1

Introduction

This book is the result of an action-research exercise jointly undertaken by a number of international and local non-government organisations (NGOs) based in four continents. The research was initiated by Oxfam from Great Britain and Novib from the Netherlands (hereafter referred to as Oxfam and Novib). The book is an attempt to share with others the lessons learned from that experience, as well as some of the challenges that arise from thinking about and practising impact assessment.

It is aimed particularly at practitioners, and evaluation specialists may find that I skim over some of the more theoretical issues. This is not because these are unimportant, but because our research showed that there is a need to demystify the subject. The aim is to make impact assessment accessible without being simplistic about it. The book also explores those elements of impact assessment which look beyond the project level at organisational processes. I hope therefore that those involved in the development of monitoring and evaluation systems at an organisational level will find this book helpful. However, it is essentially about the insights gained from the case studies, rather than an attempt to cover the abundant literature on the subject.

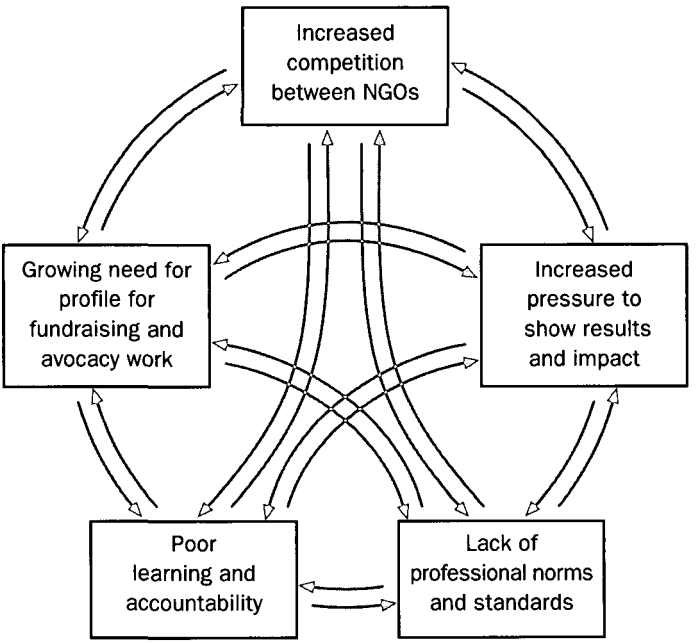
Impact assessment – making the case for development aid

Despite the statistics in recent UNDP Human Development Reports and in the World Bank's annual reports, which record a marked improvement in a number of indicators of human well-being, the scale of world poverty remains a scandal which shames us all. In many parts of the world inequality, insecurity, and conflict are growing at alarming rates. Bilateral development aid has had its critics for many years, but during the past decade we have also seen a growing number of critiques of NGOs (Smillie 1995, de Waal 1996, Sogge 1996). These critiques together describe a vicious circle which the

NGO sector, particularly in the North, faces and which it has helped to create. This circle has five main elements (see Figure 1.1): there is increasing pressure on NGOs to demonstrate results and the impact of their work. Moreover, there is increased competition between NGOs, and a growing need for a high profile and press coverage in order to raise funds and to facilitate advocacy work. Poor institutional learning and weak accountability mechanisms are characteristic of many NGOs, which both leads to and is the result of the absence of professional norms and standards.

These elements combine to produce a growing gap between the rhetoric of agencies and the reality of what they achieve. They also fuel growing scepticism about the value of aid, and lessen trust between agencies — some argue that agreements and partnerships based on shared values have been replaced by bureaucratic trust based on plans, budgets, and accounts. Moreover, the elements of this vicious circle to perpetuate the tired old image of aid going from donor to 'victim', and a view of development as something that is done to other people, far away. This analysis, although based on little empirical evidence, holds a certain truth which NGOs ignore at their peril.

Figure 1.1: The vicious circle



One recent study of NGO impact which had a stronger empirical base concluded that the true impact of NGO development work remained unclear and that there was little consensus on which tools and methods were the most appropriate to find out. This study, undertaken by Riddell et al. (1997) for the OECD/DAC Expert Group on Evaluation, reviewed 60 reports covering 240 projects in 26 countries, and undertook 13 country case studies.

Both the critique outlined above (mainly of Northern-based NGOs) and the findings of this OECD study point to the inadequacy of most current attempts to promote institutional learning, impact assessment, and greater accountability in the NGO sector — accountability to those who NGOs seek to support, as well as to those who fund this work. In a climate of increased competition, individual organisations and the sector as a whole tend to exaggerate the case for support just as their opponents tend to exaggerate the case against. This can have two consequences, the dangers of which have been pointed out for some time (Cassen 1986, Riddell 1987). First, support for development aid depends on the public's belief in its effectiveness. The moral case for providing support rests upon its achieving its objectives. However, a reluctance to admit that the effectiveness of much of that is done is unpredictable and difficult to assess, makes not just NGOs, but also international co-operation programmes, vulnerable to public criticism and the odd polemic attack.

Second, those making the case for co-operation must not create the belief that aid flows constitute the sole, or even principal, means available to donors and governments of improving the welfare of people living in poverty. Often, changes in policy and practice, for example improved terms of trade or greater debt relief, may be more beneficial.

In the long term, the case for aid can only be sustained by more effective assessment and demonstration of its impact, by laying open the mistakes and uncertainties that are inherent in development work, and by an honest assessment of the comparative effectiveness of aid vis-à-vis changes in policy and practice. The research project that forms the basis of this book was designed to make a contribution to this process, as well as to the institutional learning of the agencies involved. A broad range of intermediary and grassroots organisations in nine countries participated alongside the Northern NGOs Oxfam and Novib. More specifically, we had the following aims:

- to develop greater clarity about the key elements of impact assessment;
- to explore how unequal levels of power and participation of the various stakeholders can affect an impact assessment process;
- to test a range of approaches to undertaking impact assessment;
- to look at the organisational context in which impact assessments take place.

These issues are discussed in more depth in Chapters 2 and 3, and I will return to them in the final chapter.

The structure of this book

The second half of this Introduction describes the case studies and the organisations involved in preparing them. The case studies form the foundation of the book, and I have attempted to retain their richness and diversity. Chapter 2 discusses Oxfam and Novib's overall approach to impact assessment. It explores issues relating to ethics and participation, and emphasises the importance of ensuring that gender and other aspects of difference are embedded into all processes of impact assessment. Chapter 2 also sets out some of the assumptions and hypotheses which emerged from a literature review at the beginning of the research and which were subsequently tested in the case studies. For example, we confirmed the hypothesis that impact assessment must be seen as an integral part of development work throughout the project- or programme-cycle. Finally, this chapter summarises some of the key dilemmas regarding how to attribute any observed change to an intervention, and how to synthesise diverse and sometimes contradictory findings.

Although Oxfam and Novib feel that the distinction between emergency, development, and advocacy work is increasingly redundant, these terms do provide a convenient short-hand for different aspects of NGO work. So although Chapters 3 to 6 are structured along these lines, there is much overlap between them. Chapter 3 explores how the case studies were designed for assessments in single communities, across several communities, and across a range of projects. It also gives some guidance about basic operational matters which must be addressed in designing an impact assessment, such as how to go about sampling, what to do if there is no baseline data, and how to cross-check results. It also looks at how some of the most difficult issues such as setting indicators and measuring attribution were tackled in our case studies. In Chapter 4, I look at the wide range of different tools, methods, and approaches used as well as the problems encountered in the case studies. The reader will find summaries of the lessons learned about various kinds of tools and methods such as surveys, interviews and workshops, direct observation, participatory tools, and case studies. Chapters 3 and 4 make up the core of the book. Many of the chapters that follow refer back to some of the basic lessons described in these chapters.

Chapter 5 on impact assessment and emergencies refers mainly to one of the case studies as well as to some recent material. It explores some of the specific difficulties and challenges of undertaking impact assessment in crisis situations and compares them with the lessons described in previous chapters.

Because of the limited number of case studies directly dealing with emergency situations, this chapter should be seen as a preliminary exploration of the issue; it mainly focuses on relief interventions. The chapters on advocacy and organisations and impact assessment are also relevant to work in emergency situations. Recent events such as Hurricane Mitch and the Kosovo crisis may offer opportunities to work on more longitudinal and holistic approaches to emergency situations and their assessment.

In Chapter 6, which focuses on impact assessment and advocacy, I outline current approaches to this growing field of NGO activity. Drawing on the case studies and other experiences, I discuss both how to assess the impact of advocacy and how to use impact assessment in advocacy work.

Chapter 7 considers how we can assess the impact of projects and programmes on development organisations, and at how these organisations manage impact assessment processes. It also considers the impact that development organisations have on people's lives.

I draw some conclusions from this research in Chapter 8 and indicate areas which NGOs need to invest in if they are serious about impact assessment. I end by noting key policy implications that emerge from the case-study findings, and thinking about what these might mean for the future of NGOs. The appendices include a bibliography, further details of the organisations and individuals involved in the case studies as well as a list of acronyms.

The case studies

When Oxfam and Novib embarked on this study in 1994/5, we undertook an initial literature review as well as a rapid survey of what sort of work on impact assessment was happening in a number of countries where we supported projects and organisations. A number of reports resulted (Hopkins 1995 a and b, Dawson 1995 a, b and c) which gave a useful overview of existing concepts and tools as well as suggesting a number of principles and hypotheses for impact assessment, which are discussed in Chapter 2.

However, the literature review concluded that a more 'hands-on' approach was needed to test a range of approaches in diverse contexts with different types of organisations.

The most important challenge is to promote, through case studies, a systematic application of the methodologies on impact assessment. There has been in the past few years a significant production of tools of project evaluation ... but what is really lacking is both a vigorous and systematic application of techniques and methodologies and to learn from experience. (Hopkins 1995a)

Oxfam and Novib then approached a number of partner organisations and their own field staff to explore the potential for undertaking action-research on impact assessment and developing a number of case studies. As is common with this sort of exercise, willingness to participate became the most important criterion. So although we also wanted to ensure a good geographical balance, to include emergency and advocacy as well as long-term 'development' work, and to balance forward-looking with retrospective studies, these criteria were inevitably secondary.

In the end, case studies were carried out in Africa (Ghana, Kenya, Zimbabwe and Uganda), in South Asia (Pakistan, Bangladesh and India), in Latin America (El Salvador) and in Europe (the United Kingdom). They represent a balance of prospective work, mid-term assessments of ongoing work, and retrospective reviews — although it became clear during the exercise that these distinctions are in reality often blurred. Unfortunately, only two of the studies can be described as assessing an emergency programme. However, some of the others deal with work that emerged from responses to emergency or crisis situations. The case studies do not include rescue or acute emergency operations, but I refer to some material on assessments in these circumstances in Chapter 5.

Given the nature of the research and the diversity of programmes and partners involved, the case studies did not follow a set methodology. We adopted an open and flexible approach which was designed to encourage innovation, creativity, and 'learning by doing'. We hoped that this might lead to greater variation in the approaches, tools, and methods used and maximise the potential for learning across the case studies. A few of the original participants had to drop out due to unforeseen circumstances and were be replaced as other opportunities emerged: a fixed method would not have permitted such adaptation. On the other hand, this flexibility has made the job of synthesising the experiences all the more challenging!

Who was involved and why?

Table 1.1 gives an overview of the organisations involved in the case studies. These range from BRAC in Bangladesh, which has a budget of some $200 million, 20,000 full-time staff and works with nearly 2 million people, to the Matson Neighbourhood Project in Gloucester, UK, which has a budget of £300,000, 16 staff (eight of whom are part-time), and covers a single housing estate of 6,000 people. Information is provided for Oxfam GB and Novib as a whole as well as for specific country programmes. The difference in size and scope of the participating organisations is reflected in the size and scope of the impact assessment processes they have undertaken, as well as in the

human and financial investments they have been able to make over and above the financial support of Oxfam or Novib. In all but one case (Matson), either Oxfam or Novib are — or have been — funding these organisations, in some cases for many years.

Given this relationship, it is of course possible that participants felt somewhat 'obliged' to take part in the research or hoped that it would have a favourable effect on their relationship with Oxfam and Novib. Some of the tensions and difficulties that emerged during the case studies can probably be attributed to this. Nevertheless the main reasons of agencies for participating can be summarised as follows:

- to learn more about impact assessment;
- to assess and improve performance, and plan for the future;
- to integrate learning from the research into management, monitoring and evaluation, and reporting systems;
- to share experiences of impact assessment with others;
- to improve the relationship with both communities and donors;
- to motivate staff and the communities supported;
- Three areas of consensus can therefore be defined as explicit reasons for involvement: learning, performance, and accountability.

Overview of the case studies

This publication cannot look in great depth at the background to the case studies or the context in which these projects and programmes developed. Below is a short summary of each of the projects, which is intended as a quick reference guide.

Institutional capacity-building in Northern Ghana (supported by Oxfam GB, Integrated Social Development Centre, Northern Ghana Development Network)

This programme was set up in 1995/6 to strengthen the capacity of more than 45 NGOs and community-based organisations (CBOs) which make up the Northern Ghana Development Network (NGDN). The project, undertaken by Oxfam in collaboration with the Integrated Social Development Centre (ISODEC) and the Northern Ghana Development Network, seeks to assess the impact of the institutional capacity-building programme on poverty in this region by establishing a baseline against which future changes in poverty levels and in organisational development can be measured. In doing so, past trends are also being identified. Changes will be tracked at regular intervals as the programme progresses, at the individual, community, and organisational level.

Table 1.1: Organisational profiles of case-study participants (1999 figures)

Country	Ghana	India	Kenya	Uganda	Pakistan
Organisation	Oxfam GB in Ghana, ISODEC and NGDN	CYSD	Oxfam GB in Wajir	Oxfam GB in Ikafe	Oxfam GB in Pakistan
Type of organisation	Local NGO and CBOs	Local NGO	International NGO and CBO	International NGO	International NGO
Established	1994	1982	1980	1994	1989
Number of staff	15 (Oxfam) 60 (ISODEC)	126	14 (local)	150	19
Number of volunteers	–	120	–	–	2
Income (£/US$)	£249,839	£620,000	£300,000	£2.5m	£350,000
Size of population they seek to benefit	50,000	43,646	40,000	55,000	50,000
Oxfam/ Novib funded	✓	✓	✓	✓	✓

Organisation	Oxfam	Novib
Type of organisation	International NGO	International NGO
Established	1942	1956
Number of staff	2,800	237 full-time equivalents
Number of volunteers	27,000	300
Income	£100m	£85m

UK	Bangladesh			El Salvador	Zimbabwe
Matson NP	Proshika	BRAC	NK GSS	CORDES	ENDA - ZW
Local NGO and CBO	Local NGO	Local NGO	Local NGOs	Local NGO	Local NGO
1988	1976	1972	NK 1974 GSS 1983	1988	1983
16 (8 part-time)	3,783	51,442 (31,009 part-time)	5,296 (GSS 1997)	112	45
36	–	–	–	–	–
£300,000	£53m	£125m	N.A.	£1.25m	N.A.
6,000	8,639,180	2m	NK: 112,208 GSS: 114,000	66,000	2m
✗	✓	✓	✓ GSS ✗ NK	✓	✓

This approach to impact assessment (often referred to as the longitudinal approach) helps us understand the relationship between these levels. It is hoped that the member organisations of NGDN will use the methods developed in this research for their own impact assessment needs.

The methods used in this study include a literature review, participatory rural appraisal (PRA) methods, repeated individual interviews, and organisational assessment exercises.

Integrated rural and tribal development programmes in Orissa, India (Centre for Youth and Social Development, Novib)

The Centre for Youth and Social Development (CYSD), supported by Novib, is exploring both retrospective and longitudinal approaches to assess the impact of their programmes.

The Integrated Rural Development Programme started in 1988 and works mainly with tribal people who migrated from neighbouring states to Orissa 30 to 40 years ago. It has focused on poverty reduction, stimulating alternative sources of income, building community-based organisations, watershed management, and promoting sustainable agriculture in 21 villages made up of some 869 households. CYSD is now looking back at results. This programme is referred to as CYSD 1 in the book.

The more recently established Integrated Tribal Development Programme, referred to as CYSD 2, also concentrates on several tribal communities who belong to the poorest groups in the State of Orissa. The project works with 1,787 households, constituting a total population of 8,935, and with the 24 CBOs that exist in the area. Areas of work include education, health, environment and livelihood issues, participation in local government institutions and cultural communication. CYSD is establishing a longitudinal study to assess the impact of this new programme as it progresses. The design of this study builds on the learning from CYSD 1.

Both studies have used a wide range of methods including reviews of secondary data, PRA tools, focus-group discussions, and case studies.

Pastoralist development programme in north-east Kenya (Wajir Pastoral Development Project, Oxfam GB)

Oxfam has provided relief services and other forms of assistance to nomadic communities in Kenya's Wajir District since 1984. Since 1993, Oxfam development assistance has been channelled through the Wajir Pastoral Development Project (WPDP), a nine-year programme consisting of three three-year phases. The first phase, which ran from July 1994 to July 1997, reached an estimated 40,000 beneficiaries in Wajir Town and Wajir Bor

division. The aim of the WPDP is to reduce poverty and vulnerability of both pastoralists and settled communities by promoting sustainable livelihoods and creating opportunities and improving the conditions for self-reliance.

The work referred to in this book relates to an economic impact assessment exercise which has built upon continuous longitudinal monitoring and a more general mid-term review. It therefore involves a retrospective assessment of impact as well as a prognosis of likely future impact. The study used a questionnaire, focus-group discussions, case studies, and a range of participatory tools and methods.

A refugee settlement in Northern Uganda (Oxfam GB in Ikafe)

The Ikafe programme in Arua District of Uganda was developed in order to support about 55,000 Sudanese refugees who were relocated from nearby Koboko following security problems in 1994. Oxfam GB as the main implementing agency took on tasks ranging from infrastructure development, registration, health-care provision and management, land allocation, and food and water distribution, to aspects of community development, forestry, and income-generation. Oxfam's response attempted to incorporate a long-term, 'developmental' approach into an emergency response. For example, refugees are settled in small, dispersed groups, and agricultural land is allocated to enable them to develop a certain level of food self-sufficiency. Several other agencies work with the refugees and there is support from a number of donors.

This case study used a retrospective review (mainly based on participatory appraisal methods and reviews of project records) which was undertaken by a team made up of external facilitators, project staff, and representatives of the refugee and host populations. One objective of the study was to examine the various stakeholders' distinct perspectives of impact and progress, and to bring the groups together in meetings and 'assemblies' to discuss and negotiate their views.

