and internal validity of the evaluation reports?

- F) What kind of system for learning has the organization established? How are results from M&E reports transmitted? How is the follow up of reports? When in the process of evaluations are learning and use taking place? How are local partners (stakeholders and/or implementing partners) involved in the evaluation process, in relation to purpose, design, implementation and learning? What are the capacities and incentives in the organization to initiate, conduct and learn from evaluations?
- G) To what extent does the organization find the established system and the contact with Norad appropriate, and particularly related to internal and external learning?
- H) When relevant, to what extent has the organization followed up recommendations from Norad's evaluations?

Approach and methodology

The study will include desk studies of evaluation reports, program documents and agreements, and strategy documents of the organizations. A combination of in-depth interviews with staff in the organizations and in relevant departments in Norad (civil society, methods and results, and evaluation) and surveys shall be conducted. Case studies are included, focusing on organizations and comparison of evaluation functions in the organizations. A selection of organizations will be included in the study: NRC, NORCROSS, NCA, NPA, Save the Children-Norway and Digni. Country related case studies are not included. It might, however, be desirable to include interviews on Skype and surveys to capture links and communication between headquarter and field office, and other central partners for the work of the organization.

The evaluation shall refer to the DAC criteria on evaluation of international development cooperation, and the Consultant should clarify the use of the criteria. For impact evaluations please refer to the 3ie criteria. The principle of "Do no harm" and ethical considerations should moreover be considered. Reports will be assessed against the DAC evaluation quality standards. All proposals must follow the DAC evaluation guidelines.

The Consultant will be responsible for developing a detailed methodological framework for the evaluation. The Consultant is free to suggest additional methods and questions that have not been indicated above. If the Consultant leaves some of the detailed elaboration of the methodology to the inception report, the methodological design shall be sufficiently developed in the tender for the client to be able to make a proper assessment of the offer. The evaluation report shall describe the evaluation method and process and discuss validity and reliability. Limitations and shortcomings shall be explained.

Organization and evaluation team

The evaluation will be carried out by an independent team of researchers/consultants. The contract will be issued by the Evaluation Department (Norad), according to standard procurement procedures. Evaluation management will be carried out by the Evaluation Department, and the team will report to the Department. The team is entitled to consult widely with stakeholders pertinent to the assignment. The inception report, the draft evaluation report and the Final report are subject to approval based upon quality criteria by the Evaluation Department.

A group of stakeholders, a reference group representing the Norwegian organizations studied and the Civil Society Department will be established, administered by the Evaluation Department, to advice and comment on the evaluation products throughout the process.

The team shall involve stakeholders in the evaluation process with a view to making the process useful in improving their work.

The team, and especially the team leader, shall have extensive competence and experience in evaluation principles, methods and standards. Extensive knowledge of the institutional structures and functioning of CSOs in development is also required. At least one team member must be able to read Norwegian without any problems of understanding.

A system of quality assurance shall be in force, with ability to control both the formal and substantial aspects of the evaluation reports, including a high quality linguistic level for the reports. The system shall be carefully described in the tender, with a clear indication of the number of person days that will be allotted to the quality assurance functions.

Budget, work plan and reporting

The study is budgeted with a maximum input of 20 person weeks. The tender shall present a total budget, with allocation of resource to the members of the team (also describing their responsibilities), and other expenses envisaged. There shall be room in the budget for presentation of the final evaluation report in Oslo during a half-day seminar to be organized by the Evaluation Department. At least one key member of the evaluation team shall be available in Norway for Norwegian stakeholders during a full working day at the end of the evaluation in order to discuss ideas for its follow-up with them individually.

Tentative work plan and activities:

Activity	Grand Deadlines
Announcement of tender	07.05.12
Submission of tenders	5.06.12
Contract signature	06.07.12
Inception report	20.08.12
Draft final report	30.10.12
Final report	07.12.12
Publication, seminar	January 2013

During the evaluation process, the Consultant shall submit the following reports in English:

- An inception report providing a first draft mapping of the evaluation function in the organizations studied, as well as a detailed methodological proposal for the study.
- A draft final evaluation report presenting findings, conclusions and recommendations, with a draft executive summary. Principal stakeholders will be invited to comment in writing, and feedback will be provided to the team by the Evaluation Department. The feedback will refer to the Terms of Reference and may include comments on all aspects of the report.
- A final evaluation report shall be prepared in accordance with the guidelines of the Evaluation Department. Upon approval, the final report shall become available in paper version and electronically to the general public in the series of the Evaluation Department, and must be presented by the team in a form that directly enables such publication.
- It will be the responsibility of the team to deliver reports that have been proof read. Tables must be submitted both in word and excel, and all supporting material and evidence, including interview transcripts, must be collected by the team and be made available to Norad's Evaluation Department upon request.

Annex 2. List of Persons Interviewed

Digni (and Digni's organizations)

Jørn Lemvik, Secretary General
Arne Kjell Raustøl, Senior Advisor
Marianne Skaiaa, Advisor
Heidi Westborg, Secretary General, Himal Partner
Hanne Holmberg, Regional Leader (Asia) Mission Alliance
Hjalmar Bø, Board Member
Åshild Solgaard, Project Consultant, Bible Society
Marianne Næss Norheim, Mission and Project Advisor, Normission
Bo Christoffer Brekke, Project Consultant, Salvation Army

Norwegian Church Aid

Ingvild Langhus, Team Coordinator and Advisor for planning, monitoring and evaluation
Kari Øyen, Area Team Leader, former Country Representative in Malawi (up to 2011)
Elisabeth Mustorp, Security Advisor, former Country Representative in Sudan (up to June 2012)
Arne Dale, Program Coordinator for Brazil (direct follow up of partners from Oslo), also member of the NCA board as representative from staff
Anja Riiser, CSR Advisor and previously responsible for evaluations
Eivind Aalborg, Head of Department International Programmes
Tariq Mustafa Ali, Sudan Gender & Peace Programme Manager

Norwegian People's Aid

Kjersti Bjerre, Coordinator of Evaluation
Finn Erik Thoresen, Board member (former Chair)
Kjersti Jensen, Board member
Trude Falch, Program Coordinator
Eva Haaland, Adviser Southern Africa and Norad
Liv Bremer, Adviser Southern Sudan
Orrvar Dalby, Director International Department
Per Ranestad, Strategic Management, previously Regional Director for Latin America
Felipe Atkins, Country Director, Rwanda (skype)

Norwegian Refugee Council

Cara Winters, M&E Adviser
Ronny Rønning, Head of Strategic Management Support
Toril Brekke, Head of International Department
Patrik Ekløf, Head of Section Horn of Africa
Heide Solheim Nordbeck, Donor Adviser

Save the Children Norway

Ingunn Tysse Nakkim, Senior Evaluation Adviser
Yngve Seierstad, Director Strategy, Programme Quality and Finance
Sine Christensen, Special Adviser Education
Gunnar Andersen, International Programme Director
Sigurd Johns, Former Head of Policy and Programme Development
Markus Aksland, Regional Director for Asia, and former Country Director for SC in Cambodia
Tove Wang, General Secretary
Amos Chinyama, M&E Focal Point SC Zimbabwe (Skype conference)

Norwegian Red Cross

Svein Erik Bersås, Quality Assurance
Trine Moa, Coordinator Quality Assurance and Evaluation
Lars André Skari, Head of Administration and Quality Assurance
Sven Mollekleiv, President, Chairman of the Board
Torgeir Nærlie Vasaasen, Programme Coordinator Southern Sudan

Plus a workshop on PMER, utilisation and learning with IFRC staff and representatives from Canadian, German, Swedish, Danish, British, Australian, Finnish and Icelandic Red Cross Societies, in total 20 persons.