Project impact assessment in Pakistan (Oxfam GB in Pakistan)

Oxfam in Pakistan supports a range of projects and organisations through funding and non-funding support. This impact assessment work reviewed a range of Oxfam-supported small-scale projects. It tested a systematic method of conducting group interviews with project participants to ask about non-economic impact. It also tested a checklist of indicators that look at broad dimensions of human development, as well as a method for cross-project comparison. In addition, the researchers used a specific process for assessing economic impact and making comparisons between economic and non-economic impact.

This method has been tested with different Oxfam staff as well as other NGOs in Pakistan. A particular effort was made to ensure that the process did not take up too much time for both staff and the local men and women involved.

Community-managed services in Gloucester, UK (Matson Neighbourhood Project)

The Matson Neighbourhood Project, set up in 1990, works on the largest housing estate in the town of Gloucester, UK. The project runs a range of community-managed services such as education and training, a community shop, an advice centre, a mental health drop-in club, and a drug-counselling service. This review of its ongoing work involved a relatively unstructured process of participant observation and key informant interviews, carried out by an Indian community development specialist over a period of eight-weeks. First and foremost, the study looked at change within Matson through the eyes of local people and project staff. The study then explored how important these changes were for local people as well as what they attributed those changes to.

Economic and social empowerment in Bangladesh (Proshika)

Proshika is one of the largest NGOs in Bangladesh and promotes economic and social empowerment of the poor through a wide variety of interventions in more than 11,500 villages. This case study covers two interrelated components: a large-scale survey comparing 18,00 households, half of them supported by Proshika and the other half not supported by Proshika; and a participatory impact study undertaken to explore discrepancies arising from the initial survey. Subsequently, the participatory study was also used to develop a new impact assessment methodology for future studies, to explore and document new indicators, and to improve staff skills in participatory planning, monitoring, and evaluation.

Rural development work in Bangladesh (Bangladesh Rural Advancement Committee)

BRAC started in 1972 as a small relief and rehabilitation organisation in the war-ravaged, post-liberation Bangladesh. Today it is the largest multi-dimensional rural development organisation in the country. Its main goals are to alleviate poverty and empower rural people living in poverty. This case study focuses on a follow-up impact assessment (IAS II) of its rural development programme which aims to reach some 2.5 million people. The original study (IAS I), BRAC's first comprehensive impact assessment, was undertaken in 1993/4. This follow-up involved a large household survey involving both households supported by BRAC and 'non-BRAC households', as well as qualitative and case-study research with 25 village organisations, and the development of village profiles.

Social mobilisation in Bangladesh (Nijera Kori, Gonoshahajjo Sangstha)

NGOs in Bangladesh mainly use group-based approaches to development. However, for many NGOs this is a means to individual empowerment rather than collective action. This study, funded by Oxfam, reviewed group formation undertaken by social mobilisation organisations, and the impact of their work. It involved in-depth work with 17 village groups, a study of over 70 individual case histories, and an analysis of two specific campaigns undertaken by a wider range of groups. The study developed specific tools for assessing institutional development and sustainablity, such as a leadership-type analysis and indices of group cohesion and NGO involvement.

Post-conflict reconstruction in El Salvador (Fundación para la Cooperación y el Desarrollo Comunal de El Salvador)

The Foundation for Co-operation and Community Development in El Salvador (CORDES) was created during the last years of the civil war of 1979-91 by the mandate of the communities which it works with. These include demobilised combatants and repatriated and marginal communities from the areas most affected by the armed conflict. Projects now concentrate on financing and marketing agricultural production, and institutional strengthening. Supported by Novib, CORDES is looking back to assess the impact of its work in three communities, and to develop an impact assessment method with those communities and within its own organisation.

The study involved the development of a systematic process of reconstructing project history and then relating this to the current situation (working with a small number of communities in a participatory way).

Livelihood support projects in Zimbabwe (Environmental and Development Activities Zimbabwe)

The primary goal of ENDA-Zimbabwe is to assist disadvantaged communities to generate income and wealth through their sustainable use of natural resources and increased capacity to raise their standard of living. This retrospective study assessed the impact of the Zimbabwe Seeds Action Network (ZSAN) Phase 1 and the Chivi/ Zvishavane Indigenous Woodland Management Demonstration Project. Both these projects ran from 1985-92, and the studies therefore had to go back to communities some years after the projects had ended. The studies used a literature review, key informant interviews, questionnaires, and focus-group discussions, but also attempted to explore changes in other institutions which may have been influenced by the projects.

Table 1.2: Overview of case studies

Country	Ghana	India	Kenya	Uganda	Pakistan
Organisation	ISODEC and NGDN	CYSD	Oxfam GB in Wajir	Oxfam GB in Ikafe	Oxfam GB
Scale of assessment	3 villages 3 organisations	2 integrated programmes	Programme-wide	1 refugee-affected area	Variety of micro-projects and organisations
Stage of assessment	Base-line analysis and review	Ongoing	Ongoing	Ongoing	Ongoing
Year project started	1995	1989	1992	1994	various
Beneficiaries	170 households	1. 869hh 2. 1787hh	40,000	55,000	5,000
Duration of assessment	2 years	1 year	2 weeks 2 weeks	6 weeks	1-3 days per micro-project
Cost	£35,000	£6,404	£18,000	£19,000	£5,000
Staff involved	2 half-time 2 part-time staff	7 full-time 8 part-time staff	3 consultants 7 part-time staff	2 leaders 7 full-time 50 part-time staff	Advisor: 12 weeks 4 staff part-time: 8 person weeks
Support	IDS, UK; World Neighbours; Oxfam GB	Novib workshop; external advisor	DfID; ITDG; REMPAI consultants	Oxfam GB	External advisor

How the case studies differ

Differences in scale

The most obvious difference between the studies is the scale of the projects or programmes being assessed. One aims to reach a single community of 300 people, another involves more than 68,000 village organisations. Although this mirrors the organisational differences outlined in Table 1.1, it is not as simple as this: a large-scale programme, such as Oxfam GB's in Pakistan, can be made up of a large number of micro-projects. In the case of the Matson Neighbourhood Project, the organisation *is* the project; in this sense,

UK	Zimbabwe	Bangladesh			El Salvador
Matson NP	ENDA – ZW	NK GSS	BRAC	Proshika	CORDES
1 community 1 organisation	2 projects 5 sites	17 groups	1 thematic programme	1 thematic programme	3 communities
Ongoing	Review (4 years after the project ended)	Base-line and review	Ongoing	Ongoing	Review (4 years after the project ended)
1990	1985	various	1986	1993	1990
6,000	120,000	9 villages	2.5m (1998)	11,530 villages	299
8 weeks	1 year	18 months	1 year	2 years	1 year
£5,000	£14,725	£41,437	£35,000	£60,000	£28,200
Peer reviewer: 8 weeks; existing part-time staff	4 part-time 2 full-time staff	2 consultants and existing staff	10 researchers and 60 enumerators	10 full-time 20 part-time staff	1 advisor and existing staff
Oxfam GB	Novib workshop	External advisor	IDS, UK	Novib workshop; external advisor	External advisor from local university

although the organisation is small in comparison to some of the others, the project it represents is relatively large.

Some of the other main distinctions between the case studies include the different stages at which the assessments were undertaken; differences in the degree to which the assessment was perceived as an opportunity to test tools and methods for the future rather than as a means of reviewing the past; differences in the availability of information on a project's history; and differences in the scope and duration of the assessment and therefore of the approaches adopted. Table 1.2 summarises the main characteristics of the various case studies.

Stages in the project-cycle

Figure 1.2 situates the case studies in the project-cycle (which can also be seen as a continuous spiral). The arrows illustrate the approximate position in a particular project's life-cycle at which the assessment was undertaken. The Ghana and the CYSD 1 studies stand out as attempts to establish baseline data and systems of impact monitoring (also called longitudinal data collection) at, or near, the beginning of programme implementation. Most of the other studies examine ongoing projects or programmes, looking both backwards and forwards. Some of these, for example the Wajir programme and BRAC's rural development programme, made use of longitudinal data collected during the course of the project or programme in question. The CORDES and ENDA studies represent retrospective assessments of programmes that had ended, in ENDA's case some four years previously.

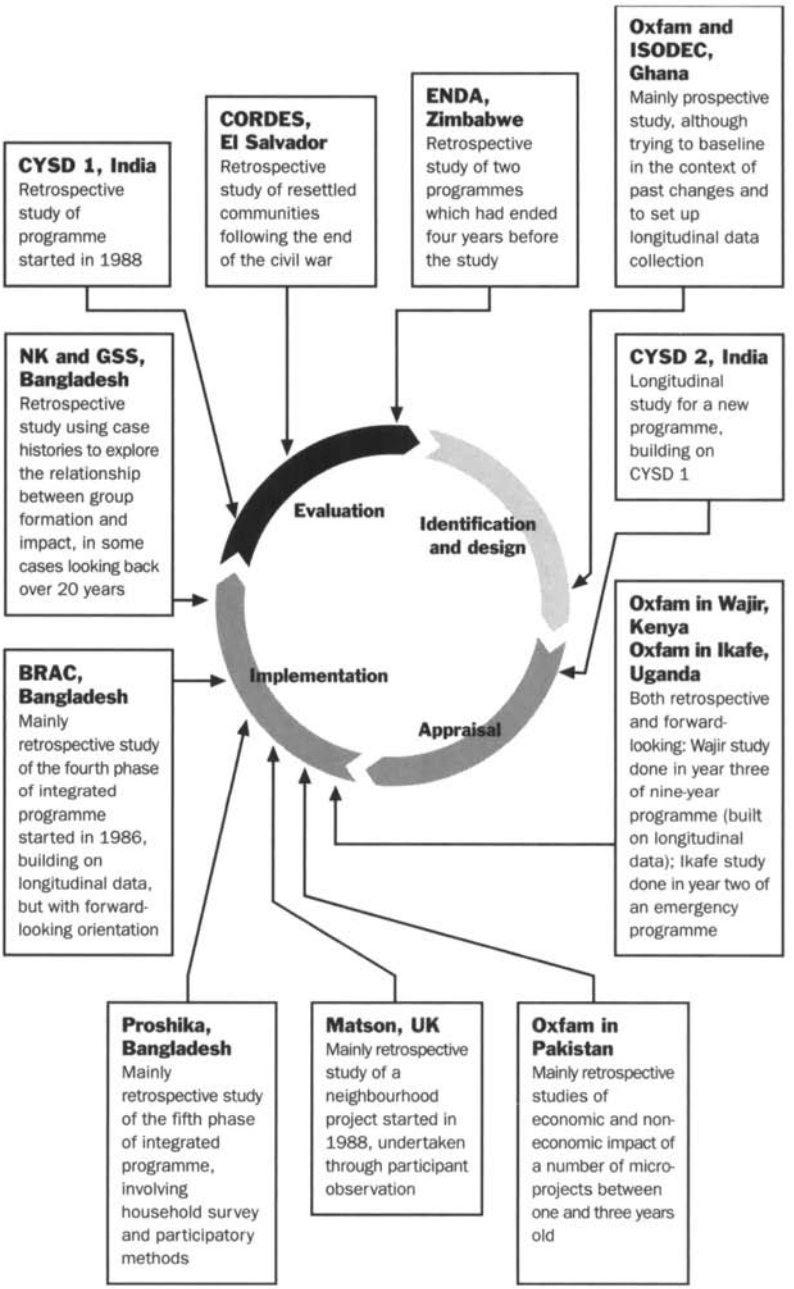
Reliance on existing data

These case studies varied in the degree to which they were based on past processes of monitoring, evaluation, and impact assessment. For example, the Wajir and BRAC studies systematically tried to build on existing impact-data collection systems. BRAC's case study is actually builds on a previous impact assessment carried out in 1993/4. Other studies, although attempting to build on past documentation, have had to expend greater effort on retracing the past using a variety of recall methods, either because existing records were weak or because previous assessments did not capture those elements pertinent to the study in question.

Different approaches

A wide variety of approaches to impact assessment and to implementation timescales, some of which are summarised in Table 1.2 above, is represented in the case studies. They vary from an Indian consultant's eight-week intense involvement in the Matson Project in Gloucester, UK, using mainly participant observation and key informant interviews, to a year-long process with BRAC in Bangladesh, which involved ten researchers and 60 enumerators in a large-scale sample survey of 1,700 households, complemented by case study and focus-group discussions. This rich diversity indicates the range of approaches, organisations, and contexts through which and in which to conduct impact assessment.

Figure 1.2: Where the case studies fit in the project-cycle



2

Our overall approach to impact assessment

This chapter sets out the theoretical foundations which the case studies were based upon. These were informed both by a review of the relevant literature[1] and by a number of key questions which both Oxfam and Novib were keen to explore.

I outline our understanding of the nature and forms of change which underlies the case studies, and our principles regarding participation and the importance of paying attention to gender and other social relations. We are also convinced that impact assessment should be useful and relevant to local staff and partner organisations, and that it should be integrated into every stage of the project or programme. In terms of methodology, I give some thought to the central issues of attribution and aggregation, and conclude with some views on the ethics of impact assessment.

Historical overview of impact assessment

Initial approaches to impact assessment date from the 1950s; development agencies began to use these approaches — which were about predicting, before the start of a project, its likely environmental, social, and economic consequences — in order to approve, adjust, or reject it. Environmental impact assessment (EIA), social impact assessment (SIA), cost-benefit analysis (CBA) and social cost-benefit analysis (SCBA) are some of the most common approaches (see Howes 1992). In recent years, there have been several efforts to integrate social and environmental impact assessments into more coherent forms (see Barrow 1997); impact analysis, on the other hand, was essentially confined to an assessment of impact several years after a project had ended.

The next generation of planning in international development agencies saw the introduction of logical framework analysis (LFA), which attempts to set out a clear hierarchy of inputs, activities, and objectives and to relate these to assumptions made about the external environment. Today the logical

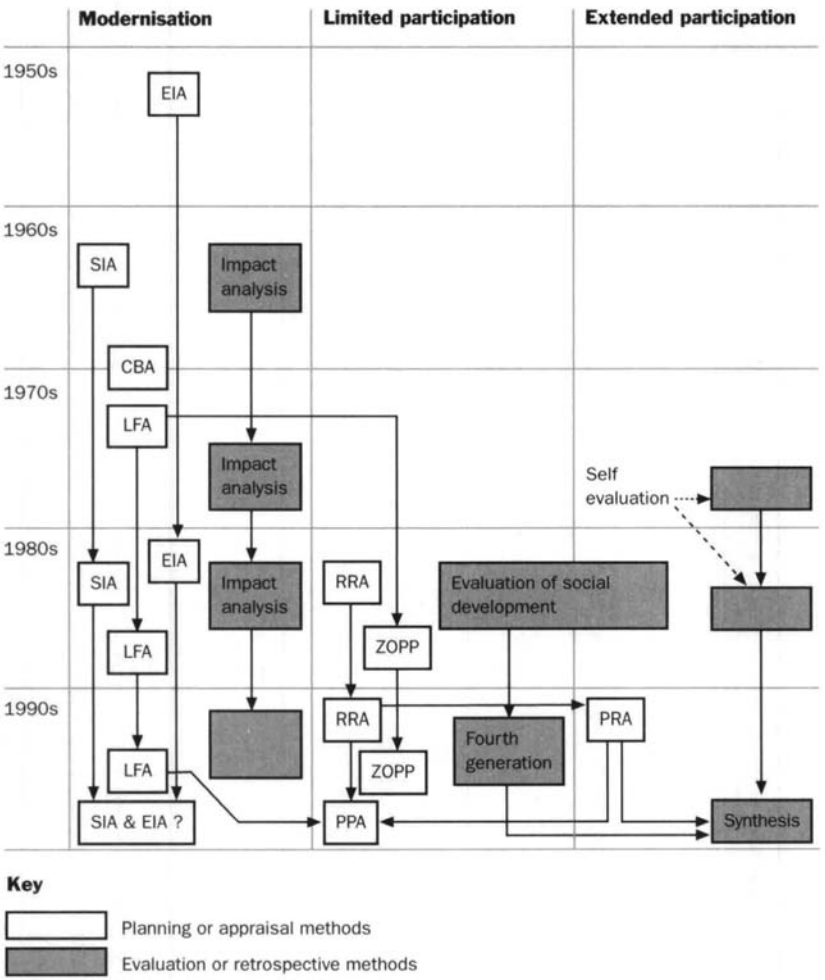
framework and its variants are the most common planning framework used by bilateral and multilateral agencies. LFA was used in several of the case studies: for instance, Chapter 4 looks at how the Wajir programme in Kenya explored cause-and-effect relationships between what the project did and what it sought to achieve.

From the early 1980s, new methods of enquiry emerged which sought to make people and communities subjects and active participants, rather than objects of impact assessment. New methods such as rapid rural appraisal (RRA), participatory action research (PAR), participatory rural appraisal (PRA) — now often termed participatory learning and action (PLA) — and other methods all blossomed during this period (Chambers 1997). This in turn led to a number of efforts to achieve a greater synthesis of these strands. The development of objectives-oriented project planning (ZOPP) in Germany attempted to introduce notions of participation into the logical framework approach (GTZ 1988a). In contrast, approaches to the evaluation of social development (Marsden and Oakley 1991) and 'fourth-generation' evaluation ideas (Guba and Lincoln 1989) have built on historical and anthropological ideas and see evaluation more as a negotiation of diverse opinions and perspectives. This latter approach, in combination with participatory methods, thus seeks to understand the opinions of various interest groups, especially those whose views are not normally heard. In recent years, national-level planning and development strategies have also begun to include participatory poverty assessments (PPAs), seeking to incorporate local perspectives and opinions by including participatory research methods within national frameworks (Norton et al. 1995).

Figure 2.1 (originally developed by Mick Howes) indicates the evolution of these different approaches. He also illustrates the origins of these approaches in relation to various development paradigms such as modernisation, limited participation, and extended participation. Modernisation here refers to an approach largely premised on promoting economic and infrastructural development as a means for 'third-world' nations to catch up with the 'first' world. By contrast, the extended participation approach begins with the belief that poverty is primarily caused by injustice and inequality and that overcoming poverty is impossible without people's full participation. This paradigm demands that outsiders relinquish control and act as catalysts for locally owned processes of empowerment and development. The limited participation approach, in Howes's view, represents a compromise between these two poles and was most apparent in the multilateral agencies' shift to embrace participatory approaches, while retaining a strong planning tradition and an emphasis on economic development.

Figure 2.1: Approaches to appraisal and evaluation 1950s-90s

(adapted from Howes 1992)



What do we mean by impact assessment?

At the outset of the research, Oxfam GB and Novib used the working definition of impact as *'sustained changes in people's lives brought about by a particular intervention'*. Impact thus referred not to any immediate outputs or effects of a project or programme but to any lasting and sustained changes brought about.

We therefore defined impact assessment as an evaluation of how, and to what extent, change had occurred. This required an understanding of the perspectives of various stakeholders in a development intervention, as well as of the social, economic, and political context in which it takes place.

Following the first stage of our research it became clear that — particularly in areas experiencing rapid and unpredictable change such as conflict zones or emergency situations — the emphasis on sustained or lasting change was a problem. It was obvious that in such situations the provision of clean water, for example, could literally save someone's life, which could only be described as a significant impact, if not a lasting one. The definition of impact therefore changed to *'significant or lasting changes in people's lives, brought about by a given action or series of action'*. In other words, programmes can make an important difference to people's lives even if that change is not sustained over time.

The consultant who was recruited to review the existing literature and to undertake some initial discussions with Oxfam and Novib partners also proposed that, given the complexity of the task, there should be two different levels of impact assessment. First,he recommended a focused appraisal with regards to the original objectives of the project, and second, a wider assessment of overall changes — positive or negative, intended or not — caused by a project. All the case studies, while recognising the importance of assessing performance against objectives, opted for the latter, wider definition.

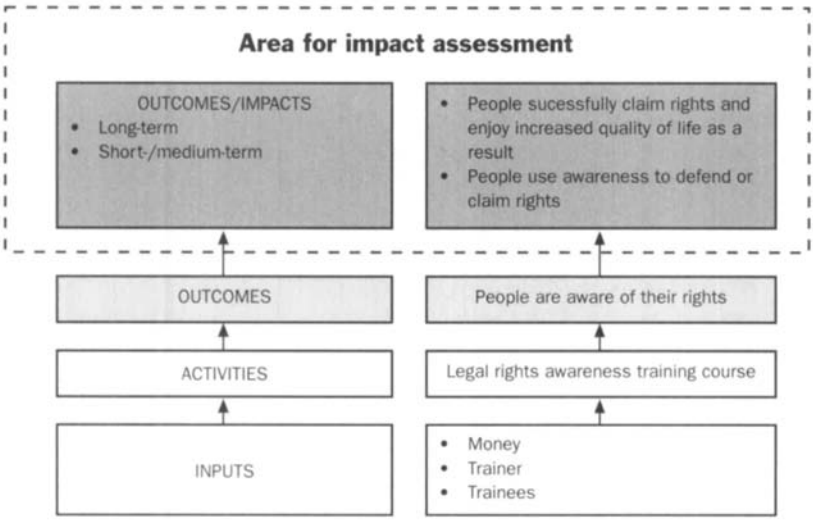
A common definition therefore emerges from the studies: *Impact assessment is the systematic analysis of the lasting or significant changes — positive or negative, intended or not — in people's lives brought about by a given action or series of actions.*

The consultant who carried out the initial research had recognised that

[i]n a certain sense, most of us have been doing impact assessment for a long time. What is new in the discussion is the emphasis on the outcomes and consequences of a project; and the attention given to the systematic *nature of such an effort.* (Hopkins 1995, p.5)

Figure 2.2 illustrates this focus on outcomes and impact, using the example of a legal rights programme. In this case, impact assessment would need to measure the degree to which an improved awareness of rights has led to people actually using the legal system and whether this has improved the participants' quality of life. It also reminds us that although the emphasis may be on significant change and not the inputs, activities, or outputs, all these will need to be examined in order to see if there is a logical link between them and the impact achieved.

Figure 2.2: The orthodox view of the focus of impact assessment

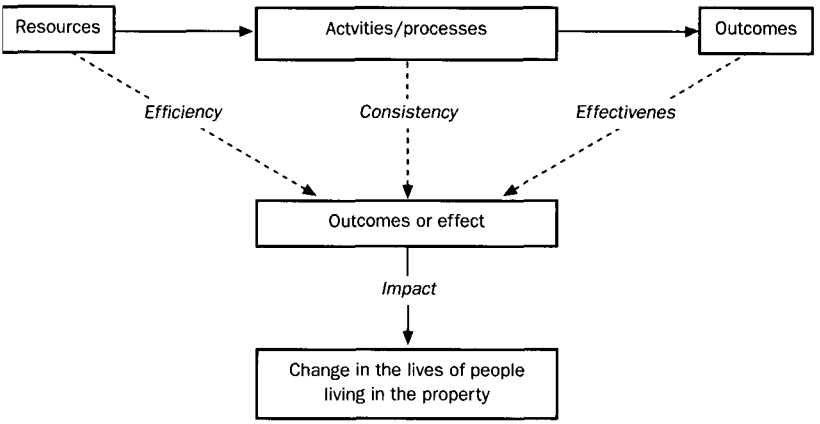


In much of the literature, and in some of the case studies which follow, a further distinction is made between outcomes or effects and impact. In the above example, the actual use of the legal system to claim a right would be considered an outcome or effect, whereas the actual change in quality of life that this brings about would be considered as impact.

These definitions[2] are often used in distinguishing between different kinds of assessment which examine the efficiency, effectiveness, consistency, and impact of an intervention (see Figure 2.3). The first of these analyses the relationship between the resources put into a given project or programme and the outputs and outcomes achieved. Thus, an efficiency assessment helps to decide whether the same results could have been achieved at less cost, or whether significantly better results could have been achieved with only a small amount of additional resources. An effectiveness assessment looks at the degree to which a project has achieved what it set out to do. Third, one can evaluate the degree to which the process or methods adopted were consistent or in harmony with the outcomes achieved: for example, a non-participatory project design and implementation would not be consistent with intended outcomes that sought to strengthen people's capacities to solve their own problems. Impact is then assessed by analysing the degree to which an intervention's outcomes led to change in the lives of those who it is intended to benefit.

Figure 2.3: Distinguishing between efficiency, effectiveness, consistency, and impact

(adapted from Paul Willot, 1985)

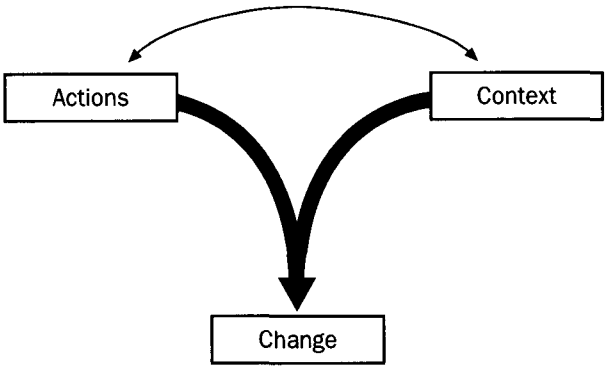


Judgements made in assessment

In practice, the distinction between outcomes or effects and impact, although useful, can be very blurred. Taking up the example of a legal rights programme, this distinction between outcomes and impact can ignore the fact that a person who is now sufficiently aware and confident to use the legal system may consider this a significant change in his or her life, even if it does not immediately lead to a positive legal result or a demonstrable change in their quality of life. This reveals that although impact assessment is about systematic analysis, it is also centrally about judgements of what change is considered 'significant' for whom, and by whom; views which will often differ according to class, gender, age, and other factors.

These judgements are also dependent on the context within which they are made. In some of our case studies, income or assets are considered key indicators of change, and these would have to *increase* over time to be assessed as an impact. In other cases, the *preservation* of current levels of food security over a number of years for the poorest households is considered a significant impact. Clearly, these judgements can only be made for each specific situation, which leads to the important conclusion that change is brought about by a combination of the activities of a given project or programme and the ongoing dynamics of the context in which these activities occur (see Figure 2.4).

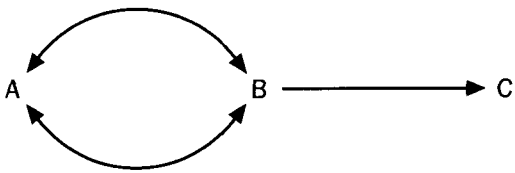
Figure 2.4: Actions and context combine to produce change



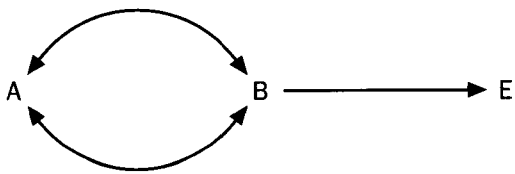
The nature of change

Impact assessment is therefore essentially about the measurement and/or valuation of change. For many years, development planning models have been associated with a linear notion of change, which assumes that an input A leads to an output B and an outcome or impact C.

However, in recent years there has been a growing interest in non-linear models of change, which recognise not only that may A and B may influence each other,



but also that this interaction may result in the same input producing divergent, and possibly unexpected, outputs over time or in different places.



This means that the change which occurs is contingent on the specific events, people, and conditions present in a given situation as well as on the action or activity undertaken. In other words, the resulting change may be due to a myriad of factors combining in a particular way. This means that using the same inputs will not necessarily produce the same results in the future. Given the nature of development projects, and of the organisations that run and support them — which are not machines but involve people who have their own ideas, dreams, and interests — it is particularly important to recognise the contingent nature of the change they produce.

The second characteristic of non-linear change is that it may be sudden, discontinuous, and unpredictable rather than drawn-out, stable, and forecastable (Uphoff 1993, Roche 1994, Fowler 1995). When a coup-d'état occurs, when currency speculation takes hold, or when an earthquake happens, a whole chain of unexpected events can disrupt the status quo. This might provoke considerable changes in social, economic, or political systems which hitherto would have seemed unlikely or impossible. Given the interdependence of environments, and the growing links between economies, communication systems, and in some cases political systems, the possibilities for systemic shocks, whose ripple effects reverberate around the world, are increasing. Such sudden changes can also occur at a local level, for example when the charismatic leader of a local organisation suddenly resigns, or when a new feeder road is built.

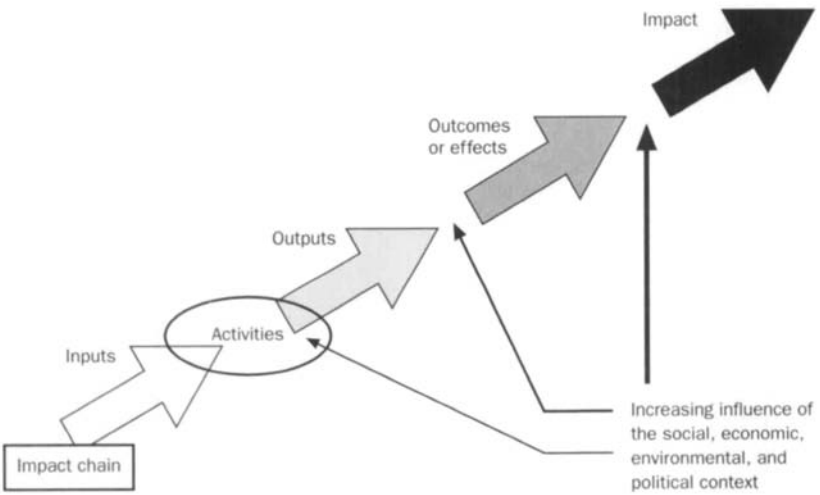
These issues are important for the purposes of impact assessment as they remind us that development and change are never solely the product of a managed process undertaken by development agencies and NGOs. Rather, they are the result of wider processes that are the product of many social, economic, political, historical, and environmental factors — including power struggles between different interest groups. Understanding these processes is important if the changes brought about by a given project or programme is to be properly situated in its broader context.

The impact chain

The degree to which the context of a project influences change will increase the further up the 'impact chain' we move (see Figure 2.5). Going back to our legal rights example, the likelihood of the training course being held is largely dependent on the organisers doing their job and the funds being available. However, whether this course results in trainees' increased awareness of legal rights depends not only on the trainers' skills and abilities, but also on whether the trainees pay attention, are able or willing to read any course texts, or whether women trainees have their husbands' permission to attend.

At the level of outcomes and impact, people's ability to use the legal system and benefit from it in turn depends on how the legal system functions in that context — whether particular groups are discriminated against, whether support and funding for plaintiffs is available, and so on.

Figure 2.5: The impact chain and the importance of context



However, although the impact chain provides us with a useful way of distinguishing between different levels of change, these differences may not be evident in practice, may be meaningless to people in communities — who assess change in their lives without distinguishing between 'outcomes' and 'impacts' — and need to be clearly situated within, and analysed in relation to, the context.

Although I may refer to examples which relate to different parts of the impact chain in the following chapters, most impact assessments will need to explore the entire chain if reliable conclusions are to be drawn about the degree to which any observed change in people's lives can be attributed, at least in part, to a given project or programme.

Impact assessment in relation to monitoring and evaluation

Some of the case studies suggest the following distinctions between impact assessment, monitoring and evaluation.

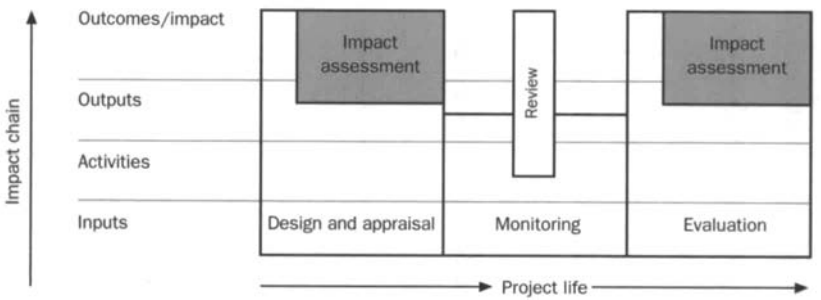
Timing: Monitoring occurs frequently and evaluation periodically. Impact assessment, however, occurs infrequently, usually towards or after the end of an intervention.

Analytical level: Monitoring is mainly descriptive, recording inputs, outputs, and activities. Evaluation is more analytical and examines processes, while impact assessment is mainly analytical and concerned with longer-term outcomes.

Specificity: Monitoring is very specific and compares a particular plan and its results. Evaluation does the same but also looks at processes, whereas impact assessment is less specific and in addition considers external influences and events.

Figure 2.6 represents what one might call the 'classical' view of the distinctions between monitoring, evaluation, and impact assessment. From this perspective impact assessment has been seen either as part of the appraisal stage (in order to anticipate potential environmental, social, or economic consequences and to redesign a project if necessary) or as a specific type of evaluation which occurs near or after the end of a project (in order to review past impact and to inform future plans or revised policies). In addition, in this view impact assessment is understood to examine only the outcome or impact level of an intervention (illustrated in Figures 2.2, 2.3 and 2.5). Recently, the importance of mid-term reviews or evaluations has been increasingly recognised; these are intimately linked to monitoring, and are increasingly addressing issues of impact. However, in the classical model monitoring, i.e. the ongoing collection and analysis of information during the course of the project, is generally seen as referring to input, activity, and output data, not to impact.

Figure 2.6: The 'classical' view of the difference between monitoring, evaluation, and impact assessment



Although the clasical model can be useful in distinguishing between different elements of recording and reviewing progress, there can be major limitations in adopting this framework too rigidly. Indeed some of the studies conclude that 'impact monitoring' is an essential part of impact assessment and that too rigid a distinction between monitoring, evaluation and impact is unhelpful. What is clear is that monitoring, evaluation, and impact assessment are all intimately connected and draw on each other.

In view of the complexities of definition, we have been guided in this research by the following considerations. Although various stakeholders will have their own perspectives in judging the impact of a given project, far greater attention must be paid to the views of those who are intended to benefit. Second, it is important to assess all changes (whether these are positive or negative), including unintended and unexpected ones, rather than simply looking at the intended impact of an intervention. These changes may also affect people other than those who the project intends to benefit. Moreover, among the groups intended to benefit, as well as among those outside an intervention's reach, people will have different experiences and perceptions of what change has been achieved. Views will diverge particularly between men and women, but also between other social groups.

While the intended impact of a development intervention is often defined in terms of long-term, sustainable changes that it wishes to bring about, there may be short-term results which are attributable to the intervention and which people judge to be significant too. For example, a dam-building project which displaces people from their land has a significant impact on them from day one, whether this is intended or not, and whether they eventually regain some land! One of the most important questions to answer is whether the project or programme in question has brought about results that would not have occurred otherwise. This involves not only an assessment of the input-output-impact 'chain', but also of the context in which the project evolved and how this context interacted with the project activities to produce change.

Power and participation

At the outset of the research Oxfam and Novib stated that they saw the development of participatory monitoring and review processes as the basic building block for impact assessment. This was based on the view that significant and lasting change in people's lives must take account of *their* values, priorities, and judgements; projects cannot be deemed to have been a 'success' or 'failure' if the perceptions of those who the intervention aims to benefit diverge seriously from those of the project staff or an external evaluator.

In addition, the literature review concluded that monitoring and evaluation systems have tended to be 'top-down and bureaucratic' and that a 'frequent way to impose authority has been to introduce sophisticated jargon' which now needs demystifying (Hopkins 1995a). Valuing the wisdom and judgement of ordinary people is therefore a critical element of any impact assessment process. 'The main challenge facing organisations such as Oxfam and Novib is how to incorporate these opinions, especially those of women, into our monitoring and evaluation systems.' (Dawson 1995)

However, we also recognise that there are other stakeholders who require information and that it is often extremely difficult to satisfy everybody's needs. Thus while insights and lessons can be drawn from participatory processes of assessment, a large organisation can only use these in a selective way. It is therefore suggested that organisations should not seek 'to aggregate and summarise the richness of this type of information, or to seek standard indicators across diverse projects and programmes' (Hopkins 1995a). Rather than simply adding up results, project-level findings can be incorporated into wider processes of review and learning. This approach explicitly acknowledges a number of interest groups, who have different and possibly conflicting objectives, are involved in any process of intervention. As these interest groups will judge impact in different ways, the search for an 'objective truth' or reality that all can agree on is, at best, fraught with difficulty and, at worst, impossible. In this scenario, impact assessment becomes far more of a negotiation of different interpretations of what has happened or will happen.

It was not made explicit in the literature review that these various interest groups bring different status, resources, and ultimately power to the negotiation process, and that some group's views on what has occurred can therefore become dominant. The question, as Robert Chambers put it, becomes 'whose reality counts?' or, as Estrella and Gaventa ask in relation to impact assessment, 'who counts reality?' (Estrella and Gaventa 1997). The meaning of the now widely used term 'participation' therefore needs greater clarity. Among the questions that need to be addressed are:

- What do we mean by participation?
- Who participates and on what terms?
- How can we trust processes which claim to be participatory?

I will return to these questions in the final chapter.