Norad, Civil Society Department

Gunvor Skancke, Deputy Director
Vigdís Wathne, Senior Adviser
Anne Merete Ødegaard, Adviser

External Evaluators

Nora Ingdal, Nordic Consulting Group
Arne Tostensson, Christian Michelsens Institute

Annex 3. Data Collection Instruments

Interview Guidelines

1. Is there a policy for M&E in the organisation
2. Brief description of the M&E policy:
 - Definition
 - Purpose
 - Organisation
 - Strategy
 - How does the strategy differentiate monitoring and evaluation
3. How is the policy communicated and how well known is it in the organisation? Compliance? What's the history of the policy?
4. Is the M&E function organised into a specific unit?
5. If there is an M&E unit:
 - What functions, if any, is it grouped with
 - Whom does it report to
 - How are personnel assigned to it
 - How is the evaluation plan decided
 - Who approves of the budget
 - What are the systems of interaction with other units
 - Where in the organisational hierarchy is the evaluation unit found
 - How many people are employed there
 - What are their qualifications
6. Does the function report to the Board, and if so, how?
7. Is M&E a line function?
8. Describe the level of centralisation/decentralisation of monitoring and evaluation?
9. How is the evaluation function linked to other functions in the organisation?
10. Is there an evaluation library/database

11. Resources spent on M&E as percentage of the organisation's total turnover?
Is that sufficient and what would be appropriate allocations?
12. How are evaluations initiated?
13. Which types of activities are evaluated?
14. How is evaluation linked to monitoring?
15. Are evaluations done by external experts?
16. Are participatory evaluations encouraged? Participation in monitoring?
17. How is an evaluation budget decided?
18. How are evaluation consultants engaged?
19. Are there norms and standards for terms of reference?
20. Is the evaluation process defined and regulated?
21. Can an average duration of the evaluation process be seen?
22. Are there any norms for interaction between evaluators/evaluation team and the organisation?
23. What is the process for receiving reports? Is there a management response system? If not, what other means are there to make sure evaluations are used?
24. Is there any quality control of evaluation reports and if so, what does it look like?
25. Are evaluations made public? Presented in a report series? Are evaluations made public on the web, do they feature in other forms in social media?
26. Are evaluations presented in other ways than through reports?
27. Is there any training on evaluation, for example how to write ToRs, how to actually do evaluations, how to use evaluation results? If so, who takes part in such training?
28. What do the networks look like, is the organisation a member of any evaluation networks, for example the Norwegian Evaluation Society, the EES, or other professional networks? If no, why not?
29. Are evaluations considered to be useful tools for learning? What would the obstacles be to learn from evaluations?

30. What is the single most important strengths of the M&E system
31. What is the greatest weakness of the M&E system?
32. Is there an evaluation culture in the organisation? If so, what are the most prominent features of it?
33. What are the significant changes of M&E in recent years and to what extent have they been established in dialogue with, or on recommendations from, Norad.
34. How does the dialogue with Norad serve the organisation, does it encourage learning or in other ways serve to strengthen the organisation. In what other ways can the organisation benefit from the experiences in development cooperation in Norad's possession, for example, evaluation systems, studies, etc.?

Survey (Starting next page)

Part 1 - Background and Engagement with Monitoring and Evaluation

The survey is completely anonymous: we will be able to track who has submitted a response but we won't be able to link responses to email addresses. So please respond frankly and sincerely.

1. Which organisation are you working for?

- Digni (or one of Digni's organizations, including partners)
- Norwegian Church Aid
- Norwegian Red Cross
- Norwegian Refugee Council
- Norwegian People's Aid
- Norwegian Save The Children

2. How long have you been employed by that organisation?

Number of Years

3. Have you worked for any other CSO organisation?

- Yes
- No

4. Have you had any training in M&E activities?

- Yes
- No

5. If yes, what kind of training have you had?

- University programme
- External workshops/seminars eg. 'bistandstorget'
- Internal training activities

Other (please specify)

6. Do you take part in any M&E network(s)?

- Yes
- No

7. If yes, which one(s)?

Part 2 - Values, Beliefs and Attitudes

The survey is completely anonymous: we will be able to track who has submitted a response but we won't be able to link responses to email addresses. So please respond frankly and sincerely.

8. To what extent would you agree/disagree with the following statements:

	Strongly Agree	Agree	Disagree	Strongly Disagree
Evaluation is a good learning tool	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Everybody needs evaluation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation is distinct from other organisational processes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You need to be independent to evaluate properly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Only those specifically trained for the purpose can evaluate properly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation is fun	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation reports tend to be tedious and bureaucratic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is the process of evaluation, not the reports, that are useful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation is mainly for top management	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is those who are close to the activity that can evaluate it properly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluations must be standardised and their quality controlled	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluations tend to become empty rituals	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluations are done for others, not for ourselves	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluations are usually of high quality and are reliable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resources spent on evaluation are often wasted	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Our organisation makes good use of evaluations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Few decisions are taken without any evidence base from evaluations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Many evaluations are not trustworthy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A penny invested in evaluation is a penny	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

invested in quality and learning				
Too few get properly involved in evaluation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We must be careful so we don't evaluate too much	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluations build the legitimacy of our organisation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Without evaluation we would not know much about results	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluators do not communicate well with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is not necessary to have a report when you evaluate	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation can be many different things, as long as it comes to a judgement about value	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation is above all distinguished by the scientific method	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is necessary to talk about evaluation results	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluations live their own life for their own purposes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation is mostly useful for external audiences	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation is interesting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation is a very specialised skill	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluations must be formally organized and regulated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How we work with evaluation changes over time	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
There is no one best way of evaluating	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluators in particular need to understand people and relations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation is essentially about measurement	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
An evaluator needs to be a technical expert	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation must be allowed time and get time for reflection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluations need to arrive in time in order to provide decision support	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If possible, I'd like to	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

attend evaluation training	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I read evaluations as often as I can	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would like to be more involved in evaluation work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Integrating evaluation activities enhance learning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Doing more evaluations would convince partners of needed changes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The recommendations should be either formally accepted or rejected and if accepted, there should be a system to ensure they are followed through	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluation ought to be part of the training and induction of all members of staff	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The findings from evaluations need to inform, shape and influence future projects	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The findings from evaluations need to help shape the messages that we use to communicate what we do to our members and the general public	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We can usually build our evaluations on high quality monitoring data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Part 3 - Practical Experience with Monitoring and Evaluation

The survey is completely anonymous: we will be able to track who has submitted a response but we won't be able to link responses to email addresses. So please respond frankly and sincerely.