The importance of gender and other social relations

Oxfam and Novib both believe that social relations are a critical determinant of well-being or poverty, a view shared by most of the case-study participants.

Confronting gender-related inequalities is seen not only as a pre-requisite to 'achieving sustainable development and alleviating poverty' (Oxfam Gender Policy, 1993) but as a objective for social justice in its own right. It is well known that within communities, factors such as gender, class, ethnicity, religion, ability/disability, and age are important elements and that communities do not share a single identity, goal, or ambition.

We also know that in recent years, the understanding of the household[3] has changed enormously from an approach which stressed sharing and co-operation to models which include the possibility of negotiation, bargaining, and conflict. The old approach viewed the household as a single unit, rather than one made up of individuals who are connected to wider structures and networks, such as kin or age groups, mutual support groups, or more formal organisations. Moreover, the household was described as if it were the same everywhere, rather than recognising the large differences between households within and between societies as well as over time.

Given these insights about power and participation, those planning an impact assessment must reflect carefully not only on what needs to be assessed and how this is done, but who is involved, and what level and unit of analysis is most appropriate.

In the past few years increasing attention has been paid to gender issues in the design, implementation, and evaluation of development projects. Several frameworks have been developed in order to assist better gender analysis, including Practical and Strategic Needs, the Harvard Framework, the Capacities and Vulnerabilities Framework, and Social Relations Framework (see March et al. 1999 for detailed discussion of the advantages and disadvantages of these).

The case studies made different assumptions about gender and other factors that shape social relations; they therefore explored the issues in different ways and emphasised different aspects. I will explore these and present some of the tools and methods used.

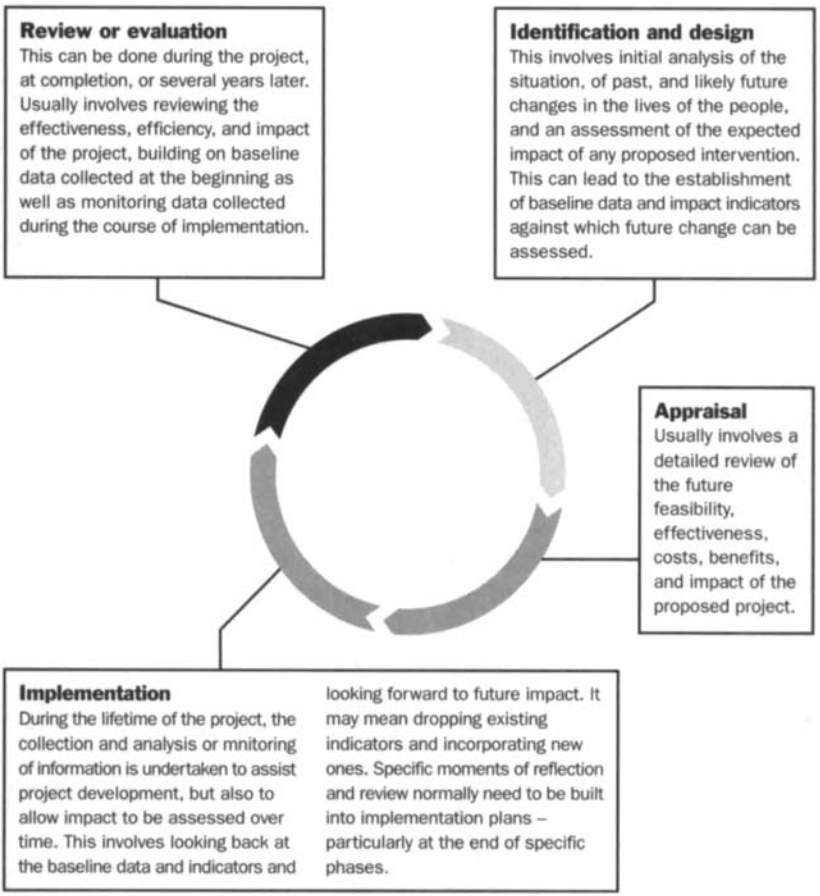
Impact assessment and the project- or programme-cycle

The literature review stressed that 'it is not possible to carry out impact assessment activities successfully if the more basic task of evaluating activities and their immediate effects is not done properly' (Hopkins 1995a). A better assessment can be made if a baseline study has been carried out, indicators have been monitored efficiently, and clear objectives have been defined, with corresponding activities and indicators. This means that impact assessment is an activity that must be done *throughout* the project cycle; what

changes is the nature of the exercise. In the preparatory stages, it is anticipatory or prospective. In the process of implementation, it checks what consequences a project is currently having. After completion, the emphasis is on examining what effects the project has had.

Moreover, because most interventions aim to improve the quality of life of a target group of people, their input needs to be captured at all stages of the project-cycle. Figure 2.7 details the typical form of various elements of the impact assessment process at specific stages of the project-cycle.

Figure 2.7: Key elements of impact assessment at specific stages of the project-cycle



The image of the project-cycle has a number of well-known weaknesses: projects do not endlessly repeat the same 'cycle' as they evolve; rather, the process can be described as a 'spiral' with continual adaptation and modification. Another weakness is that a project approach tends to be rigid, inflexible, and to ignore the context. Third, while the project tends to dominate the lives of those who manage and support it, it may make up a very small part of the lives of those it seeks to benefit.

With this in mind, I will explore what the case studies tell us about the utility of a project-cycle conceptualisation of impact assessment, and what the findings suggest in relation to more programme- or process-oriented approaches.

Impact assessment and organisations: The problems of attribution and aggregation

All organisations, be they a community-based group, a local NGO, or an international agency, need to make sense of what they are doing. They also generally want to know what difference they are making. This raises two central problems: how organisations can synthesise or summarise what they are doing (aggregation), and how they can find out to what extent any changes they observe were brought about by their actions (attribution). These issues are further complicated if the organisation has to communicate their achievements to many other people, both inside and outside the organisation.

Attribution

'Success has many parents ... failure is an orphan.'

In the current climate, despite the vogue for development organisations to proclaim partnership and the need for 'synergy', raising money and gaining in reputation largely depend on demonstrating what a single organisation has achieved, what difference it has made, and what value it has added, usually to the exclusion of other actors or agencies. As Neil Thin points out, NGOs, like other organisations, tend to blame others and/or the external environment when things go wrong (Thin 1992). Similarly, they tend to neglect to mention the role played by other organisations or concomitant events when things go right.

Yet one of the most problematic parts of impact assessment is determining causality, because in real life, a combination of several factors is likely to have caused any observed change. In the laboratory, different possible sources of change are isolated so as to observe, and precisely measure, their different

effects. Outside the laboratory, and particularly in the area of social development or emergency work, it is not usually possible, or appropriate, to rely solely on this form of scientific observation, for the following methodological and ethical reasons:

- It is usually impossible to find a control group similar to the group of people benefiting from an intervention, who are subject to exactly the same influences, except for the specific agency input, and whose situation mirrors that of the beneficiary group over the life of a given project.
- Withholding support from a control group in order for it to remain 'uncontaminated' is not only difficult (you might have to persuade other agencies not to support them!), it is in most cases unethical and may contravene their human rights.
- A particular intervention does not occur in isolation from those of other organisations, from the local context, wider economic and political policies and, more importantly, from the actions and reactions of those who it aims to support. It will interact with these in different ways in different places. Unpicking what caused an impact in this situation in any scientific way is likely to be very difficult and usually very expensive (Hulme 1997).

Often the most that can be done is to demonstrate through reasoned argument that a given input leads logically towards a given change, even if this cannot be proved statistically. This in turn can be cross-checked against the views of other actors to see if important areas of agreement or disagreement emerge. In some cases, quasi-experiments with control groups or non-beneficiaries used in combination with other methods may be appropriate.

But it should always be remembered that a common principle of development is that agencies such as Oxfam or Novib should be trying to 'work themselves out of a job'. In doing so we usually aim to enable people to take ownership of achievements: this directly affects how one might look at attribution. If a group says 'we did it ourselves', this is an indicator of their empowerment. While they may recognise the role that others have played in supporting them, if others start claiming the success as their own, then this might undermine the group's sense of self-worth and self-confidence[4].

In the case studies, various approaches are used to deal with the problem of attribution such as the use of control groups, statistical analysis, and cross-checking. I explore the advantage and disadvantages of these different approaches and their appropriateness for different context and types of organisation in Chapters 3 and 4.

Aggregation and synthesis

All organisations need to organise, categorise, synthesise, and aggregate information. The larger the organisation, the bigger the problem. The problem can be posed in terms of practical issues — how does an organisation cope with the large amounts of information generated from participatory processes? It can also be defined in terms of utility or rigour: does the aggregated data actually mean anything? And is it of any use?

Organisations cope with this by demanding that certain categories of information (such as budgets) are included in, for example, project proposals, reports, and evaluations. In addition to formal reporting structures, staff also communicate to others in their organisation in ways which 'filter' information: how they share information depends on what use they assume will be made of it. For instance, when an Oxfam staff member from Africa was asked how he knew that a tomato-growing project was successful, he replied 'because the men in the village are marrying second wives'. He was then asked why he did not report this fact in his annual report, and commented: 'Because that's not what you want — you want "10 tonnes to the hectare" or "income up 50 per cent"'. It is beside the point that this quantitative data had not been demanded — he *assumed* that it was required. This underlines the importance of recognising the informal signals that an organisation may send out to its staff or partners, in addition to its formal procedures and guidelines.

The case studies aggregated and synthesised findings in different ways. I shall explore these as well as the strategies which some of the organisations involved have adopted in order to meet the needs of various stakeholders.

Avoiding bias

The case studies take different starting points regarding the assumed objectivity or subjectivity of particular approaches to impact assessment. This is not surprising, given the intensity of the debate that sometimes surrounds these notions.

The BRAC team decided not to rely solely on participatory methods.

[A]lthough the criteria used in the participatory approach tell us more about local people's multidimensional conception of poverty than conventional consumption and expenditure measures, these are, by their very nature, subjective and only definable within the norms and customs of a given society. (BRAC case study)

In contrast, the Matson study deliberately started with the 'subjective' opinions and perceptions of local people, precisely because the researcher believed that understanding their views counted most.

No one can be entirely objective, of course, and the measures that researchers develop, even quantitive ones, always contain a degree of subjectivity. For example, who decides that $1 a day is an accurate poverty line? In addition, the choice of approach is often the result of rather pragmatic concerns: BRAC was attempting to assess the impact of a programme that covered over 60,000 village organisations, and therefore felt that they needed to be able to compare results across communities, whereas the Matson study covered a single housing estate.

We will return to this issue at several points in the next chapters, because it goes to the heart of several issues related to the design and implementation of impact assessments, the choice of assessment methods and teams, and the importance of cross-checking or triangulation.

The ethics of impact assessment

The case studies highlighted several ethical questions associated with evaluation processes which need to be borne in mind, some of which become particularly acute in emergency or conflict situations.

At the community level, the dangers of raising expectations and of taking up too much of people's time are commonly raised in the studies. This suggests that a critical ethical issue is the degree to which an impact assessment is meaningful and intrinsically worthwhile to those involved. If impact assessment work is purely interested in extracting information and does not allow *all* participants to gain insights, to reflect on their observations, and to look to the future, then the cost to some of those involved will far outweigh the benefits. In this case, it may be appropriate to recompense them in some way. This issue is explored further in relation to different approaches to impact assessment.

A further ethical issue relates to the danger of impact assessment processes exacerbating existing conflicts of interest within communities. For instance, in a programme which is having negative effects for women, or another group within a given community, it may be difficult for them to express their feelings about this, because a more powerful group is benefiting. Anybody assessing impact therefore must make efforts to understand who the winners and the possible losers might be, and seek the opinions of both. Promoting the continuation of activities that negatively affect certain groups clearly has an ethical dimension.

There are some specific issues realted to work in conflict zones or with traumatised populations such as refugees. In these cases, simply talking to evaluators may be dangerous (particularly if the evaluation subsequently leads to unpopular decisions being taken) or disturbing (if it leads to people having

to recall traumatic events). In very tense situations the evaluation process may well lead to the emergence of conflicting views, which precipitates more conflict and generates risks for local populations, agency staff, and others. Agencies or individuals commissioning assessments need to be aware of these dangers and seek to minimise them. They should also recognise that evaluators may be particularly at risk in conflict zones, especially given that they may uncover facts that some would like to remain unknown.

However, these ethical concerns must be balanced against the moral question of accountability to others such as donors or supporters, who have often given money on the assumption that a particular impact or outcome will be achieved. There are further ethical considerations in relation to future performance, in terms of the degree to which an assessment can contribute to future positive impact for other people. Trying to balance these various concerns involves complicated trade-offs, which I shall look at particularly in Chapters 4 and 5.

3

Designing an impact assessment process

This chapter explores how to design an impact assessment process drawing upon the experience of the case studies. It attempts to cover some of the critical issues that must be considered when planning an assessment process, be it before a project starts, during its course, or after it has finished. I discuss what preparatory steps need to be taken, the questions of baseline data and indicators, as well as the difficulty of how to attribute change to a given intervention. The chapter concludes by analysing how the studies approached the need to cross-check findings and synthesise them in such a way that the results could be used and communicated. Some of the major lessons highlighted in this chapter relate to the importance of all stakeholders being clear about the purpose and focus of an impact assessment exercise; and to the necessity of deliberately exploring negative and unexpected impacts, of proactively involving those whose voices may not normally be heard, and of preventing bias by cross-checking findings.

Preparatory steps

When designing an impact assessment, whether it is at the beginning of the project, in the middle, or at the end, time spent in preparation is rarely wasted. Being clear about the purpose of the assessment, who it is for, who should be involved, and what resources are available, are just some of the questions that need to be addressed. The case studies answered them in different ways.

Defining the purpose of the impact assessment

This may seem an obvious point, but it is often overlooked. The purpose of most impact assessments is to demonstrate past impact and to improve future practice, and there may be tensions between the two. These tensions will be more difficult to handle if there is a large number of diverse stakeholders with an interest in the exercise, but who have different expectations of it.

Clarifying expectations and what different interest groups mean by impact can help to create a shared understanding of the process.

In most of the case studies, introductory workshops or meetings were held to discuss understandings about impact assessment and to define the purpose of each study. In some cases this involved Oxfam and Novib, who in most cases not only funded the impact assessment work but also the organisations involved. In the case of the Wajir study, representatives of the British Department for International Development (DFID), the funder of the Oxfam programme, were also present. In only one case (Ikafe) were representatives of intended beneficiaries directly involved at this stage. In many cases, the introductory workshop was held before it was decided where and with whom the impact assessment was to be undertaken. However, in some cases (such as the CORDES study) once communities had expressed interest in getting involved, similar workshop were held with them.

In Zimbabwe, an introductory workshop was held over five days and was attended by two members of staff from NOVIB and ten ENDA staff members.

The workshop provided a forum for debate and understanding of the [impact assessment] project [and] played a very important role in educating the participants on the expectations of the project. It is unlikely that a total understanding of the concepts of the impact assessment study could have been taken up easily without such a workshop. Of particular concern among staff were issues of whether future funding was related to the outcome of the impact assessment. Again the suspicion of the study being an evaluation may have been lingering in some minds. This was settled by having the participants choose which projects they would like to assess. (Bowdin King, ENDA case study report)

The issue of funding relations was touched upon explicitly and implicitly at several of the workshops and indicated the slight suspicion which many felt about this research, despite Oxfam and Novib's insistence that this was a mutual learning process, designed to find out about ways of assessing impact. These workshops did allay some of these fears, but at least in some cases, the question remained, because relationships between agency and partner organisation and decisions about future funding were being negotiated during the same period. It was inevitable that in some cases it was impossible to 'insulate' the impact research from these other processes — we shall return to this issue in Chapter 7.

Another important element of the workshops was to define the boundaries and focus of the case studies. This meant taking into account the financial and human resources available, the time scale envisaged, the scale

and stage of the project being assessed, the types of activities undertaken and how long they had been running and the supposed impact of the project in question. As a result of these considerations, some components of projects were dropped from the assessment. For instance, BRAC deceided to exclude the issue of institution-building from their study, because staff felt that not enough time had elapsed since they had adopted a new strategy for this.

Workshops of this type can be a useful step to determine and agree the nature, scope, and purpose of impact assessment studies. In several of the case studies they also proved important in overcoming suspicions and mistrust between the stakeholders. The workshops also encouraged a sense of ownership of the assessment work within the organisations involved, particularly when more staff than those who would implement the project were involved, as was the case with ENDA, CYSD, and CORDES.

Perhaps most importantly, the workshops gave participants the opportunity to define what models of change they were using and what kinds of impact they envisaged.

Defining models of change: What has changed and why?

All impact assessment processes make assumptions about how change happens and why it has occurred, or that it will occur in the future. Sometimes these assumptions are made explicit but often, as David Hulme states, 'in many smaller scale exercises the framework is implicit and may be seen as 'common sense' (Hulme 1997). In those case studies where these assumptions were made explicit they were found to be based on certain theoretical frameworks relating to poverty or well-being.

In their study, which covers some 68,000 village organisations which are a part of its rural development programme, BRAC took a conventional approach to poverty measurement (looking at consumption and expenditure) because, although they recognise that poverty is multidimensional, they believe that 'its characteristics are sufficiently well correlated with consumption and expenditure to allow us to focus on these two variables'. This approach was then complemented by a range of qualitative techniques which aim to explore people's own perceptions of poverty. In their study of Nijera Kori (NK) and Gonoshahajjo Sangstha (GSS) — local Bangladeshi NGOs whose work emphasises social mobilisation — Rao and Hashemi explicitly rejected the assumption 'that poverty stems from a lack of access to scarce resources and that once a channel to such resources has been established it can continually be harnessed. In contrast, we begin with the assumption that poverty is created and reproduced through structural inequalities'. This led Rao and Hashemi

to put less emphasis on assessing individual income and consumption and more on understanding the extent to which village organisations and their federations 'pit their interests first against the local power structure and then against the state'.

The Wajir study team, building on the programme documents, concluded that the reduction of poverty and vulnerability in a remote pastoral zone in north-eastern Kenya is dependent on strengthening and diversifying pastoral livelihood strategies. They agreed that this could best be achieved through improved access to market and state services (particularly concerning human and animal health, and provision of water, credit, and education servies); through a legal and policy framework which endorses common grazing and enforces existing property rights; and through stronger pastoral associations which can take collective action in order to claim and defend their rights. As a result, the Wajir study looks at change in pastoral livelihoods, at possible differences in pastoral livelihoods (and therefore in welfare) between project and non-project sites, and at the performance of service-delivery by local state and non-state institutions.

The Pakistan study on the other hand, which explored the impact of a number of micro-projects, started from an assumption that human development or well-being is multi-dimensional. The approach they adopted was explicit about what these dimensions are, in order not to miss important elements, and in order to ensure that the study focused on the ends of development, not just the means. The study uses a number of dimensions of human development from the field of ethics which are considered universal, irreducible, and non-hierarchical. These are life (which encompasses health, nutrition, security and so on), knowledge, excellence in work and play, relationships, beauty/ environment, inner voice/ peace, religion, and empowerment. The Pakistan study thus focused on the degree to which the projects in question produced change in these dimensions, and which ones the beneficiaries considered the most important. This was done without any prior judgement about the relationship between the dimensions, nor about the relationship between the project and its stated objectives.

This range of assumptions led to important differences in emphasis between the studies, particularly in terms of the main units of assessment chosen (individual, household, community, organisations, institutions), the areas or dimensions of change which were investigated, and the tools and methods chosen. Given how important these assumptions are in shaping the study, it makes sense to spend time clarifying and agreeing them. Summary 3.1 outlines some of the lessons learned in the course of this process, as well as other important issues that need to be clear at this stage.

Summary 3.1: Key lessons about preparing for an impact assessment process

- It is important that all stakeholders are clear about the purpose and focus of the assessment as well as the resources available — initial workshops or meetings can help in creating this shared understanding.
- Difficult issues that may need to be raised at this stage include how the results may or may not influence future funding.
- It may be necessary to clarify what is meant by impact assessment as opposed to any other evaluation or review exercise.
- At this stage, one should make explicit different stakeholders' assumptions about how change happens, what area of work they consider important to explore, and how they view the context within which an impact assessment exercise will be undertaken.

What areas of change and indicators are to be assessed?

Once the purpose of the assessments was clear and the assumptions about change made explicit, most of the case studies then defined the areas or kinds of impact to be assessed. The choice of indicators is often seen as one of the most crucial steps in the process of impact assessment, but there is no agreed method of doing so. Three different approaches were used in the case studies.

The first approach, used in restrospective studies, is largely project-led and involves an initial review of the current state of both the project in question and its external environment, in order to determine which indicators are most important to explore. In some case, new indicators were identified in this review, or previously established ones modified. This sometimes involved a process of consultation with communities and focus groups in order to determine the indicators.

For BRAC, whose team had already undertaken an impact assessment study (IAS 1) in 1993/4, this meant reviewing the indicators they had used to see if they remained relevant for IAS 2. Given that they wanted to compare the two studies, this posed certain problems.

Comparing with IAS 1 required the retaining of as many indicators of IAS 1 as possible. This raised a dilemma. We finally decided to exclude redundant

or poorly specified variables included in IAS 1 ... [o]nly those indicators were retained which were of substantial value and relatively easy to repeat. (BRAC case study)

However, in other cases the projects being studied had either not previously established any indicators, or those that they had established proved difficult to assess or were simply not relevant. In cases where indicators were not available, for example in the CYSD retrospective study of an integrated project supporting tribal people in Orissa, India, researchers decided to engage with local communities to establish a set of indicators — related to the sectors which the project had worked on such as agriculture, literacy and so on — which could then be used to assess the project.

The project personnel have not given much effort in identifying/ visualising long term tangible/ intangible indicators and standards in participation with target population during different stages of the project. In this context there were no proper yardsticks (indicators) for the researcher to assess the level of achievement. Hence a set of indicators were framed by the research team from the villagers perspective through focus-group interview (with representatives from all social strata). Here they were asked to express their indicators (according to sector) by which they expect the project can withdraw from the area. These intangible indicators were then considered as best criteria for assessing the level of achievement. (CYSD 1 case study)

The second approach is more open and first and foremost explores changes brought about by the project as stated by various stakeholders, using generic checklists which are broader than the original scope of the project. The checklist in the Pakistan study, for example, included health and security, knowledge, excellence in work and play, relationships with others including gender relations, inner voice and empowerment, and beauty and religion. Men and women in the communities involved were then asked if change had occurred in these areas and to specify what the change was. As such these 'indicators' represent people's own examples of specific changes that has happened in their lives, rather than functioning as general, verifiable indicators.

The third approach is similar to the second, but attempts to 'turn the telescope round'. Rather than looking at impact solely from the perspective of an organisation or the project it supports, it starts by looking at what change is considered most significant in people's lives, irrespective of any project. It then goes on to explore with those concerned what are the processes bringing about such change, among which NGO-initiated projects and programmes may be one, and possibly minor, element. This approach was used in the Matson and Ghana studies.

In the first approach, indicators are established and the exercise of assessment is then essentially about verifying the degree to which these indicators can be confirmed. In the latter two, however, indicators exemplify broader areas of change in people's lives.

The importance of looking beyond predetermined indicators was deemed important in some studies due to what some have called the indicator dilemma (Goyder et al. 1998). This dilemma, as illustrated in Figure 3.1, notes that indicators which are used to verify impact can, by definition, only capture expected change and will only reflect those areas of change which can be made explicit or are agreed upon by key stakeholders. This leaves out situations where unexpected change occurs, or areas of change that have not been agreed or are left hidden by one or more stakeholders.

Figure 3.1: The indicator dilemma: What kind of change is captured?

Source: Action Aid Participatory Impact Study, Goyder et al. (1998)

	Expected	Unexpected
Agreed	✓✓✓	???
Not agreed	???	???

For instance, a review of a goat-rearing project (carried out as part of the Pakistan case study) revealed that the project had an important impact on women's' religious life because it enabled them to sacrifice animals for the Muslim festival of Eid. This had not been made an explicit area of change at the beginning of the project, and although the women may always have viewed it as an indicator of success, it was not seen as such by Oxfam. If Oxfam had only assessed change in line with the project indicators, it might well have not discovered this. Clearly this sort of surprise can be prevented by better pre-project discussions between communities and those supporting their projects. However, it may often be the case that these sorts of impact are unexpected, and there will be areas which some groups may feel uncomfortable or unwilling to declare as future indicators of success, for a variety of reasons. An evaluation of an Oxfam-supported credit and income-generation programme in Chad some years ago revealed that one of the women's own success criteria was that as a result of the project, they were now able to purchase bigger cooking pots. This in turn had allowed them to

participate more fully in community celebrations, a critical element in feeling part of a social network. As Bridget Walker, formerly of Oxfam, observes, 'I wonder whether Oxfam would have initially supported a project to purchase bigger cooking pots for weddings and circumcision ceremonies!'

The indicator dilemma also illustrates the problem of assessing negative impacts. Although indicators may provide information about poor levels of attainment of agreed and expected changes, these are nearly always phrased in a positive way: increasing income, reducing animal mortality, and so on. Unexpected negative impacts will not necessarily be exposed. As similar studies have found (Goyder et al. 1998), the unequal relationship between an NGO and those who it seeks to support, may make the expression of negative impacts less likely. The method used in the Pakistan case study underlines this.

There will be positive impact and negative impacts. There are negative impacts even of very good activities. It is nothing to be embarrassed of. For example, if a women gets married, she spends less time with her parents; if I have a good job, I have less time to drink tea with friends. (Pakistan case study)

In the Pakistan study, the methodology developed put great emphasis on raising the issue of negative impacts at the beginning of any meeting with communities, and repeating questions about negative impacts at the end. However, this study also stresses to the researchers (local Oxfam staff) that their attitude and willingness to hear about negative impacts is also critical.

You yourself must really *want to know what has gone wrong, or is going wrong — if you don't, how can you diffuse it? And you must welcome criticism.* (ibid.)

The importance of developing trust and openness is brought out in the CORDES study. It seems that due to the very close relationship between CORDES and the community, the negative effects of a failed collective credit programme were clearly expressed. This in turn allowed important lessons to be learned not only about the technical aspects of credit management, but also about the knock-on effects in terms of the community's increased tendency to act as individuals.

Other research teams deliberately sought out those who might have been negatively affected by a project or who might have more critical opinions. For example, BRAC included individual case studies which were selected from those households that had dropped out of the programme. The Ghana study purposely chose women and men who were not benefiting from NGO projects in its sample of interviewees in order to assess possible negative effects upon them. Several studies specifically explored changing gender relations, and the effects on women, given the well-known history of negative impact that development projects have had on women's workload and status.

Types of indicator

Obviously the choice of indicator reflects the different projects undertaken. However, there are also differences between the case studies in the degree to which attempts were made to trace all stages of the projects' development, which means covering a wide range of indicators related to outputs and process as well as impact. Doing this of course depends in part on the extent to which a project has been well documented. Those projects which had to do more to reconstruct project activities and indicators (CYSD, ENDA) tended to have greater difficulty in defining and prioritising key indicators, whereas those projects with a greater emphasis on monitoring and recording progress (Wajir, BRAC, Proshika) were clearer about which impact indicators were critical to assess.

Table 3.1 lists examples of the key areas of change, within which more specific indicators were chosen, examined in the BRAC, Proshika, and Wajir case studies. This indicates that while there are important differences between the contexts of the case studies from Bangladesh and the study from Kenya, they have certain basic categories or dimensions of change in common. There are three main categories.

Material wealth: This encompasses assets (land, cattle, housing), income, credit and savings, occupational status, wages, expenditure, food security and quality of diet, dependency on money lenders or on food aid.

Social well-being or human capital measures: These terms refer to health status and more specifically infant and child mortality, water and sanitation, and education — especially literacy and school-attendance rates.

Empowerment or political capital measures: These include ownership and control over assets, perceptions of well-being and quality of life, participation in decision-making and public institutions, access to public resources, dependency and mobility, and family-planning rates as a proxy for women's empowerment. It is important to apply this category specifically to women, whose measure of empowerment will in most cases differ significantly from that of men.

Whilst these broad areas of change are similar, the means by which they are assessed may differ. For example, in the pastoralist community in Wajir, Kenya, which has been severely hit by drought and insecurity, animal mortality, dependence on food aid and the law and order situation were deemed critical indicators of the security of livelihoods. In rural Bangladesh access to land and credit were assumed to be critical determinants of general well-being, and of women's well-being in particular, in addition to their control and ownership of assets.

Table 3.1: Key areas of change within which specific indicators were selected

BRAC	Proshika	Wajir
Economic well-being • land holding • occupation • assets • housing status • household expenditure and consumption • food security • credit and savings • ability to cope with crisis	Economic empowerment • indebtedness • assets • income • savings • investment • market mobility and power	Change in welfare/ livelihood • animal mortality • occurrence of peri-urban destitution • need for food aid • quality of diet • rate of return to investments provided through credit • law and order
Social aspects of well-being • literacy and educational level • health, sanitation, and family planning • demographic and other household characteristics	Social empowerment • literacy • health education and awareness • family planning • environmental awareness and practice • infant mortality	Social empowerment • school-attendance rates • parental satisfaction with education quality • reliability of water supply • child mortality
Women's empowerment • involvement in income-generating activities • ownership and control over assets • perceptions of own well-being • economic dependence on husbands • mobility	Women's empowerment • access to public resources • participation in local institutions	Women's empowerment • perceptions of changes in quality of life

By contrast, in the ENDA study each part of the project process was analysed at three levels: output, effect, and impact. The research team at first selected more than 20 indicators, which were subsequently found to be inappropriate and difficult to use as a basis for assessing impact. They therefore decided to begin indicator selection all over again in order to 'assess impact with a smaller number of relevant and manageable indicators. This gave some prospect of understanding the change that had taken place, rather than an unmanageable and ambitious list' (ENDA case study). ENDA's initial approach was to look for impacts and influences at every stage of the project-cycle. One of the assumptions they made was that each activity has an output that may lead to an impact.

The model therefore pointed to getting as much detail about the implementation process as possible. In retrospect, we feel that... a lot of time could be spent in the reconstruction and understanding of the project. Although an understanding of the situation and conditions that prevailed during the project is required, we found that despite the masses of data that were generated these data were difficult to relate to tangible project impacts. (ENDA case study)

The Pakistan study, mentioned above, uses a generic checklist of possible impacts (see Table 3.2) and then systematically examines changes in each area to determine if there have been positive or negative changes in each category, to record specific examples of those changes, and to assess the relative importance of each category.

Table 3.2: Checklist for dimensions of impact used in Pakistan case study

Dimensions of impact	Description
Life — health — security	• changes related to physical survival
Knowledge	• technical, practical, about others, about themselves
Excellence in work and play	• impact on skills used at work, and at home during relaxation
Relationships, especially gender relations	• within community, family, with outsiders, within group • between men and women
Inner voice	• at peace with themselves, with their conscience, sense of harmony
Empowerment	• ability to make meaningful choices and decisions and to influence others
Beauty/environment	• impact on environment, sense of harmony with nature: has the intervention created or destroyed things of beauty or culture?
Religion	• impact on deeper values, sources of meaning

This technique of offering groups a broad framework of dimensions, which they then define and explore from their own experience, seems to offer several benefits. First, it solves the problem of people being more likely to report those impacts that they think will be of interest to the NGO or researcher — in a purely open-ended process this can be a real danger. Moreover, this technique can also be a starting point for the development of participatory indicators to be monitored in the future, or for planning how the community can deepen certain positive impacts and address negative impacts. Finally, if the facilitator is accompanied by a discreet note-taker, this method can help the NGO to capture the impacts in the words of the beneficiaries.

Properties of indicators

The desired properties of indicators will depend very much on the approach adopted and the nature of the project. For the purpose of verification and planning, establishing appropriate indicators is a critical part of the process. The acronym SMART (specific, measurable, attainable, relevant, timebound) is a commonly used shorthand to describe the necessary properties of these indicators. Table 3.3 defines each of the characteristics in detail.

Table 3.3: SMART properties of indicators

Properties	Definition
Specific	Indicators should reflect those things the project intends to change, avoiding measures that are largely subject to external influences
Measurable and unambiguous	Indicators must be precisely defined so that their measurement and interpretation is unambiguous
	Indicators should give objective data, independent of who is collecting the data
	Indicators should be comparable across groups, projects thus allowing changes to be compared and aggregated
Attainable and sensitive	Indicators should be achievable by the project and therefore sensitive to changes the project wishes to make
Relevant and easy to collect	It must be feasible to collect data on the chosen indicators within a reasonable time and at a reasonable cost
	Indicators should be relevant to the project in question
Timebound	Indicators should describe by when a certain change is expected

Several of the case studies underline the difficulties that arise if some of these criteria are not met. For instance, in the CYSD study a household survey undertaken in 1992 stated that literacy levels of 60 per cent had been achieved in the village Berena by the end of 1992. But a repeat survey in 1997 has observed this to be only 38 per cent. On further examination, the researcher learned from the former survey team that a person who can sign their name was considered literate. However the 1997 survey defined literacy as the ability to read, write, and do minimum accounting. This lack of precision in the indicator — literacy — led to problems in assessing levels of change over time.

However, when indicators are used more as specific examples of change, different characteristics become important. These relate to the process of defining indicators as well as how they are measured. Table 3.4 outlines another set of characteristics, called SPICED (subjective, participatory, interpreted, cross-checked, empowering, diverse) in shorthand.

Table 3.4: SPICED properties of indicator development and assessment

Properties	Definition
Subjective	Informants have a special position or experience that gives them unique insights which may yield a very high return on the investigators time. In this sense, what may be seen by others as 'anecdotal' becomes critical data because of the source's value.
Participatory	Indicators should be developed together with those best placed to assess them. This means involving a project's ultimate beneficiaries, but it can also mean involving local staff and other stakeholders.
Interpreted and communicable	Locally defined indicators may not mean much to other stakeholders, so they often need to be explained.
Cross-checked and compared	The validity of assessment needs to be cross-checked, by comparing different indicators and progress, and by using different informants, methods, and researchers.
Empowering	The process of setting and assessing indicators should be empowering in itself and allow groups and individuals to reflect critically on their changing situation.
Diverse and disaggregated	There should be a deliberate effort to seek out different indicators from a range of groups, especially men and women. This information needs to be recorded in such a way that these differences can be assessed over time.

Clearly this does not mean that SMART and SPICED objectives and indicators cannot be combined. Indeed, most of the case studies attempted exactly that. In Wajir, quite specific indicators (such as the number of meals eaten in a day) were defined by the researchers on the basis of previous participatory planning exercises, and then verified by the population. These were combined with much broader, open questions such as asking people to score changes in their quality of life over time. This raises the question whether emphasis should be placed on developing indicators at the outset of a given intervention and on monitoring them or whether one should assess significant change regardless of pre-determined indicators.

In recent years there has been a growing interest in 'indicatorless' reporting, for example in the work of the Christian Commission for Development in Bangladesh supported by Rick Davies (Davies 1998). In this approach, staff are requested to report the most significant changes (positive or negative, planned or unplanned) over the last period and explain why they have chosen these. This exercise is then repeated at each level in the organisation.

Indicators and the project-cycle

The experience from the case studies suggests that the assumptions about indicator development and the project cycle-outlined in Chapter 2 are perhaps too neat and tidy. At the outset of a project it is important to understand both past *and* current change, as well as its indicators, and what has brought it about. This helps to understand people's changing circumstances and perceptions about the past before future desired change is explored (people's lives don't begin when projects start). Moreover, it is vital to understand existing trends so that any significant changes to those trends can be traced (which can make attribution easier — see p.79).

Equally, during the course of a project, monitoring current performance must be complemented by reviewing what has changed and looking to the future in order to modify existing indicators. In the same way, even an impact assessment undertaken many years after a project has finished (for example, ENDA in Zimbabwe) needs to consider future policies and practice rather than concentrating on past events — although this it may be of less benefit to communities no longer engaged with a given institution, than to the institution itself and the groups it might work with in the future.

It is not always possible or desirable to define impact indicators at the outset of a process and track these same indicators over time. In fact, the way in which people's own indicators of poverty change over time is an important element in understanding how their needs, attitudes, and values evolve. For

example, when exploring changes in women's situation, CYSD in India found that women stressed that the main difficulty they had faced before the project was male dominance and the subordinate role of females. But these women had not considered this as a problem when the project first started. Their involvement with CYSD and their changing self-analysis had led them to this understanding. The fact that women's indicators of change had evolved was an important indicator in itself that the project's efforts to change awareness and perceptions had borne fruit.

A recent impact study by Action Aid[1] also emphasises the importance of capturing *emerging* indicators. Interestingly, one reason why they came to this conclusion was the fact that women originally had not included changes in their own status as indicators, but subsequently did. The study proposed that changes in the type and nature of indicators was in itself an important indicator or 'meta-indicator'. For women, these might include change in the number of indicators they propose; change in the scale of desired change; change in the degree to which indicators are also desired by men or not.