9. Have you taken part in any evaluation assignment?

- Yes
 No

10. Have you been working with monitoring activities?

- Yes
 No

11. Do you normally read the evaluations that are commissioned by your organisation?

- Yes, all of them
 Yes, most of them
 Some of them
 No, never

12. When you are personally involved in evaluations, how much do you read?

- Usually the full report
 Usually glance through the full report and read bits and pieces
 Usually only the summary

13. Could you provide an example of an evaluation that you found useful in your work?

- Yes
 No

If yes, please provide info on which evaluation(s) you found useful

14. To what extent do you read evaluations from other organisations, such as other CSOs, Norad, DFID, etc.?

- Frequently
 At times
 Rarely
 Never

Annex 4. Format to Assess Evaluation Quality

1. Basic information about the evaluation

Title:	
Evaluation object:	<input type="checkbox"/> project/programmes <input type="checkbox"/> sector support <input type="checkbox"/> policy <input type="checkbox"/> organisation specific
Evaluation budget and/or cost of the evaluation:	
Evaluation team: (size, nationality)	
Team leader: (Gender, nationality, professional background)	
Host country participation:	
Program characteristics	<input type="checkbox"/> Simple <input type="checkbox"/> Complicated <input type="checkbox"/> Complex
Evaluation ownership	<input type="checkbox"/> commissioned by the CSO alone <input type="checkbox"/> commissioned together with other funding agencies
Timing	<input type="checkbox"/> While activities are still going on <input type="checkbox"/> At the time when activities have been completed <input type="checkbox"/> Some time after all activities were completed (at least 1 year)

Comments on background information

2. Structure and presentation

	1*	2	3	4	5	6	ND*	NR
Is there a clear and adequate executive summary?								
Is there a clear and logical structure to the chapters of the report and to the report as a whole?								
Are illustrations and figures used to facilitate reading and understanding?								
Are tables, boxes and models well designed, clear and accurate								
Does the report make use of references and is it appropriately referenced?								
Are annexes well structured and readable?								
Is the report free from grammatical and spelling errors?								
Is the language of the report precise, varied and interesting, free from jargon?								
Is the report frank, does it address issues squarely and straight on?								
Is the report impartial and does it apply different perspectives to issues treated?								

- a key to the rating is found at the end, the same key applies to all tables.
- ND stands for not done at all in the evaluation, that is, cannot be assessed, NR stands for not relevant – meaning that the question is meaningless in the context of this evaluation. A slash in this column means that we cannot answer the question.

Comments:

3. Evaluation methodology

(a) Description

Nature of any preparatory work; baseline data, standards for benchmarking, or the like:		
What design does the evaluation have?		
What methods does the evaluation use?		
Population and samples		
Does the evaluation contain a description of:	<input type="checkbox"/>	Costs of the intervention
	<input type="checkbox"/>	Time of the intervention
	<input type="checkbox"/>	Purpose and logic of the intervention
	<input type="checkbox"/>	Description of activities and outputs
Does the evaluation treat*:	<input type="checkbox"/>	Efficiency
	<input type="checkbox"/>	Effectiveness
	<input type="checkbox"/>	Impact
	<input type="checkbox"/>	Relevance
	<input type="checkbox"/>	Sustainability
Does the evaluation use the following instruments:	<input type="checkbox"/>	Qualitative indicators
	<input type="checkbox"/>	Quantitative indicators
	<input type="checkbox"/>	Rating scales
	<input type="checkbox"/>	Benchmarks for assessment
	<input type="checkbox"/>	Comments:

* note whether in focus (1), secondary issue (2), or not treated at all (3)

(b) Assessment of methodological choices

	1	2	3	4	5	6	ND	NR
Description of the problem								
Are the terms of reference clear and focused?								
Does the evaluation interpret and focus the task as defined in the terms of reference?								
Is the basic question clearly stated in a specific section?								
Description of design								
Is there a section that describes the design/approach								
Is there a discussion of alternative designs								
Can the reader make an independent assessment of the choice of design								
Description of methods								
Is there a section that describes the methodological choices fully?								
Is there a discussion of threats to reliability and validity?								
Can the reader make an independent assessment of the evaluation methods?								
Instruments								
Are the instruments for data collection well designed?								
Are indicators appropriate?								
Are benchmarks fair and relevant?								
Are rating scales well designed?								

Comments:

4. The evaluation's analysis and assessment of the intervention

	1	2	3	4	5	6	ND	NR
1. Analytical content								
Does the evaluation present empirical material in the report?								
Is the analysis relating to the evaluation questions exhaustive and complete?								
Are findings and conclusions supported by the data?								
2. Analysis of management?								
Is there a trustworthy analysis of leadership and governance?								
Is there a trustworthy analysis of planning?								
Is there a trustworthy analysis of financial management								
Is there a trustworthy analysis of coordination?								
Is there a trustworthy analysis of networks and linkages?								
Is there a trustworthy analysis of organisational structures?								
3. Analysis of achievements								
Is there an accurate assessment of efficiency?								
Is there an accurate assessment of effectiveness?								
Is there an accurate assessment of impact?								
Is there an accurate assessment of sustainability?								
Is there an accurate assessment of relevance?								

Comments:

5. The evaluation process

	1	2	3	4	5	6	ND	NR
Was the evaluation completed on time?								
Was the evaluation completed within the agreed budget?								
Did the evaluators consult relevant authorities in the host country?								
Did the evaluators gather data from end-users or beneficiaries of the project?								
Did the evaluators gather data from the implementing organisations								
Did the evaluators use past evaluations on similar or related topics?								
Did the evaluators involve any stakeholder groups in the evaluation process?								
Did the evaluators interact with a reference group?								
Were the evaluation resources used efficiently?								
Were the evaluation resources used effectively								

Comments:

6. Conclusions and recommendations

	1	2	3	4	5	6	ND	NR
Does the evaluation respond to the questions in the terms of reference?								
Are the conclusions in the evaluation clear and consistent?								
Do the recommendations follow from the analysis and conclusions?								
Are the recommendations practical, can they be translated into decisions?								
Are there recommendations for clearly specified groups of actors?								
Are there relevant and for an informed audience interesting lessons learned in a specific section?								
Can an informed reader identify and make sense of lessons learned through the intervention?								
Has the evaluation added to a general understanding of development cooperation?								

Comments:

7. Holistic assessment of the evaluation

	1	2	3	4	5	6
Methodology						
Presentation						
Understanding of the intervention, its substance and context						
Innovation and creativity in the process of evaluation						
Depth and quality in the information provided through the evaluation						

Comments:

Key to the rating of quality indicators:

	6 – yes excellent, very well done	
	5 – yes, quite good	
	4 – yes, it can pass	
	3 – not quite adequate	
	2 – no, significant problems	
	1 – no, very poorly done	

Annex 5. Evaluations Studied in the Quality Assessment

Save the Children Norway

- Axel Borchgrevink et. al. Evaluation report, Save the Children Guatemala, 2008
- Cliff Harber and David Stephens, From Shouters to Supporters, Quality Education Project – Final evaluation Report, 2010
- INTRAC, Thematic Evaluation of Save the Children Norway's Cooperation with Partner, Final Global Report
- Save the Children in South East Europe, A Strategic Evaluation of Save the Children Norway's Programme, August 2010
- Evaluation of Save the Children Norway in Cambodia's Education Programme (2006-2009), 2010

Norwegian Refugee Council

- Casa Consulting, Evaluation of the NRC Colombia Programme 2008-2010, 2011
- Gisle Hagen et.al. Review of Norwegian Refugee Council Information, Counseling and Legal Assistance project, Liberia, 2010
- Mary Wyckoff and Hemang Sharma, Trekking in Search of IDPs and other Lessons from ICLA Nepal, A Study of NRC's ICLA programme in Nepal, 2009
- Oliver Wilds and Anthe Herrberg, Evaluation of the Mediation Support Unit Standby team of Mediation Experts, 2011
- TFM Consult, Mid term Evaluation of DFID CHASE Support for NRC Programmes, 2012