The Wajir study came to a similar conclusion.

[T]he impact on project beneficiaries livelihoods and welfare often calls for the continuous integration of (new) impact indicators during the project implementation period. Only in this way can we know whether the project is having any impact on the beneficiaries. The constant integration of impact indicators ... is also necessary because of the uncertain and unpredictable nature of the project environment. (Wajir Economic Impact Assessment)

Summary 3.2 outlines some of the key lessons learned about indicators and their measurement.

Summary 3.2 : Key lessons about indicators

- At the outset of projects, determine key areas of change, as well as some specific impact indicators, with ultimate beneficiaries and local staff. This is necessary in order to assess progress, and can be empowering in itself, but it is also useful in meeting the demands of other stakeholders, particularly funders.
- Ensure that these indicators are sought from different groups — men/ women, well-off/ less well-off, etc. — and differences noted.
- When circumstances change, update and reformulate existing indicators, as well as introducing new indicators and dropping others. This requires monitoring the project's context and environment.

- Reduce the number of impact indicators to a manageable proportion based on key areas of change. Anecdotal indicators can be used to exemplify past changes in key areas of people's lives, even if they have not been pre-defined. These may need to be interpreted so that others can understand their significance.
- Explore significant changes which occurred as a result of the project/programme, but which lay outside these initial indicators. Use this information to develop indicators for the future.
- Deliberately set out to capture negative change and to seek out those who might report it, particularly groups who are often disadvantaged such as women, minority groups, or people who have dropped out of the project.

What are the units of assessment?

Any impact assessment exercise needs to determine what its key units or levels of assessment are. Will the study focus on change at the level of individuals, communities, organisations, or all of these? What are the advantages and disadvantages of concentrating on one level as opposed to another? Clearly this decision depends on the objectives of the programme in question and the types of impact to be assessed. However, being clear about what needs to be assessed at different levels can help to focus studies and concentrate resources, as well as helping to understand important linkages between these levels.

Although all studies looked at more than one unit of analysis, they varied in emphasis. For example, BRAC, in order to assess the impact of its programme on material well-being, used the household as the main unit of analysis as 'it is the household as a whole, not the programme participant alone, which experiences the impact of the programme intervention' (BRAC). In the NK and GGS study in Bangladesh, Rao and Hashemi emphasised village organisations and their evolving relationships with intermediary NGOs, concentrating on transformation in the community. Several of the studies explore the relationship between the individual and groups/organisations as well as on wider institutions (Ghana, Wajir, CYSD).

The following lists the levels explored in the case studies, with the main focus in italics.

Ghana: *individuals*, households, communities, *support NGOs*

CYSD: *individuals, households*, CBOs, *community*

Wajir: individuals, *households*, *CBOs*, support NGOs, institutions

Ikafe: *individuals*, household, *communities*, institutions

Pakistan: *individuals*, organisations

Matson: *individuals* and organisational

Proshika: *individuals, households, groups*, villages

BRAC: *individuals, households, groups*, villages

NK and GSS: individuals, *groups*, federated structures, *support NGOs*

CORDES: *individuals*, household, *communities*, institutions

ENDA: *individuals*, household, *communities*, institutions

It is clear that nearly all the studies focused particularly at individual and household levels. Some looked at organisational change within community based organisations and local NGOs, and fewer still attempted to look at changes in community or societal norms. Within this some, but not all, of the studies looked at women's position, status and empowerment in particular, at both an individual and household level.

While this primary focus on individuals and households is understandable, focusing on only one level provides a very partial picture of impact. Given the complexity of including all these levels in an assessment, and given the definition of impact as significant change in people's lives, there are advantages and disadvantages in focusing on a range of units or levels of assessment. As we see from Table 3.5, focusing on more than one level not only allows for a wider range of impacts to be explored, but also allows the linkages between levels to be examined. However this can also make the study more complex and time-consuming. Prioritising the most important levels and linkages and focusing on these, is necessary if the study is to remain manageable.

Table 3.5: Advantages and disadvantages of various units of assessment

Unit of assessment	Advantages	Disadvantages
Individual	• Easily defined and identified • Allows social relations and gender issues to be explored • Allows inter-household relations to be explored • Can allow personal and intimate issues to emerge	• Most interventions have impact beyond the individual • It may be difficult to speak to the most marginalised people • Difficulty of attribution through long impact chain • Difficult to aggregate findings

Table 3.5: Advantages and disadvantages of various units of assessment (continued)

Unit of assessment	Advantages	Disadvantages
Household	• Permits appreciation of income, asset, consumption and labour pooling • Permits appreciation of link between individual, household and group/community • Permits understanding of links between household life-cycle and well-being	• Exact membership is sometimes difficult to assess • Inter-household relations are often ignored
Group/ CBO	• Permits understanding of collective action and social capital • Permits understanding of potential sustainability of impacts • Permits understanding of potential transformation community in the community Exact membership is sometimes difficult to assess	• Group dynamics are often difficult to understand • Difficult to compare using quantitative data
Community/ village	• Permits understanding of differences within the community • Can act as sampling frame for household/individual assessments • Permits understanding of collective action and social capital • Permits understanding of faction and clan relations • Permits understanding of potential transformation community in the community and beyond	• Exact boundary is sometimes difficult to assess • Community dynamics are often difficult to understand • Difficult to compare

Table 3.5: Advantages and disadvantages of various units of assessment (continued)

Unit of assessment	Advantages	Disadvantages
Local NGO	• Permits understanding of potential sustainability of impacts • Permits understanding of changes brought about by capacity-building • Allows assement of performance (especially in terms of effectiveness and efficiency) • Allows exploration of links between change at the community, group, and the individual level	• NGO dynamics are difficult to understand • Difficult to compare various local NGOs
Institutions	• Permits wider change and influence to be assessed • Permits assessment of how favourable future context is likely to be in helping sustained change to continue	• Greater problems of attribution • Internal processes and dynamics are difficult to explore or understand

Adapted from Hulme (1997) and related to case studies

What information already exists?

Once it is clearer what information is required and what levels of assessment are appropriate, it is important to find out if the information that is required already exists, or there are systems in place for its collection. The collection of secondary data is important, not least in order to ensure that individual's and communities' time is not wasted collecting information that already exists. In addition this data can also reveal gaps in official records, existing trends, and contradictions between official statistics and the study's findings.

In the case studies secondary data were collected from the following sources:

Literature reviews: Researchers searched academic and aid agency sources to discover more about specific issues, such as the relationship between local institutions and poverty in Northern Ghana (Wolmer 1996), or on the issue of

impact assessment itself (BRAC). In the case of Ghana, although it took some time for this material to be absorbed by those undertaking the impact assessment, it did allow village-level findings on changing gender relations to be verified by comparing them to longer-term anthropological work. Literature reviews also provide important information on changing trends in the area as a backdrop to exploring changes brought about by NGO projects,

Official records and surveys: Undertaken by government agencies, multilateral institutions, research institutes, or other aid agencies, these are valuable sources of data that can inform impact assessment exercises. Some of the major sources of information which may be of use are listed in Summary 3.3. As I mention later in the section on reconstructing baselines, using data from external sources, particularly government agencies, on education, agriculture, and health was important in several of the studies.

Project document and record reviews: The first phase of the Wajir mid-term review consisted of project staff gathering information for their 'mid-term review status report'. This report summarised information that had already been collected by the project as part of the monitoring process, and included issues which had been raised by project staff as important and considered necessary for the review to look at. Although this preparatory work at the time seemed over-long to staff, they recognised its value at a later stage. For example, the impact assessment team in the ENDA case study prepared a 'baseline analytic report', which was then presented to all ENDA staff at an internal workshop.

Box 3 indicates some key sources of secondary information identified by Renata Lok of the United Nations. Certain ministries, such as Health, Education, Social Welfare, Agriculture, Employment, and Public Works can all provide useful information in addition to National Statistical Offices which may have a role in synthesising information and therefore may already have done some of the analysis of data that may be important for impact assessment purposes. In addition, a number of relevant specific surveys, often undertaken by International Agencies in co-operation with local ministries, may also exist. The kind of survey pertinent to impact assessment exercises includes the household survey; specialised surveys which analyse income, consumption, agricultural production, prices, employment rates, and so on; and hybrid methods which combine household or individual data-collection with other survey methods.

Of particular relevance among the hybrid surveys are sentinel studies (for example, those undertaken by UNICEF which track changes in the health of a given group of people over time); knowledge, attitudes and practices (KAP) studies, which try to look at changes in areas that are often more difficult to

assess; and participatory poverty assessments (PPAs), which have been used mainly by the World Bank, and attempt to combine participatory approaches of assessing poverty and its causes with data gathered and analysed in more formal ways.

Summary 3.3: Key external sources of information

Government sources

Ministry of health: May have data on health services provided through hospitals, clinics, and health campaigns (such as vaccination drives), including information on child nutrition status, disease incidence, inpatient and outpatient visits, and so on.

Ministry of education: Source of data from schools on numbers and profiles of students and teachers, educational attainment, literacy levels etc.

Ministry responsible for social welfare: May have synthesised local records on the poverty status of households which some countries maintain.

Ministry of agriculture: May hold records of poverty-related services.

Ministry of labour or employment: Source of unemployment and wage statistics.

Ministry of finance: Will have financial data, consumer price trends, and so on.

Ministry of public works: Has data on water provision, sanitation, and distribution of electricity.

National statistical office: In some countries, this institution may collate a lot of the above data.

Surveys carried out by international agencies and government ministries

Multi-topic household surveys

- Living standard measurement surveys (LSMS)
- Surveys associated with social dimensions of adjustment (SDA) undertaken by the World Bank: the integrated survey (IS) and the priority survey (PS)
- 'Core welfare indicators' questionnaire

Specialised surveys

- Household income and expenditure surveys (HIES)
- Demographic and health surveys (DHS)
- Labourforce and employment surveys
- Food-consumption and nutrition surveys
- Consumer price surveys
- Agricultural surveys, agricultural sample surveys, agro-economic surveys
- Other specialised household surveys

Hybrid survey methodologies

- Sentinel site surveillance (SSS)
- Knowledge, attitudes, and practices (KAP) studies
- Poverty assessments and participatory poverty assessments (PPA)

Source: Renata Lok Dessallien (1999)

Who should be involved?

Chapter 4 gives a detailed account of the tools and methods which were used to involve a wide range of stakeholders, particularly beneficiaries, in the case studies, and discusses some of the associated problems. Here I will simply describe the main groups involved in the studies and look at issues relating to the composition of the assessment team. I return to the question of how to assess the quality of participation in the final chapter of the book.

Men and women, rich and poor

As stated in Chapter 2, issues of power and participation are central not only to impact assessment but also to understanding impoverishment and injustice. Some of the case studies were more concerned with distinguishing between project and non-project groups (see below). Others made more effort to ensure the involvement of both men and women or worse-off groups; other aspects such as age and ethnicity were also relevant in several studies. In some cases, specific research was undertaken in order to identify who to involve, such as wealth-ranking exercises which determined the least well-off households (see next chapter). In others, targets were set for the proportion of women to be interviewed in a given community, and reserachers ensured that various members of a household were interviewed. Those studies which considered gender-related issues particularly important made greater efforts to ensure not only that these issues were explored

(which was done in most studies), but also that the choice of assessment team members, the methods used, and the timing of the exercise were all designed to making women's involvement easier.

Beneficiaries

All the case studies involved the ultimate beneficiaries in some way. However, both the proportions of beneficiaries involved and their levels of participation varied enormously. In the CORDES study, almost the entire population of the three communities that are part of the project participated throughout the study and were involved in its design. In the studies involving much larger programmes, such as BRAC or Proshika, only a small sample of the many thousands of village organisations they support were involved. Table 3.6 indicates the level of participation at different stages of the impact assessment process in the case studies. It indicates that despite the willingness to involve beneficiaries, few of the studies managed to do this at *all* stages of the process. Although beneficiary groups were involved in most or all of the case studies during the reconstruction of project histories and in determining and verifying indicators, they participated much less in defining what was meant by impact assessment, in the design stage, and in the analysis and feedback stages towards the end of the process. This suggests that despite the professed desire of all the organisations to involve communities in the impact assessment process, there were still significant differences in the depth and quality of participation.

Non-project respondents

Non-project respondents are individuals who have deliberately not been involved in the projects assessed, but who have similar characteristics to those which these projects seeks to support. They were interviewed in a number of case studies which used them as 'control groups'. The lessons from comparing changes in the lives of project and non-project respondents are described in detail in this chapter's section on attribution.

Excluded and drop-out groups

It is well known that some members of the community which an invention is intended to benefit may well be excluded from or even disadvantaged by it. Some of the case study teams sought to identify these groups and individuals in order to understand why this had happened and what negative effects it might have had. In both the Ghana and the UK case studies, efforts were made to contrast the views of these 'excluded' groups with those of project participants. In other cases, project participants who had left the project or dropped out of groups the project was supporting, were identified and interviewed.

Table 3.6: Participation of beneficiaries at different stages of the impact assessment process in the case studies

Stage of impact assessment process	Occurrence in case studies
At the outset, defining what is meant by impact and designing the impact assessment study	Beneficiaries were fully involved in this stage in only one of the case studies, although partially in one other.
Reconstructing project history and changes in the context	Benficiaries participated in most case studies, although in some, assessment of changes was limited to changes directly associated with the project in question.
Determining the areas of change, objectives, or indicators necessary to assess past or future impact	Beneficiaries were involved in most case studies. In some cases, people did not participate in prioritising the updated indicators, although they might have been involved at an earlier date.
Verifying if project objectives had been achieved, or indicators met	Beneficiaries took part in this in nearly all the case studies, although in a few cases, objectives and indicators were unavailable or not sufficiently specific to do this.
Analysis and interpretation of results	In some of the studies, benficiaries participated as a matter of course; in others in some aspects; and in some hardly at all or on a very limited basis.
Feedback of results	This occurred in a few of the studies, both informally and as a result of a systematic attempt to feed back to beneficiaries.
Presentation and communication of findings	This really happened only in one case study, where it was part of the study design.

Project staff

The involvement of project staff varied particularly according to the size of the project, its stage in the project-cycle, and the degree to which the exercise was also perceived as a capacity-building process for staff themselves.

However, as the ENDA study revealed, the involvement of project staff — although it has many advantages in terms of building 'ownership' of the process, in providing the exercise with unique knowledge and insights, and contributing to feasible future plans and policies — also has a down side. Staff may fear that the evaluation will affect their jobs, reputations, or future resource allocation, in which case they may feel defensive, become obstructive, or seek to manipulate results so as to promote themselves.

Donors

Although donors were only explicitly involved in a few case studies (for example, agreeing the terms of reference and scope of the impact assessment or attending introductory workshops), their unseen presence was very much evident in considerations of study design and implementation. This particularly affected the selection of the impact assessment team, which projects were selected, and the degree to which independence and 'objectivity' were sought.

Others

Local government, other NGOs, and international agencies also were important stakeholders in some of the case studies. This was particularly true of the refugee programme in Ikafe in Northern Uganda. Chapter 5 explores the issue of how this impact assessment exercise attempted to cope with the diverse perspectives involved.

The impact assessment team

Depending on the purpose of the study, the information required, and its likely sources, decisions must be made about whether an impact assessment team needs to be established and who it should include. The case studies vary in their approaches, in their use of local project staff or external specialists, and in the degree to which external staff were perceived as evaluators, people offering support, facilitators, or peers. Table 3.7 summarises the degree of involvement of outsiders, ranging from purely external teams in two cases, mixed teams in two cases, to internal teams supported by external specialists or internal research units in six cases. The Matson project team in the UK invited an external reviewer as a peer and facilitator rather than as an evaluator.

Clearly the use of external specialists will also depend on the level of experience available 'in-house', the degree of independence or 'objectivity' certain stakeholders may demand, and the extent to which the exercise is seen as a training or capacity-building process in its own right. The Ikafe and the CORDES studies were the only ones where representatives of beneficiaries were included in the assessment team.

Table 3.7: Degree of involvement of outsiders in the case studies

Degree of involvement of outsiders	Case study
External evaluation team	Wajir, Kenya; NK and GSS, Bangladesh
Internal evaluation or research team:	
• supported by external specialists	Pakistan; Ghana
• run by, or supported by, internal research/ evaluation unit or impact assessment co-ordinator	ENDA, Zimbabwe; BRAC and Proshika, Bangladesh; CYSD, India (sometimes with external support or with specific tasks contracted out)
Mixed evaluation team involving outsiders, staff, and beneficiaries	Ikafe, Uganda; CORDES, El Salvador
Peer review by outsider	Matson, UK

There are well-known advantages and disadvantages to various combinations of 'insiders' and 'outsiders' in a team (Rubin 1995). However, the case studies reveal that the individual's attitude and skill, as well as their ability to build rapport, are more important than their affiliation. For example, Stan Thekaekara from India, who undertook the Matson study in the UK, was able to gain important insights from informants due to good interpersonal skills (not least with children thanks to his magic tricks!), many years' experience in grassroots work, and an inquisitive yet humble approach. Over an eight-week period, Stan mainly used participant observation techniques, which allowed him sufficient time to get to know the community and vice versa. The message that Stan was a colleague involved in similar work elsewhere and that he also wanted to learn from Matson meant that he was not seen as an expert coming to judge the project. He offered the project staff some very useful insights into their work, but he was essentially a peer who had come to listen, learn, and share what he had learned with project staff, community members, and other stakeholders.

The definition of who is an 'outsider' varies enormously. For example, in the ENDA and Ghana studies the impact researchers were 'outsiders' to the communities where the assessments took place, although they were local staff of the agencies involved in the assessment (ENDA, Oxfam, and ISODEC). They too faced difficulties, in having to overcome the attitudes of both staff and community members who saw them as a threat to the continuation of the project, in sometimes having to work through translators,

and in having to overcome their lack of specific knowledge of the projects or communities in question. However, there also were advantages: in Ghana, people felt able to talk about certain aspects of their lives with outsiders, precisely because they were not members of the community or project staff. Yet because this sort of information is confidential, great care must be taken to ensure that it remains so not only in the final report, but also in any feedback sessions held with the community.

In most case studies care was taken to ensure a gender balance on the research teams, or to include at least one woman. The benefits of this are well known. A review workshop on the impact assessment work in Ghana revealed that the female researcher managed to elicit sensitive information in individual interviews related to child-spacing, household relations, and female income which would have been difficult, if not impossible, for a man to find out. But workshop participants also pointed out that the researcher's attitude and approach were equally critical (being a woman may be a necessary but not sufficient condition!). Some of the information was revealed not simply because the researcher was female, but also because she was an outsider.

Most of the impact assessment teams were made up of two to three people, with the exceptions of BRAC, Proshika, Ikafe, CORDES, and Matson. In Matson, the main reason for this was the approach adopted: participant observation by a peer over an extended period. In Ikafe, the complexity and insecurity of the situation (involving refugees and host populations) and the number of stakeholders involved (central and local government, several relief agencies, United Nations specialised agencies, refugees and host populations, and private contractors) prompted the researchers to establish a large team with representatives from all the major actors. At the end of the study the managers acknowledged the risk inherent in this choice.

The large number of team members / facilitators produced very much information and analysis, especially through working in sub-teams. The team functioned well and was able to discuss, cross-check and reach consensus on many issues, but such a large team and process if not successfully managed can also result in unmanageable, unfocused amount of information from which no consensus analysis arises. On reflection a smaller team would be recommended for future work of this kind. (Neefjes et al. 1996, p.48)

In creating a team, the challenge is to combine individuals whose skills complement each other. There are therefore several considerations for establishing an impact assessment team.

Questions of training and support

External support at the outset

Two of the case studies (Ghana and Pakistan) used external trainers or facilitators at the outset of the studies in order to develop team skills for impact assessment work. In both cases this involved field-based exercises using a variety of participatory research tools (see Chapter 4). This approach has the advantage of providing hands-on experience and a practical understanding of the advantages and disadvantages of different tools and methods. However, care must be taken to ensure that staff do not feel they have to use the same number of tools and methods in each community, but that they are familiar with a range of ways in which questions about impact can be asked and answered.

Support from internal staff

In five of the case studies (BRAC, CYSD, ENDA, Proshika, CORDES), most of the training, support, and advice was provided by internal research or monitoring and evaluation units, sometimes supported by sympathetic local academics. Clearly internal support offers major advantages such as continuity and an understanding of the context, programme, and organisation. If these units have links with wider networks of expertise, they can also communicate critical external experience, in a way that is appropriate to the local situation and the organisation.

However, there can also be problems with internal support: satff may be busy with other tasks, which means that the necessary advice is not available when needed; they may lack the expertise relevant for the research in question, and feel reluctant to admit it; they may have personal interests in the organisation which are threatened by building the capacity of others or by the discovery of certain findings; they may represent certain interests according to their gender, class, ethnicity, race, faith, and so on; they may be seen as the representatives of senior management and project staff may therefore be less likely to divulge problems, errors, or mistakes in the fear that this may threaten their jobs or affect future resource allocation. Of course external trainers also have their own personal interests and biases, but internal units may have institutionalised dominant personal interests which can influence the support and advice that is given, particularly to contentious impact assessments.

External team which includes capacity-building in their work

Three of the case studies (Ikafe, Wajir, Proshika) involved external evaluators whose role it was, either implicitly or explicitly, to undertake the assessment in

such a way as to enhance local staff and partners' impact assessment skills. This is difficult, and there is a risk of doing neither the capacity-building nor the impact assessment properly. In Wajir, however, the evaluation team gave staff some helpful advice on how to improve their 'impact tracking' which also validated a more flexible approach that recognises the importance of capturing emerging impact indicators as well as monitoring existing ones. The way in which the assessment was conducted also provided useful experience in determining how the project could monitor impact, rather than just inputs and outputs.

Equally, the Proshika experience highlights the important role that credible external support can play in institutionalising new (in their case participatory) approaches to impact assessment, especially if there is some scepticism about how these compare with existing approaches.

Peer review/ mutual learning

The Matson review was set up in a different way to all the other studies. Oxfam acted as a go-between by putting a neighbourhood project team, who wanted to learn more about impact assessment, in touch with an Indian community development worker who wanted to explore cross-cultural community exchanges. Therefore the relationship was one of peers seeking to learn from each other and gain experience. This involved no training as such, but a lot of observation, debate, and exchange of ideas — some of it relating to the impact assessment, but a lot of it developing in wholly unpredictable directions, such as the possibility of Matson residents bagging tea sent from Southern India for sale in the UK. This review is now seen by both parties as a first step in developing longer-term relations between the two communities, which may include training Matson staff and volunteers in community animation techniques developed in India.

Summary 3.4 lists the main lessons learned about establishing an impact assessment team.

Summary 3.4: Questions to consider when establishing an impact assessment team

- What balance of insiders and outsiders is appropriate? Is the best way to ensure that the views of some stakeholders are voiced to havie them represented on the team? Are there alternative, better ways of doing so?
- What is required to achieve an adequate gender balance on the team?
- What are the attitudes and behaviour required of all team members? How can these be verified?

- What technical/ sectoral knowledge is required?
- What conceptual and methodological skills are required in order to solve problems specific to impact assessment, such as attribution?
- What balance of quantitative and qualitative methods will be used and what skills will these demand?
- What co-ordination, facilitation, and diplomatic skills are required, particularly in managing the assessment?
- How will the findings be communicated? What writing and communication skills are needed for this?
- What is the best way of ensuring that the findings actually make a difference to people's lives, to the projects being assessed, or to the organisations involved? Who might need to be involved in order for this to happen?
- What kind of training and support will team members require? What resources are available for this? What degree of capacity-building of project staff, partner organisationsm and beneficiary groups is possible or appropriate?

Sampling

There is very rarely the time or the money to talk to, or contact, everybody in a given community, organisation, network, or region. This means choosing or sampling from a larger 'population'. In the following, I describe some of the types of sampling that exist. Most observers distinguish between sampling that is done in order to generalise findings about a given group of people or organisations (which usually involves random sampling), and sampling undertaken in order to identify specific groups of people or organisations about which more information is needed (which usually involves non-random sampling). Again, it is crucial to be clear about the purpose of the study, because this will determine the sampling frame or the 'population' under study.

The sample population may involve different units of assessment such as individual women, men or children, households, communities, organisations, or institutions. For example, the Ghana study team wanted to explore the relationship between organisational development and poverty reduction; they therefore chose a sampling frame of the 30 organisations which are part

of the Northern Ghana Development Network and the communities they supported. By contrast, the Matson study's sampling frame included the entire population of Matson neighbourhood.

Once the type of sampling and the units of assessment have been identified, the next step is to decide on sample size. Cost, staffing levels, and availability of researchers, and logistics often are the key factors in determining sample size, as is the willingness of organisations or communities to get involved. However, one must also consider how valid particular sample sizes will make the study. If the study primarily aims to confirm that existing small-scale findings apply to the entire population, and to do so in a way that is statistically significant, then there are agreed statistical procedures for estimating how confidently this can be done for different sample sizes (Paul Nichols, pp.53-56). If, however, the aim is to understand in depth how a given impact was achieved or to explore possible impacts, a much smaller sample size will probably be appropriate. But again, sample sizes must be set without losing sight of practical concerns.

[A] small sample, properly managed and carefully analysed is always better than a poorly supervised, large sample which is never fully analysed for lack of time. Will the sample size you can afford answer your research questions? If not, then consider how you might either get more resources, or scale down your research questions to something more realistic. (Nichols, p.53)

There are numerous methods for selecting the sample once the sample size is clear. Summary 3.5 describes some of the best known types of random and non-random sampling as well as repeat sampling methods.

The case studies mainly used non-random sampling methods, although larger studies (Wajir, Proshika and BRAC) did use stratified sampling methods such as random sampling within particular sub-groups. In addition, most studies went through staged sampling processes, although not usually on a random basis. The Ghana study team first selected three organisations from the 30 members of the Northern Ghana Development Network, according to criteria determined by the researchers and the Network. They then selected three communities from among those the organisations worked with. Within each village, ten men and ten women were selected following wealth-ranking exercises, in order to include each wealth grouping in repeat interviews over a longer period. This process was chosen in order to develop tools and methods for impact assessment which could capture the opinions of those whose voices were seldom heard, as well as to understand the relationship between organisational support and poverty reduction.

Summary 3.5: Types of sampling

Random sampling

Simple random sampling: A group of people are selected at random from a complete list of a given population.

Stratified or systematic random sampling: This ensures that sub-groups within a population are included in the sample, by randomly sampling within each of these sub-groups.

Cluster sampling: By selecting geographic clusters of villages or households within a given population, time and money is saved; this technique thus allows more people or groups to be contacted in the time available.

Staged sampling: For large populations, one may need to sample within samples. For example, BRAC's Rural Development Programme works with 63,846 village organisations through 372 area offices. Each of these villages contains a number of households. Therefore BRAC selected a sample of areas, within which a random sample of households was interviewed.

Random walk: Instructions are given to the interviewer to follow a random route and interview individuals ('take first road right, interview at second house on your left, continue down the road, interview tenth household on your right', and so on).

Non-random sampling

Quota sample: Based on information about a population, quotas of certain types of people or organisations are selected for interview; common criteria for quotas are age, gender, occupation, and whether people live in project or non-project areas.

Genealogy-based sample: Select entire families and their relatives rather than households.

Chain sampling or snowballing: Select a first contact and then ask them who you should talk to next. This method is useful for identifying minority groups or occupations within communities.

Matched samples: Similar pairs of villages, projects, or groups of people are selected in order to compare them (project groups and non-project groups are an example).

Repeat sampling methods

Panel or cohort surveys: A set of people or organisations is contacted several times over a relatively long period.

Repeat survey: The entire survey process is repeated, including sampling.

Rotating survey: This is a combination of the panel and repeat survey methods: one fraction of the sample is changed each time the survey is repeated, another fraction remains the same.

For more details see Nichols, P (1991) *Social Survey Methods: A field guide for development workers*, Oxfam Development Guideline No.6

The Proshika study used a two-stage cluster-sampling technique. As the organisations works in over 4,000 villages, obtaining a complete list of all households in these villages was considered unnecessary and too costly. Instead, they first 'clustered' the villages into units of about eight villages each. As a result, they reached a number of 500 clusters in areas where Proshika worked and 380 clusters where they did not. The researchers selected nine of these clusters from areas they worked in and ten from those areas where they were not involved. The second stage was to compile a complete list of households for these 19 clusters, from which 190 households were selected at random — 100 from Proshika areas and 90 from other areas.

Although the larger studies used quite sophisticated sampling procedures to generate representative findings, they also included non-random samples for more in-depth, qualitative assessments and for case-study material. Thus BRAC, in addition to the 1250 BRAC sample households and 250 non-BRAC households which were randomly selected, also surveyed 200 households who had demonstrated very high economic performance as well as 25 village organisations. The Wajir study for example added women who were beneficiaries of a loan scheme, and families who had been involved in an animal restocking project, to its original sample as well as selecting a number of Pastoral Associations and Women's Groups for workshop and focus-group discussions.

Given that impact assessment usually attempts to combine both quantitative and qualitative questions, sampling is likely to include a number of important 'purposive' criteria — criteria that are deemed critical because of the research team's knowledge of the area, the organisations involved, and the project objectives and history. Some of the key 'purposive' criteria used for sampling in the case studies are included in Table 3.8. There are three main groupings:

context-related criteria, which are about ensuring that social, economic, and environmental differences are taken into account; organisation-related criteria, which range from the pragmatic — looking at capacity and willingness to get involved — to questions of stakeholder participation; and project-related criteria, which aim to ensure that both beneficiaries and non-beneficiaries are included, and that those involved in different ways and for different lengths of time are adequately represented.

Table 3.8: Key 'purposive' criteria used for sampling in the case studies

Context-related	Organisation-related	Project-related
• Gender and age • Wealth and well-being • Geography and agro-ecology • Proximity to markets and roads • Cultural/ ethnic composition	• Willingness and capacity to get involved • Balance of types of organisations, for example community-based organisations and intermediaries • Stakeholder balance • Mix of targets for advocacy/ influencing	• Project and non -project sites • Participants and non-participants of the project within the same community • Good and poor performers, including disbanded groups and drop-outs • Length of involvement in projects/ programme (asking whether an intervention has been running for long enough to have made an impact, or selecting from a scale of length or intensity of involvement in order to compare) • Involvement at particular moments of the intervention (a community was the project entry point, or the last area to be included in a programme) • Involvement in particular interventions • Availability of information • No, or not too many, previous studies

Sampling represents a critical stage in most impact assessment work and can affect the results as well as the perception of the study by others. Key questions that might be asked at this stage are included in Summary 3.6.

Summary 3.6: Some key questions related to sampling

If the sample claims to be representative of a larger population:

- What might have occurred to make the sample atypical of the wider group?
- Could certain types of participant be less likely to be selected than others?
- Could pragmatic criteria such as cost or time constraints introduce bias into the sample selection?

If the sample's main purpose is to identify particular groups or people and to find out more about qualitative aspects or impacts, other questions may be more relevant:

- Does the sample cover those whose views and opinions are particularly important or normally overlooked, in particular women and the poorest groups?
- Whose views and opinions will not be covered by a given sample, and does their exclusion matter?
- Does the sample cover all groups likely to have differing opinions or views? Does the sample help us understand the linkages between different units of analysis (such as individuals and organisations)?

Timing of assessment

The timing of the assessment is another important issue and relates to the project-cycle, to seasonal conditions, and to people's daily rhythms and work schedule.

Timing and the project-cycle

When is it appropriate to carry out an impact assessment? As we have seen, in an impact assessment exercise, certain activities have to be undertaken throughout the project-cycle in order to determine objectives and indicators, monitor progress to date, and to monitor and adapt to changes in the external

environment. However, there are probably moments when intense reflection on and analysis of impact are more appropriate then others.

The first such moment is the initial situation analysis and appraisal, when likely future impact is assessed and baseline data collected. The second occurs towards the end of a specific phase in the project-cycle, when a new phase is being planned, or at other key stages of an intervention. Such interim assessments may be combined with lighter, more frequent periodic reviews of impact, say, on an annual basis. As the Wajir study shows, 'before' and 'after' studies between which a long time has elapsed are unlikely to reveal much about the sequence of cause and effect, nor will they permit 'impact tracking' to occur, which would allow an almost continuous integration of new and adapted indicators during implementation. His lack of continuity particularly applies in areas of high uncertainty, for example regarding security, and great fluctuations, for example in rainfall. Thus ongoing impact monitoring is vital in order to gather evidence about causality and to adapt to changing circumstances. The frequency with which it needs to occur during implementation will clearly depend on the context and how stable it is.

Lastly, there are 'terminal' evaluations which occur soon after a programme has finished or several years later. As the ENDA case study reminds us, extra care must be taken to (re)-gain people's confidence when the research team is going to a community where a project has not been running for a number of years — even if the researchers are from the same agency which originally supported the project.

Figure 2 in Chapter 1 on the case studies illustrates where each of the case studies 'fits' in terms of timing.

Seasonal issues

It is important at what time of year information is collected, not only because people's answers to some questions will change during the year, but also because their availability and willingness to respond will vary with the seasons. For example, BRAC had to delay field work during Ramadan, the Muslim month of fasting. While the 'hungry season' in Ghana is a critical opportunity for understanding people's coping strategies, it is also a time of great stress and peaking labour demands. Involving people in lengthy interviews or workshops at this time of year, even at times suitable to them, may not only lead to poor quality information but can be ethically dubious.

Particular activities may also show different results in different seasons, which can affect the impact assessment process. For example, in order to conduct an assessment of a well-digging component of their work, researchers in the CYSD study visited well sites in the summer, when the water table was at its lowest. They found out that the water-lifting devices they had introduced

could not lift water efficiently at this critical time of year due to the low depth of water in the wells. If they had visited at another time of the year the problem may not have been so obvious. In order to overcome some of these problems, BRAC collected certain items of information (on consumption of and expenditure on non-durable goods) twice — once in the lean season and again in the peak season — in order to take into account seasonal fluctuations.

The question of seasonality is also important if baseline data or previous surveys are to be compared with new information. This is particularly the case if questions you include questions such as 'how many times a day did you eat in the last week?'. If this sort of information is not compared to equivalent periods of the year it will be pretty meaningless. It should also be remembered that even if similar periods are compared, fluctuations in rainfall, the economic situation, or social conditions may be much more likely to explain any differences than the project that is being assessed.

What day of the week and time in the day to use for assessment

Times that are most convenient for villagers may be least convenient for researchers and vice-versa. Some of the studies made particular efforts to overcome such problems.

- Researchers must recognise that there is a gendered aspect to people's time availability and that researchers must make special efforts to talk to women. A team of women who carried out an evaluation for Oxfam in Tigray observed that women there had no words in their language for leisure.[2]
- The research team should identify when people might be available for interview or workshops and arrange them accordingly.
- Particular days in the week, festivals or social events should be avoided.
- Researchers should make observervations and engage in conversations in informal surroundings (at a well, in markets, at tea stalls, in bars) without interrupting people's daily routines or meetings. In some cases, interviewers can help interviewees with their work, for instance by sitting with them and picking stones out of grain for cooking.

Despite such attempts to be considerate, several of the studies indicate that time availability for both researchers and respondents, particularly women, was a real issue and in some cases this led to the studies being simplified, particularly time-consuming methods being abandoned (see next chapter), and the worth of group meetings which only the 'time rich' could attend being questioned.

Summary 3.7 shows some of the key questions that need to be asked about the timing of different elements of impact assessment work.

Summary 3.7: Some questions related to the timing of an impact assessment process

- What stage of the project-cycle is the most appropriate to reflect on and analysise impact?
- How regular does the collection of impact data need to be to cope with changing circumstances and the difficulties of attribution?
- What season or time of year makes most sense for the collection of impact data? Does certain information need to be collected during more than one season?
- Are there particular days of the week , or times of day, that should be avoided when collecting impact data? Who might be excluded if certain times of the day (or night) are chosen to carry out interviews or meetings?

Where there is no baseline

The importance of having a starting position from which to measure change is perhaps one of the most common points made in impact assessment texts (Oakley et al. 1998). Within the case studies we find a variety of situations: where no baseline study had been systematically carried out (CORDES, Pakistan), where a baseline study or survey had been done but was inaccessible or unavailable (CYSD), where a baseline study did exist but did not contain all the necessary information to assess impact (BRAC, Wajir, Ikafe, Proshika), and where the impact assessment exercise was self seen as a means to establish a baseline for the future (Ghana).

Therefore all the studies, in one way or another, had to reconstruct the past. This not only underlines the need for baselines, but also illustrates what their limitations are. There are two main factors that make baselines a problematic tool. First, it is impossible to predict all the information that might be needed because information needs will change over time. This may occur because of changes in the environment (for instance, drought affected the ENDA-supported projects in Zimbabwe and the Wajir programme in Kenya), or as a result of changes brought about by the project. For example, in Matson the success of the community in resisting council-house sales encouraged them to get involved in improving other aspects of their lives, such as child-care or mental health, for which no baseline existed. A third reason why the requirements for baseline information may change is that a better *understanding* of the information is actually needed, or a better understanding of ways in which it

can be collected. For example, Proshika discovered that their original way of collecting information on the physical abuse of women was grossly underestimating its incidence. Any baseline using this data would therefore have been a very poor guide. As a result, they adopted different means of exploring this issue, moving away from individual interviews using questionnaires to greater use of participatory and group methods.