Norwegian Church Aid

- Acacia Consultants Ltd, Norwegian Church Aid, programme Evaluation report, Somalia Programme Activities in Gedo, Puntland and Mogadishu
- Christian Balslev-Olesen, Review Norwegian Church Aid-Darfur Programme, 2011
- Javier Martinez, Improved Health Training Education in Malawian Nursing Schools, Independent Mid-Term Review, 2008
- Luciano Nunes Padrao, Norwegian Church Aid, Brazil Country Programme Plan Evaluation, 2010
- Nora Ingdal, GSP Evaluation, Great Lakes Programme 2005-2007, 2008

Norwegian People's Aid

- Norwegian Institute of Urban and Regional Research; Review of the organisation of the Gender Equality (GE) work in the International Programme Department, Norwegian People's Aid (NPA), 2010
- Alvin R. Anthony; South Africa Program. Mid-term Evaluation, 2011
- ABC Consulting; External Project Evaluation of Youth Projects. 2011
- Elling N. Tjønneland and Arne Tostenson; Making Partnerships Work. Vision and Implementation of a Development Programme. A Review of the International Development Strategy of Norwegian People's Aid 2008 – 2011
- Ethiopia Final report 2008 – 2011

Norwegian Red Cross

- Towards Local Solutions: NorCross – RCSC OD partnership Impact Evaluation. 2010
- Penny Bardsley; Review: Together We Can Peer Education Programme; 2009
- Anders Eklund; The Somali Red Crescent Society Rehabilitation Programme, 2011
- Anna Höybye; Impact Evaluation Report: VNRC / NRC Partnership and Associated Projects, 2010
- Gomo, E.G., Gardner, B. Munyati, S., Matoka, S., and Toto, P. ; Southern Africa Regional HIV/AIDS programme, 2011

Digni

- Gadissa Bultosa, Hiwot Workneh, Emanu Getu, Evangelical Church Mekane Yesus - Development and Social Services Commission Western Ethiopia Integrated Environment and Food Security Development Programme. Mid-Term Review, 2011
- TAABCO R&D Consultants; Capacity building within Health Care, End of Term Evaluation Report; 2012
- Elsa Döhlle and Mulbah Kackollie; Evaluation of Liberia – Norway Partnership Community Development Programme, 2011
- Linda Forsberg och Lorentz Forsberg; Facing Mount Elgon: Mid-term evaluation of the Peace and Rights Programme, 2012
- Kate Halvorsen, Nguyen Trang, Tu Ngoc Chau Evaluation of Norwegian Mission Alliance Vietnam Development Program, 2011

Annex 6. Review of Past Evaluations

1.1 Increasing demand for evaluations

There is a changing climate for civil society also in Norway. Over the last twenty years and in particular during the 1990s, the profile, number and budgets of Non-Governmental Organisations (NGOs) or later CSOs expanded dramatically – based on the notion that the value of their interventions was different from that of governments – their engagement was people-centred, participative and built on partnerships. Yet studies of CSO performance gave rise to increasing scepticism about their added value and their assumed comparative advantages. By the end of the twentieth century, robust and conclusive evidence was simply not available to confirm this proposition and the high expectations that the CSOs' unique contribution to development and poverty reduction was being realised in practice.

Evaluations and progress reports often show that short-term project objectives have been achieved with positive, but often scattered micro-results. Yet, these studies have repeatedly told us little beyond the more immediate effects. In spite of a growth in the number of impact studies conducted, much less is known about the long-term impact and the wider effects of CSO development interventions beyond these projects and their direct and often limited effects on the beneficiaries who were assisted directly. In other words, there is knowledge about results about Norwegian CSO projects in the South from external studies and evaluation, but also from reviews and evaluations commissioned by the organisations themselves. The question is what do we know about the evaluations and the evaluation function in the Norwegian organisations?

1.2 Current knowledge about CSO evaluations

Norwegian CSOs have registered 519 evaluations/reviews in the Norad evaluation base covering the period 1998 to 2012¹. This number is far lower than the actual number of evaluations carried out – partly because the organisations have not entered all reports in the database and partly because few organisations have a complete overview of what reviews and evaluations they have carried out. Sometimes, it is also difficult to know when a report is an evaluation, a review or a travel report since there are no clear and agreed definitions. It is up to the organisations to decide when a report can be categorised as an evaluation and upload it in the database. The Norwegian

1 <http://www.norad.no/en/tools-and-publications/publications/reviews-from-organisations>

CSOs have also implementing partners and some operate through regional and country offices. This means that some evaluations are commissioned by the CSO HQ in Norway, others by the regional/country offices and others by the partner organisations themselves.

The number of evaluations commissioned by Norwegian CSOs have increased radically over the last 10-15 years, but we don't know how much. We also know that while a few organisations have full time staff working in evaluation units and with evaluations, it is much more common with evaluation focal points – staff to have part time responsibility for commissioning evaluations. However, there is so far no systematic knowledge about the evaluation function in the organisations. The organisational reviews of the CSOs commissioned by Norad don't provide much solid information on the evaluation function in the organisations (cover mostly monitoring and reporting) and even less on the quality and relevance of the evaluations carried out except for the reports on NPA and Save the Children. According to Norad requirements, the CSOs should have an evaluation policy and plan as part of their multi-year application, but it is up to the organisations to decide form and content. It is true that evaluations play a much more significant role in Norwegian CSOs today than ten years ago, but it is not an organisational priority for most of the organisations and driven by a demand from senior management and board of director. It is more a priority for a small group of particularly interested. We know much more about the substance of the reports – based on evaluations commissioned by Norad assessing civil society support – evaluations synthesising findings from the existing M&E systems. The following observations are primarily based on three studies:

- Two meta-studies of evaluations from Norwegian CSOs (Kruse 2003). The studies are in principle outdated, but we believe many of the findings are still relevant and confirmed in later programme and project evaluations. However, there are no similar meta-evaluations carried out more recently..
- The recent study by Norad's Civil Society Panel: Tracking impact of Norwegian CSOs in countries in the South (Norad 2012).

1.3 The CSO Meta-Evaluations

The purpose of the meta-evaluations was to provide Norad with better knowledge about CSO achievements based on reports submitted to NORAD over two years. Each year there was about 100 reports – covering 27 countries and 21 different Norwegian NGOs. Total number of CSO projects supported by NORAD each year totalled around 1000, so the sample analysed was about 10%. The first part of the report summarised results along different dimensions. The methodological chapter covered issues like:

- The evaluation process (what type of reports, when were they carried out, main focus).
- The scope of the evaluation (type and number of questions).
- The evaluation team (size, composition, external/internal, choice of team leader).

- The evaluation report (language, length, TOR, recommendations and summary included or not).
- Data collection methods (use of structured/unstructured, participatory/non-participatory, quantitative/qualitative methods).
- Sources of information (use of documents, staff, participants, field visits, monitoring data).
- The types of analysis carried out (discussion of limitations in methods, relevance to national needs and plans, cost-effectiveness analysis, analysis of capacity development efforts, lessons learned and policy/strategy discussions).

The main findings about short- and long-term results were:

- Data and information about long-term outcomes and impact were missing. 60% of the reports did not provide any information about impact. Some reports assessed impact and the assessments were in general positive, but the findings were not based on any systematic collection of data. Findings were mostly anecdotal. The reports were as such of limited value as a source of information about CSO long-term development effectiveness. Several reports explained well why impact data were missing and hard to collect.
- The reports assessing long-term effects were overall positive. None were categorised as very satisfactory, but nearly 40% as satisfactory in terms of achievements – and none “very unsatisfactory”.
- More and better information was available about outcomes – or short-term achievements. As in the OECD study (Riddell 1997), the more specific project objectives were to a large extent achieved. Two thirds of the projects had reached their short-term outcomes successfully. The CSOs delivered and they delivered well. Serious deviations from project plans were relatively rare – including “crisis projects” (only 8%). No reports described any form of corruption or serious misuse of funds possibly because they were not asked to.
- Most and best data and information were available about outputs. There was as expected most and best information about the lowest level of results – outputs in terms of services provided and products delivered. 80% of the evaluations could report satisfactory achievements. This meant that Norwegian CSOs delivered what they were expected to deliver in terms of specific project outputs. This information could to a large extent be retrieved from progress reports. This has also been the type of information that NORAD and other donors traditionally have asked for. More recently, NORAD has as most other donors changed its reporting requirements and demands more data and information about results.
- There were still few specific and measurable targets. Objectives were often unclear and targets often not measurable. CSOs seek “to contribute to”, “to improve, increase or enhance”, but it was rarely specified on what dimensions changes were expected to occur and how much. The objectives represented the guiding vision and intentions.
- The evaluations were focused on project benefits and not broader political and socio-economic impact and relevance. Almost all the evaluations were

project focused – with a few thematic studies or broader programme evaluations. There were also few efforts in the reports to change, broaden the project perspective and discuss the work of CSOS in a wider context.

- Results were described, but not explained. Most of the evaluations tried to document results – and covered as mentioned well the lower levels in the result chain, but the achievements were rarely analysed and explained. There was little analysis of why some projects succeeded, while others failed.
- Baselines and use of indicators were weakly developed. Few projects had carried out any baseline and end line studies in order to measure change over time more systematically. Only a small minority had used indicators for regular monitoring of changes and results and selection of indicators were not systematic part of the project design. Financial and activity monitoring was much more common.
- Capacity development – was weakly discussed. Capacity development or strengthening of civil society in the South is an important goal for Norwegian CSOs. They should not only channel funds, but assist in strengthening administrative and technical capacity among partners. More than 70% of the reports did not have any assessment of capacity development. The reports covering capacity building were largely positive, but how and to what extent the Norwegian NGO's represented “added value” in capacity strengthening – over and above being a donor - was not well explained and documented.
- Institutional sustainability – promising results. Institutional sustainability covers the Southern CSOs administrative and technical capacity - the ability to manage and plan, implement and follow up projects without external technical and managerial support – even if financial support was necessary. Approximately half of the evaluations could report about satisfactory institutional sustainability – while 20% still needed external support. This is an indication of successful capacity building and promising changes in the area of technical and administrative capability.
- Financial sustainability – extremely weak. While the institutional sustainability was found to be promising, financial sustainability was extremely weak (30%). Financial sustainability was found satisfactory in only 12% of the projects. Most of the reports did not assess issues of sustainability.

1.4 Can the evaluations be trusted?

The quality of the reports varied considerably. 45% of the reports were characterised as “good”, while 30% as “adequate” according to a list of quality criteria such as methodology, presentation, understanding of the intervention and quality of information. Several reports revealed major weaknesses as for instance: no terms of reference, executive summaries or questionnaires (instruments). It was also often unclear who had commissioned the evaluations and how findings and recommendations would be used. The large majority of the evaluators were men. 70 % of the team leaders were said to be external, but it was difficult to know how external/internal the team leader and the team were. The CSOs tend to recruit people they knew well in advance.

60% of the evaluations followed a systematic and structured approach (specifying steps and procedures for data collection and analysis), while in others it was difficult to assess the plausibility of findings and recommendations from lack of data and unsystematic collection of information. The majority of the reports had no financial analysis. There were also few efforts to extract lessons learned and provide strategic guidance based on findings from the evaluations. A more thorough discussion of relevance could have led to a discussion about to what extent objectives and strategies were still appropriate or were in need of change.

CSOs have often been critical of what they perceived as traditional donor evaluations – evaluations with too much emphasis on specific evaluation criteria, the use of external teams and low level of participation of target groups. In the sample, there were only a few examples of “alternative evaluations” - for instance using participatory methods. More innovative approaches were exceptional. Save the Children made efforts to involve children in evaluations. There were also examples where CSOs evaluated complex capacity building projects with systematic organisational assessment instruments.

Almost all evaluations (90%) used qualitative information, while 40% was not based on any systematic quantitative data. The most used method of data collection was interviews (90%), review of documents (70%) and observation (60%). 70% presented no baseline data and could not use a systematic before and after or with and without design. Change was primarily assessed through informed judgement.

1.5 Tracking wider impact

Norad’s recent Civil Society Panel provides a state of the art overview of current CSO evaluations. The report points to the fact that the first wave of studies examined the impact of individual discrete projects, especially the relationship between the aid inputs provided and the more immediate outputs, using project-specific data to try to draw out sketchy implications concerning wider impact.

Many of these early studies (undertaken from the mid-1980s onwards) were commissioned by donor agencies, which may explain in part why most CSOs were initially quite sceptical about the benefits of undertaking in-depth impact assessments of their projects. In successive “country evaluations” of NGO development activities, evaluators have remained largely reluctant to use this project-based evidence to draw wider conclusions about the overall impact of NGO development interventions at the country level.²

Increasingly over the last 10 to 15 years, however, more and more CSOs and in particular the larger ones have seen the value of evaluating and assessing the impact of their work. Most are now undertaking or commissioning their own evaluations, with the larger ones using external evaluators as well. Although this

² See Oakley and Folke (1999a and 1999b) and the most recent Norad NGO impact evaluation, Ternström Consulting AB (2011).

has led to an explosion of CSO impact assessments, the work has continued to be dominated by evaluation and assessment at the project level, with lively debate on the best methods and approaches and continuing recognition that the data typically remain inadequate for in-depth evaluations to be undertaken and robust conclusions to be drawn.

These NGO impact assessments are still largely output-focused, and even recent “country studies” largely comprise assessments of individual projects or clusters of projects, with evaluators reluctant to draw firm conclusions about wider or longer-term impact or the relationship between aid input and broader outcomes. At the same time, increasing attention is being given to examining the impact of CSO activities beyond the project, trying to assess the impact of the lobbying, advocacy, campaigning, policy and general influencing work of CSOs. These assessments, too, have tended to focus on discrete interventions and have not been used to make judgements about the wider or longer-term impact of these types of CSO interventions.³

Other important changes have been occurring, not least in the literature on the nature of the discourse about CSOs and development. These changes have profound implications for how one judges the impact of CSOs and the activities they support or promote. Thus, firstly, it has become quite common in recent years to talk about the role and impact of Civil Society Organisations (CSOs) in development, and of civil society more generally. For some, the term CSO has replaced the term NGO, though for many, the words NGO and CSO are used interchangeably.⁴ Secondly and related to this, contemporary development discourse speaks of the importance of “strengthening civil society”. This, in turn, has led to a massive increase in the funds northern agencies channel to southern agencies for the immediate purpose of building and strengthening the capacities and capabilities of CSOs in the south.

It has also been increasingly recognised that the effectiveness of NGO development work would be enhanced by improving the capacity and capabilities of the CSOs that implement projects and programmes. If this was the only or core purpose in strengthening southern agencies, then a key way to assess the impact of such activities would be to examine the difference this investment has made to the overall impact of the work of such agencies.

1.6 Methodological challenges

An accurate assessment of the contribution made by CSOs and civil society more generally to development outcomes, depends critically upon two related factors:

- The quality and reliability of the data which assessments will be based on
- The robustness of the methods used.

³ For an overview of these developments, see Riddell (2007), Riddell (2008: 259-324) and Riddell et al. (2008).

⁴ The dominant view in the literature is that the term CSO encompasses a wider group of organisations than the traditional NGOs. See, for instance, Van Rooy (1998) and Edwards (2004).

There are major challenges on both counts. The recent wider literature draws attention to major data problems. A long-standing concern about how to accurately assess the impact of discrete CSO and NGO projects has been the combined effect of common weaknesses: a lack of clarity concerning the precise objectives of projects and how they might best be assessed; poor or non-existent base-line data; inadequate monitoring and project completion reports; the low priority given to assessment and the related problems of inadequate in-house skills (Riddell et al., 1997). These concerns persist. For example, one of the key conclusions of the recent Norad evaluation of NGOs in East Africa was that “most projects lacked the data and information required to be able to measure changes in indicators for key results accurately” (Ternström Consulting AB, 2011: xvii, 50-66 and 76-7). As the wider impact of CSOs in part is based on extrapolating outwards from project impact data, this remains a fundamental problem for understanding the wider impact of CSO development efforts. All too often, attention is focused on what evaluations tell us, while far less attention is given to assessing the quality of the methods used to draw these conclusions. “Evidence” ought to be treated with scepticism – arguably even discounted entirely - unless one is sure that the methods used are robust enough to guarantee that sound conclusions can be drawn.

A related problem is that assessments of the wider impact of CSO development activities require not merely information on inputs and outputs, but also information on outcomes and the relationship between these. Poor quality data comprise major and recurrent problems that are highlighted in the recent wider literature. For example, the recent Dutch study on capacity development noted that because organisations gathered insufficient data and information on outcomes, it was impossible to judge whether capacity building initiatives had resulted in attainment of core organisational goals (IOB, 2011: 15-6) - a fundamental weakness.

Weak data add to other methodological challenges that face those trying to assess the link between CSO activities and wider development outcomes. The evaluation of discrete NGO and CSO projects faces challenges in trying to judge to what extent the outcomes achieved can be attributed to the project, when a range of other external factors also are likely to have influenced outcomes. It is possible, however, to address many of these challenges through a range of different methods, including in-depth evaluations and Randomised Controlled Trials (RCTs). The CSO sector is currently awash with debates and discussion about what methods to use and how appropriate and costly they are (see Woolcock, 2009 and Karlan, 2009).

Annex 7. Qualitative Comparative Analysis

The QCA analysis undertaken in this study has two main purposes: identifying the differences among the organizations that were hardly clear cut from the survey and other data collection activities; and explaining them. QCA is a tool for data analysis based on the dichotomization of a set of variables that, according to the working model built by the research, potentially explain an outcome. The yes/no variables thus constructed are more properly called conditions because rather than representing a varying quantity they refer to the presence / absence of a qualitative feature of the case. For the purpose of this study both the explanatory and outcomes conditions are organizational characteristics and behaviours, which are indicated in Table 1, along with their abbreviated name used in the software analysis of the data.

When QCA results are presented, capital letters have a different meaning from small letters: the capital letters indicate presence (of the condition), while the small letters indicate absence.

The last five conditions can be regarded as representing the broad, historical context the CSOs are embedded in; they refer to characteristics that are not easily changed in the short-medium term. A quick analysis of these contextual characteristics aimed at reducing complexity by identifying typologies has returned three organizational types, which are described below.

The first is exemplified by the configuration “FUNDPUB*FUNDNAT*MEMBER*FIELD”, which describes a member organization rooted on the Norwegian territory, but also having a field presence around the world as a Norwegian organization, or with affiliated partners, which draws more than most other organizations on Norwegian public national funds. The CSOs that are closer to this typology than to others are DIGNI (including its organizations) and Norwegian Church Aid.

The second is labeled “FUNDPUB*fundnat*orgs*FIELD”, which refers to an organizational type that draws more than most other organizations on international public funding, is rooted in the field as a Norwegian organization but cannot rely on any mother/sister organization. The CSOs that are closer to this typology than to others are Norwegian People’s Aid and Norwegian Refugee Council.

Table 1. Descriptors in the QCA Model

Descriptor	Values	Condition QCA name
Use of evaluation	for learning vs. for accountability	USE1
Most influential donor environment	national vs. international	DONOR
Type of evaluation use	instrumental vs. open and less systematic	USE2
Location of the evaluation function	high, low, both	UNIT
Main activity of evaluation unit	system development and on demand consultancy vs. training	ACTIV
Presence of a central evaluation budget	Yes/no	BUDG
Characteristic of the evaluation function	concentrated, i.e. dedicated M&E staff vs. dispersed i.e. staff working part-time in M&E	STAFF
The extent to which CSOs value independence rather than closeness to the activities and the context	Independence vs. closeness	INDCLO
Extent of board involvement	High/low	BOARD
M&E closely linked to planning and reporting	Yes/no	LINE
Characteristics of Evaluation Policy	explicit policy vs. policy-like instruments and documents	POL
Presence of management response system	Yes/no	MRESP
Monitoring of outcome indicators	Yes/no	MONOUT
Monitoring of project performance	Yes/no	MONIT
Funding sources	public by far the main source vs. considerable private sources	FUNDPUB
Funding location	mainly national by far vs. considerable international funding	FUNDNAT
Presence of mother or sister organizations	Yes/no	ORGS
Type of organization	member organization vs. private foundation	MEMBER
International Field Presence as a Norwegian organization (equivalent to not having a mother organization)	Yes/no	FIELD

The third and last is represented by the configuration “FUNDNAT*ORGS*MEMBER”, which describes a member organization drawing on Norwegian national (but not necessarily public) funding more than most others, which can rely on mother/sister organizations. The CSOs that are closer to the third typology than to others are Save The Children Norway and Norcross.

The distinction between these three types could be interesting because it allows us to identify specific behaviours that we can connect to the organizational type, in the sense that belonging to that type implies the behaviour. This doesn't mean that other types cannot adopt the behaviour, but just that other types do not guarantee its adoption (in causal terms: they are not “sufficient” for the behaviour). Below are the connections between organizational types and a number of behaviours. The table below shows how some of the variables combine with each other. However, the fact that there is a combination does not in itself allow for immediate generalization. What the combinations do is that they might point us in the direction of links that may need to be explored further.

Table 2. Issues/Connections that Could be Explored

Organization type	Organization characteristics	Organizational behaviour implied by the organization type
TYPE I	Member organization rooted on the Norwegian territory, having a field presence around the world as a Norwegian organization, drawing more than most other organizations on Norwegian public national funds	Absence of a Management Response System Influenced mainly by the National donor environment
TYPE II	Draws more than most other organizations on international public funding, is rooted in the field as a Norwegian organization but cannot rely on any mother/sister organization	Use of evaluations for learning rather than accountability Open and less systematic use of evaluations Main activity of Evaluation Unit is training
TYPE III	Member organization drawing on Norwegian national (but not necessarily public) funding more than most others, which can rely on mother/sister organizations and is not present in the field as a Norwegian organization	M&E closely linked to planning and reporting Influenced mainly by the International donor environment

Another model that was explored attempted to relate organizational behaviour specifically to composition of funding. Three new funding conditions were introduced: tendency to use national public funding (fundnatpub), tendency to use international public funding (fundint), and tendency to use private funding (fundpriv).

A number of types emerged, each connected to one funding typology:

Table 3. Relating Configuration to Funding Typology

Configuration	System Description	Funding Typology
LINE	Organizations that tend to look for private funding more than the others tend to have decision making closely connected to evaluations	FUNDPRIV (NRK, NSC)
USE1*use2	Organizations that tend to look for international public funding more than the others tend to use evaluation for learning rather than accountability and in an open and less systematic way rather than instrumentally	FUNDINT (NRC, NPA)
mresp	Organizations that tend to look for Norwegian public funding more than the others tend to lack a management response system	FUNDNATPUB (DIGNI, NCA)
LINE*mresp	Most of the organizations that tend not to look for international public funding use evaluations in a way that closely connects them to decision making and planning, but have no management response system.	Fundint (DIGNI, NCA, NRK, NSC)
USE1*use2*mresp	Most organizations that tend not to use private funding use evaluations for learning rather than accountability, their use is open and less systematic and they lack a management response system.	Fundpriv (DIGNI, NCA, NPA, NRC)
USE1*use2*LINE	Most organizations that tend not to use Norwegian public funding (or use it less compared to the others) use evaluations for learning rather than accountability, in an open and less systematic way, and at the same time their decision-making is closely connected to evaluations.	Fundnatpub (NRK, NRC, NPA, NSC)

Other organizational behaviours seem to be implied by other behaviours and organizational characteristics in the sense that, based on the data from this six organizations, they are sufficient to guarantee the behaviour. The table below summarizes the implications of organizational characteristics and behaviour on organizational behaviour.

Table 4. Implications of Organisational Behaviour

Organizational behaviour and characteristics	Implications
More private funding than other CSOs	Decision-making closely connected to evaluations
Less private funding than other CSOs	Absence of a management response system
Less international public funding than other CSOs	Evaluations closely connected to decision-making
	Absence of a management response system
More Norwegian public funding than other CSOs	Absence of a management response system
Less Norwegian public funding than other CSOs	Evaluations closely connected to decision-making
Use of evaluation for accountability rather than learning	Absence of a management response system
M&E loosely linked to planning and reporting	Open and less systematic use of evaluations
	Use of evaluation for learning rather than accountability
	Policy like instruments and documents rather than an explicit evaluation policy
	Absence of a management response system
Instrumental use of evaluation	Evaluation function dispersed rather than concentrated
	Absence of a management response system
Open and less systematic use of evaluations	Use of evaluation for learning rather than accountability
Policy like instruments and documents rather than an explicit evaluation policy	Absence of a management response system
Use of evaluation for learning rather than accountability	Open and less systematic use of evaluations

The overall picture that would seem to emerge sees CSOs that are less dependent on Norwegian public funding to be more ready than others to implement and use a management response system and to connect evaluations more closely to decision-making.

EVALUATION REPORTS

- 7.99 Policies and Strategies for Poverty Reduction in Norwegian Development Aid
- 1.00 Review of Norwegian Health-related Development Cooperation 1988–1997
- 2.00 Norwegian Support to the Education Sector. Overview of Policies and Trends 1988–1998
- 3.00 The Project “Training for Peace in Southern Africa”
- 4.00 En kartlegging av erfaringer med norsk bistand gjennomfrivillige organisasjoner 1987–1999
- 5.00 Evaluation of the NUFU programme
- 6.00 Making Government Smaller and More Efficient. The Botswana Case
- 7.00 Evaluation of the Norwegian Plan of Action for Nuclear Safety Priorities, Organisation, Implementation
- 8.00 Evaluation of the Norwegian Mixed Credits Programme
- 9.00 “Norwegians? Who needs Norwegians?” Explaining the Oslo Back Channel: Norway’s Political Past in the Middle East
- 10.00 Taken for Granted? An Evaluation of Norway’s Special Grant for the Environment
- 1.01 Evaluation of the Norwegian Human Rights Fund
- 2.01 Economic Impacts on the Least Developed Countries of the Elimination of Import Tariffs on their Products
- 3.01 Evaluation of the Public Support to the Norwegian NGOs Working in Nicaragua 1994–1999
- 3A.01 Evaluación del Apoyo Público a las ONGs Noruegas que Trabajan en Nicaragua 1994–1999
- 4.01 The International Monetary Fund and the World Bank Cooperation on Poverty Reduction
- 5.01 Evaluation of Development Co-operation between Bangladesh and Norway, 1995–2000
- 6.01 Can democratisation prevent conflicts? Lessons from sub-Saharan Africa
- 7.01 Reconciliation Among Young People in the Balkans An Evaluation of the Post Pessimist Network
- 1.02 Evaluation of the Norwegian Resource Bank for Democracy and Human Rights (NORDEM)
- 2.02 Evaluation of the International Humanitarian Assistance of the Norwegian Red Cross
- 3.02 Evaluation of ACOPAMA n ILO program for “Cooperative and Organizational Support to Grassroots Initiatives” in Western Africa 1978 – 1999
- 3A.02 Évaluation du programme ACOPAMA n programme du BIT sur l’« Appui associatif et coopératif aux Initiatives de Développement à la Base » en Afrique de l’Ouest de 1978 à 1999
- 4.02 Legal Aid Against the Odds Evaluation of the Civil Rights Project (CRP) of the Norwegian Refugee Council in former Yugoslavia
- 1.03 Evaluation of the Norwegian Investment Fund for Developing Countries (Norfund)
- 2.03 Evaluation of the Norwegian Education Trust Fund for African the World Bank
- 3.03 Evaluering av Bistandstorgets Evalueringsnettverk
- 1.04 Towards Strategic Framework for Peace-building: Getting Their Act Together. Overview Report of the Joint Utstein Study of the Peacebuilding.
- 2.04 Norwegian Peace-building policies: Lessons Learnt and Challenges Ahead
- 3.04 Evaluation of CESAR’s activities in the Middle East Funded by Norway
- 4.04 Evaluering av ordningen med støtte gjennom paraplyorganisasjoner. Eksempifisert ved støtte til Norsk Misjons Bistandsnemnd og Atlas-alliansen
- 5.04 Study of the impact of the work of FORUT in Sri Lanka: Building Civil Society
- 6.04 Study of the impact of the work of Save the Children Norway in Ethiopia: Building Civil Society
- 1.05 –Study: Study of the impact of the work of FORUT in Sri Lanka and Save the Children Norway in Ethiopia: Building Civil Society
- 1.05 –Evaluation: Evaluation of the Norad Fellowship Programme
- 2.05 –Evaluation: Women Can Do It – an evaluation of the WCIDI programme in the Western Balkans
- 3.05 Gender and Development – a review of evaluation report 1997–2004
- 4.05 Evaluation of the Framework Agreement between the Government of Norway and the United Nations Environment Programme (UNEP)
- 5.05 Evaluation of the “Strategy for Women and Gender Equality in Development Cooperation (1997–2005)”
- 1.06 Inter-Ministerial Cooperation. An Effective Model for Capacity Development?
- 2.06 Evaluation of Fredskorpset
- 1.06 – Synthesis Report: Lessons from Evaluations of Women and Gender Equality in Development Cooperation
- 1.07 Evaluation of the Norwegian Petroleum-Related Assistance
- 1.07 – Synteserapport: Humanitær innsats ved naturkatastrofer. En syntese av evalueringsfunn
- 1.07 – Study: The Norwegian International Effort against Female Genital Mutilation
- 2.07 Evaluation of Norwegian Power-related Assistance
- 2.07 – Study Development Cooperation through Norwegian NGOs in South America
- 3.07 Evaluation of the Effects of the using M-621 Cargo Trucks in Humanitarian Transport Operations
- 4.07 Evaluation of Norwegian Development Support to Zambia (1991 - 2005)
- 5.07 Evaluation of the Development Cooperation to Norwegian NGOs in Guatemala
- 1.08 Evaluation: Evaluation of the Norwegian Emergency Preparedness System (NOREPS)
- 1.08 Study: The challenge of Assessing Aid Impact: A review of Norwegian Evaluation Practise
- 1.08 Synthesis Study: On Best Practise and Innovative Approaches to Capacity Development in Low Income African Countries
- 2.08 Evaluation: Joint Evaluation of the Trust Fund for Environmentally and Socially Sustainable Development (TFESSD)
- 2.08 Synthesis Study: Cash Transfers Contributing to Social Protection: A Synthesis of Evaluation Findings
- 2.08 Study: Anti- Corruption Approaches. A Literature Review
- 3.08 Evaluation: Mid-term Evaluation of the EEA Grants
- 4.08 Evaluation: Evaluation of Norwegian HIV/AIDS Responses
- 5.08 Evaluation: Evaluation of the Norwegian Research and Development Activities in Conflict Prevention and Peace-building
- 6.08 Evaluation: Evaluation of Norwegian Development Cooperation in the Fisheries Sector
- 1.09 Evaluation: Joint Evaluation of Nepal’s Education for All 2004–2009 Sector Programme
- 1.09 Study Report: Global Aid Architecture and the Health Millennium Development Goals
- 2.09 Evaluation: Mid-Term Evaluation of the Joint Donor Team in Juba, Sudan
- 2.09 Study Report: A synthesis of Evaluations of Environment Assistance by Multilateral Organisations
- 3.09 Evaluation: Evaluation of Norwegian Development Cooperation through Norwegian Non-Governmental Organisations in Northern Uganda (2003–2007)
- 3.09 Study Report: Evaluation of Norwegian Business-related Assistance Sri Lanka Case Study
- 4.09 Evaluation: Evaluation of Norwegian Support to the Protection of Cultural Heritage
- 4.09 Study Report: Norwegian Environmental Action Plan
- 5.09 Evaluation: Evaluation of Norwegian Support to Peacebuilding in Haiti 1998–2008
- 6.09 Evaluation: Evaluation of the Humanitarian Mine Action Activities of Norwegian People’s Aid
- 7.09 Evaluation: Evaluation of the Norwegian Programme for Development, Research and Education (NUFU) and of Norad’s Programme for Master Studies (NOMA)
- 1.10 Evaluation: Evaluation of the Norwegian Centre for Democracy Support 2002–2009
- 2.10 Synthesis Study: Support to Legislatures
- 3.10 Synthesis Main Report: Evaluation of Norwegian Business-related Assistance
- 4.10 Study: Evaluation of Norwegian Business-related Assistance South Africa Case Study
- 5.10 Study: Evaluation of Norwegian Business-related Assistance Bangladesh Case Study
- 6.10 Study: Evaluation of Norwegian Business-related Assistance Uganda Case Study
- 7.10 Evaluation: Evaluation of Norwegian Development Cooperation with the Western Balkans
- 8.10 Evaluation: Evaluation of Transparency International
- 9.10 Study: Evaluability Study of Partnership Initiatives
- 10.10 Evaluation: Democracy Support through the United Nations
- 11.10 Evaluation: Evaluation of the International Organization for Migration and its Efforts to Combat Human Trafficking
- 12.10 Evaluation: Real-Time Evaluation of Norway’s International Climate and Forest Initiative (NICFI)
- 13.10 Evaluation: Real-Time Evaluation of Norway’s International Climate and Forest Initiative. Country Report: Brasil
- 14.10 Evaluation: Real-Time Evaluation of Norway’s International Climate and Forest Initiative. Country Report: Democratic Republic of Congo
- 15.10 Evaluation: Real-Time Evaluation of Norway’s International Climate and Forest Initiative. Country Report: Guyana
- 16.10 Evaluation: Real-Time Evaluation of Norway’s International Climate and Forest Initiative. Country Report: Indonesia
- 17.10 Evaluation: Real-Time Evaluation of Norway’s International Climate and Forest Initiative. Country Report: Tanzania
- 18.10 Evaluation: Real-Time Evaluation of Norway’s International Climate and Forest Initiative
- 1.11 Evaluation: Results of Development Cooperation through Norwegian NGO’s in East Africa
- 2.11 Evaluation: Evaluation of Research on Norwegian Development Assistance
- 3.11 Evaluation: Evaluation of the Strategy for Norway’s Culture and Sports Cooperation with Countries in the South
- 4.11 Study: Contextual Choices in Fighting Corruption: Lessons Learned
- 5.11 Pawns of Peace. Evaluation of Norwegian peace efforts in Sri Lanka, 1997–2009
- 6.11 Joint Evaluation of Support to Anti-Corruption Efforts, 2002–2009
- 7.11 Evaluation: Evaluation of Norwegian Development Cooperation to Promote Human Rights
- 8.11 Norway’s Trade Related Assistance through Multilateral Organizations: A Synthesis Study
- 9.11 Activity-Based Financial Flows in UN System: A study of Select UN Organisations Volume 1 Synthesis Volume 2 Case Studies
- 10.11 Evaluation of Norwegian Health Sector Support to Botswana
- 1.12 Mainstreaming disability in the new development paradigm. Evaluation of Norwegian support to promote the rights of persons with disabilities.
- 2.12 Hunting for Per Diem. The uses and Abuses of Travel Compensation in Three Developing Countries
- 3.12 Evaluation of Norwegian Development Cooperation with Afghanistan 2001–2011
- 4.12 Evaluation of the Health Results Innovation Trust Fund
- 5.12 Real-Time Evaluation of Norway’s International Climate and Forest Initiative. Lessons Learned from Support to Civil Society Organizations.
- 6.12 Facing the Resource Curse: Norway’s Oil for Development Program

Norad
Norwegian Agency for
Development Cooperation

Postal address
P.O. Box 8034 Dep. NO-0030 OSLO
Visiting address
Ruseløkkveien 26, Oslo, Norway

Tel: +47 23 98 00 00
Fax: +47 23 98 00 99

No. of Copies: 300
postmottak@norad.no
www.norad.no