Lastly, there may be a different demand or divergent standards for baseline data. In one organisation, where the research team 'pointed out the shortcomings in the reporting and the lack of data in the files, former project staff objected. Most of these staff members claimed that they had the baseline data that we were looking for'. However, when the researchers did come across this information they felt it 'was largely in anecdotal form' and inadequate for impact assessment purposes.

The second difficulty with baselines is that there are major difficulties in not just the collection of relevant information but also in the analysis, storing and recovering of that information at a later date. In some cases, researchers effectively stumbled across crucial information by accident, often by meeting with former project staff who pointed them in the right direction but even in these cases they still had problems.

The researcher met the first anchor person ... and learned that a detailed household survey was conducted ... [b]ut the inferences from the survey were not converted into a report. Some of the survey sheets were found from the Project store. Efforts to trace the remaining sheets from the heap of old files proved futile. (CYSD case study)

Those case studies which had useable baseline data were able to make comparisons that recall or retrospective techniques (designed to help people remember the past situation and compare it to the present day) would struggle to achieve. For example, information collected from BRAC members when they enrolled allowed meaningful comparisons to be made between their status before joining BRAC and after a number of years. In addition it allowed BRAC to compare well-being indicators for members of various durations of membership. They were therefore able to determine that there was 'a significant positive relationship ... between the net change in members' involvement in income-generation activities and increase in membership length' as well as in their control over assets.

One of the weaknesses with recall methods is that people look back at a situation with the benefit of hindsight. As we saw earlier in the section on indicators, the CYSD case study revealed that women looking back at their relationships with men reached different conclusions than they would have done at the start of the project. If changing attitudes, awareness, and

perceptions are important elements of impact that are being assessed, then this should be borne in mind.

In order to overcome some of these problems, a number of the case study programmes such as Wajir and BRAC have identified 'panel groups' or 'cohorts': a group of individuals or households who are tracked over the lifetime of the programme. They provide a 'rolling baseline', which allows changes in people's lives and priorities to be identified on a regular basis, and their views and opinions on project performance to be known. As was noted above in the section on indicators, this sort of approach also makes 'impact tracking' possible, which can help gather better evidence of cause-and-effect relationships between a given project and significant changes in people's lives.

Certain lessons that need to be taken into account for baseline data collection are included in Summary 3.8.

Summary 3.8: Key lessons for baseline data collection

- Aim to collect only those data that is seen to be particularly relevant to assessing the outcome of the project and that will be difficult to identify through recall or baseline reconstruction methods (see below).
- Aim to collect only as much information as the organisation actually has the capacity to analyse, organise, and store. Don't be tempted to match the kind of data and analysis that a larger organisation might collect.
- Recognise that it is impossible to predict all the information that might be needed and that any baseline will need to be updated.
- Explore the possibility of creating 'rolling baselines' by following the progress of a particular number of people or groups throughout the course of the project.
- Investigate the possibility of using new individuals, groups, or communities which become involved in a programme as a baseline for comparison with existing participants.
- Ensure that collected data is properly recorded, filed, and stored. Make sure that the organisation knows where these files are held and what they contain, and that there is an adequate system for retrieving the information when it is needed.

Baseline reconstruction

How can the lack of baseline data be corrected? There are three main sources which proved useful in reconstructing baseline data in the case studies: project documents and records, other organisations, and key informants (whose knowledge was elicited through interviews and participatory recall methods).

Baseline information from project documents and records

As we saw in the earlier section 'What information already exists?' several of the case studies needed to review existing project documents and records in order to prepare a consolidated summary of the project. This included background information and baseline information that not been used in combination before. For instance, the CYSD study found a 'situational analysis' which was prepared during the second phase (1992-93) of the project particularly useful, although by that time the project had already been going for four years. They also found case-study reports of self-help groups and traditional birth attendants which had, for example, specifically explored women's status and health practices before the project intervention. These project documents also proved very useful for reconstructing baseline information, but although they provided important qualitative information, supporting facts and figures were often found to be missing.

Baseline information from other organisations

Some of these supporting facts and figures may be available from other organisations. CYSD used statistical records from a wide variety of official agencies, including village schools, health centres, local government departments such as tax and forestry, which went back to before the project had begun. Although this was time-consuming (it took almost 15 days to reconstruct data on children's school attendance at seven village schools) it did allow key comparisons to be made with the situation prior to project intervention. Similarly, data on cropping patterns and intensity was readily available from agricultural centres.

Although CYSD recognise that the authenticity of government records is sometimes questionable, they are an important source of information base which can be used in conjunction with others. They also provide important insights into how official functionaries and organisations are likely to perceive trends.

It should be remembered that official information may not be available to other organisations (or it may take some time to get hold of it), and decision-makers may not use it to inform their policy and practice. In the Matson case

study in the UK, the team asked official agencies to confirm trends on crime, health, and employment identified by the community by supplying relevant statistics. It became clear that there were problems with the comparability of the data over a period, because of a change in the way that data was collected, and that key staff within those agencies actually had great difficulty getting hold of this information. It also suggested that statutory authorities' perceptions of change were not actually based on an analysis of the quantitative data they collected.

Baseline information from key informants

School teachers, agricultural extension officers, chemists, traditional birth attendants, tax officers, health workers, community activists, former project staff, and the men, women, and children within communities have been invaluable sources of information in the case studies. Simple questions about comparing different years, or the situation before and after a project intervention, have provided very useful quantitative and qualitative information relatively quickly. This information relates to changes in the project environment as well as to changes more directly associated with the project in question. It is striking how many exercises in the case studies could have asked respondents to compare perceptions about current project performance or the current situation with the past but did not. For impact assessment — which is essentially about change — this comparison is critical.

On the other hand, ENDA found that they spent perhaps too much time and effort trying to reconstruct events that took place during the implementation of the project. At the time this was felt to be important if they were to follow the change model they had adopted. In retrospect, they felt that it was unnecessary to search out all of the details of project implementation. Those case studies which started simply by identifying the areas of change seen as significant by different stakeholders, and then exploring these in more depth, seem to have been more successful in reconstructing meaningful baselines.

CORDES developed a very systematic process of reconstructing baselines and then relating these to the current situation and to the evolving context. This method is based on what they term a 'triple diagnostic', asking the communities involved to answer the following questions:

- What was actually done?
- What were the objectives or motives which led us to do this?
- What was the impact on the community and what were the positive (helpful) and negative (constraining) effects of the context?

This line of questioning allowed communities to reflect on whether the original objectives had been met and whether the project, or other external factors, had brought about the change.

How to deal with attribution

One of the critical questions that impact assessment traditionally sets out to answer is what caused any identified change. For most projects this means trying to determine if the changes that have occurred would have happened anyway, and the degree to which an observed change can be attributed to a given project or programme. This is notoriously difficult.

Control groups

One of the ways in which the case studies have dealt with the problem of attribution is through control groups. The control group method requires a comparison between a population that has been targeted by a particular intervention and one that has not. Ideally, this assessment should be done before an intervention occurs and again afterwards, in order to determine if there is any difference between the two populations in question.

During its first impact assessment exercise in 1993, BRAC selected a control group from villages which no programme had yet reached, but which had similar traits in terms of land-holding to villages supported by BRAC. They hoped to be able to compare the well-being of these two groups at a later date during the 1996/7 impact assessment and thus deduce the difference between progress made by BRAC and non-BRAC members. However, BRAC's experience 'shows that getting a true comparison group is a tough problem' (BRAC 1998). First, during the three-year period between the two studies, 39 per cent of the comparison group households had dropped out from the list because they participated in NGOs or for other reasons. Second, even if they were not directly involved in other projects, several comparison households lived in villages where NGOs intervened. They may therefore have experienced a 'spillover' effect from these interventions, especially in the areas of health-care, sanitation provision, and reproductive behaviour, which all have an impact on material well-being. Third, it became clear that although comparison group households were similar in terms of land-holding to those benefiting from the BRAC programme, they differed in terms of age, sex, occupation of the household head, and education levels. This is significant, since the 1996/7 impact assessment found that initial land-holding and occupational status (self or wage employed) of the household head were the two major indicators influencing asset accumulation, apart from BRAC inputs. Last, although some attempts were made to control these factors (by applying

multivariate econometric regression analysis, so as to separate out the different initial endowments of the comparison households and the contribution of external effects) it is recognised that there are major difficulties in quantifying many important socio-cultural aspects which may be critical factors[3].

Table 3.9: Comparison between BRAC and non-BRAC members
Source: BRAC case study, 1997

	BRAC members	**Non-BRAC members**
Poverty		
Below the poverty line	52%	89%
Living in extreme poverty	22%	37%
Assets	Higher for BRAC households both in 1993 and 1996. In 1996, 50% higher net worth (asset plus savings less outstanding loan)	
Rate at which asset value increased 1993-97	22%	84%
Savings	Twice as high as non-BRAC members	
Health		
Households using sanitary latrine	24%	9%
Rate of contraceptive use	40%	27%
Food	Average per-capita calorie consumption and total food and non-food expenditures were significantly higher	
	Higher food stocks, consuming more vegetables, fish and meat	Higher food deficits
	Low seasonal fluctuation in monthly food expenditure (3%)	High seasonal fluctuation in monthly food expenditure (18%)

BRAC therefore advises that the comparison group be treated with caution; on occasion, they have questioned the validity of retaining the comparison group (Mustafa et al. 1996). Table 3.9 outlines some of the data that emerges from the comparison between the control group and BRAC members. Although it clearly points to BRAC members being better off in terms of material poverty, assets, and food security, some points remain unclear because of the dilemmas cited above. One is the starting position of comparison members — were BRAC members better off in the first place? Another is whether changes during this period were associated with BRAC's inputs – indeed, the much faster rate of growth in asset value among non-BRAC members opens up a number of questions.

Non-project respondents

An alternative to establishing a control group at the beginning of the project is to do so at the time of the retrospective assessment by identifying non-beneficiaries and asking both project and non-project groups to recall their situation from a date before the intervention occurred and compare this to their current situation. Both the Wajir and Proshika case studies adopted this approach.

A good example from the Oxfam-Wajir case study in Kenya is illustrated below. In this case, households in project and non-project sites were simply asked to score out of ten their quality of life and ability to withstand drought, ten years ago and now. The average response for each group is shown in Table 3.10 and indicates marked differences between project sites, where people perceive things to be improving, and non-project sites, where people perceive things to be getting worse.

Table 3.10: Perceptions in changes in quality of life over the past ten years by 200 pastoral households in project and non-project sites (on a scale of zero to ten)

	Project		**Non-project**	
	Ten years ago	Now	Ten years ago	Now
Quality of life	4	6	6	4
Capacity to withstand drought	3	7	7	3

Source: Wajir Economic Impact Assessment Report

In the Proshika study this approach produced a number of findings which suggested that among those supported by the programme literacy rates, use of

family planning, participation in local institutions, and economic assets were all higher, and the number of dowry marriages and infant morality rate lower compared to non-beneficiaries. However, there seemed to be little difference in immunisation and divorce rates, and in the use of organic farming methods — despite the project's emphasis on environmental awareness.

This retrospective method of comparing groups in and outside the project, rather than a 'before and after' control-group comparison limits the risk of wasting energy by collecting data that is subsequently considered almost useless because control-group members drop out. In addition, this approach may be more appropriate in areas where there are fewer NGO or governmental interventions and thus less danger of 'spillover' effects from other interventions. Yet this method does not overcome some of the problems associated with the 'before and after' approach such as being able to attribute differences precisely to specific variables, and to find strictly comparable groups. However, there is potential for greater comparison by further disaggregating the groups in question. So in Wajir for example, responses to questions about milk consumption (a critical element of well-being in pastoral areas) were not only explored from the point of view of project and non-project sites but also in terms of wealth and age.

Table 3.11: Average number of times milk is consumed in a day by 200 rich and poor households in project and non-project sites

	Adults		Children	
	Rich	Poor	Rich	Poor
Project sites	0.21	0.15	0.8	0.5
Non-project sites	0.18	0.09	0.7	0.5

Source: Wajir Economic Impact Assessment Report

This breakdown of results allows better comparison of different groupings across project and non-project sites, indicating that the frequency of milk consumption of poor adults in project sites is approximately 65 per cent greater than the frequency of milk consumption of poorer households in non-project sites. However, there is little difference in milk consumed by children in project and non-project sites. Such an approach allows for a greater understanding of who is receiving the greatest benefit from a given intervention.

In another approach, CYSD used a community they had just started working with as the control group. They found that because of this relationship, people were co-operative. In effect, they were using work

carried out with this group in programme identification (through PRA exercises) as a baseline and as a comparator with existing groups and communities. This process allowed questions to be asked about levels of awareness on topics which CYSD had been aiming to change, for example about the use of neem pesticide or the availability of maternity allowances from primary health-care centres. The team also used direct observation to make comparisons; for example, they noted the lack of income-generation production units in the control village compared to project sites.

CORDES adopted a similar approach by comparing three communities which had worked with them for different lengths of time and at different levels of intensity, and measuring the various impacts achieved. Interestingly, CORDES also brought these groups together so that they could compare their own findings and draw their own conclusions.

The methods adopted by CYSD and CORDES overcome many of the ethical problems associated with the conventional control-group approach, although CYSD were concerned that the groups' expectations might bias results. The team feared that because the group was about to become involved with them, they would perhaps exaggerate their poverty and problems in order to be assured of future support. When the community were informed of the idea, they were reported to have questioned how they could be compared to the operational villagers. According to them, the targeted population had had opportunity to learn and act under the continuous guidance and with the assistance of the project functionaries, whereas they had not (CYSD, p.88). Of course, this is precisely the reason why they do in fact provide a useful basis for comparison, although the question with why such groups might want to get involved is an important one.

In fact, this motivation of the respondent is a critical issue for all types of impact assessment. As David Hulme notes

A 'rational actor' confronted by an impact assessor asking standard I[mpact] A[ssessment] questions ... would soon tell the interviewer where to put his/her survey instrument. Fortunately in the world of practice more polite responses are the norm but the issue of how to persuade respondents to spare the time for an interview, and provide accurate and honest answers, is an important one that is rarely mentioned in I[mpact] A[ssessment] methodological statements. (David Hulme 1997, p.19)

So although beneficiaries may put up with an endless stream of questions — because they understand that it is part of the price they have to pay for being involved with an NGO — they may answer questions in a way which they feel is most likely to ensure continued support. However, control groups may have even less incentive to co-operate (as they are by definition liable to

remain excluded) or a greater incentive to exaggerate their needs (in the hope that someone responds). Some suggest that a reward, bribery, or payment of some form should be considered, as is the practice in the UK or USA when market researchers convene focus groups (Hulme 1997).

Lastly, impact assessments also need to take into account what effects a project has had on groups which were intended to benefit but did not. These differ from non-project respondents who are deliberately excluded from a particular intervention, but in effect also function as a control group. For instance, the Ghana study revealed that certain clan groups within the villages studied were systematically excluded from a number of development projects supported by local NGOs. However, one muct be careful in making this sort of comparison, because impact is not neutral: the project may actually have had a negative effect on some people's lives, which makes the comparison with those who have benefited less clear. Moreover, the reasons why certain groups may have been excluded — on the grounds of gender, ethnicity, or age — are precisely what makes them different from those groups that did benefit, which again makes the comparison problematic. This does not undermine the importance of understanding these possible negative impacts or processes of exclusion — quite the reverse — but it does suggest that as far as attribution is concerned, the existence of marginalised groups may not tell one much about what might have happened if the project in question had not occurred.

Using secondary data and other key informants

While it may be considered too difficult, too costly, or unethical to use control groups or non-respondents it may often be possible to use other sources of information about project and non-project areas or groups. As mentioned in the section on reconstructing baseline information, key informants and other organisations can sometimes provide this information.

CYSD for example collected information on land under paddy cultivation, vegetable, pulses, and oil seed cultivation from government agricultural offices, for areas in which they were working as well as for neighbouring areas. This threw light on the relative differences between the area where they had been promoting changes in agricultural practice and those where they had not.

The Ghana study on the other hand deliberately decided not to attempt to prove cause and effect; instead, they sought 'confirmation of attribution from different stakeholders in order to cross-check our analysis'. Data collected from individuals and communities about trends and project impacts was compared with anthropological studies in other communities in the area, a World Bank participatory poverty assessment, and other relevant data. This allowed some comparison between trends which were visible throughout

the region, likely to be attributable to broad economic, environmental, political, and social phenomena, and changes which seemed to occur only in the communities in question, 'bucking' the trends elsewhere.

By looking for other explanations

Another important, connected, means of exploring attribution is deliberately to explore other possible explanations for an observed change or a difference between various populations or communities. For example, Cordes ensured that one key element of their 'triple diagnostic' method was to assess the contribution of external factors in promoting or inhibiting observed changes. In their case study, CYSD discovered many factors, which they called 'interference effects', that revealed why project impacts in one community varied from those in another. These included political interference, drought, crop destruction by wild animals, inter- or intra-village conflict, and donor-driven factors. CYSD undertook an 'external influence study' to understand these issues in more depth. This allowed them not only to compare a project village and a non-project control village, but also to understand why, for example, a project may have had a significant impact on people's lives in one village while it had much less impact in another. This in turn can help in estimating the degree to which differences in a project's context contribute to its success or even failure.

In another instance, BRAC had an indication during their first impact assessment exercise in 1993 that village-level infrastructure, and the distance from metalled roads in particular, played an important role in the economic performance of the communities where the village organisations they supported were located. They recognised that if they did not take this into account in the follow-up study in 1996, they could exaggerate the impact of programmes in villages with good economic infrastructure and possibly underestimate it in villages with poor infrastructure.

BRAC attempted to 'control' this and other factors by developing village profiles based on standard data, and derived from key informants in each village covered by the household survey.

The data include distance of the village from nearest city and metalled road, number of households, existence of socio-economic infrastructure such as haats, bazaars, educational institutions, health centres, NGOs, electricity, and so on, and access to various socio-economic institutions. (BRAC case study)

Although it is difficult to apportion the degree to which project or external factors contribute to a given change, it will add to the credibility of any impact study if the most obvious possible alternative explanations for observed changes are at least referred to and added to the balance of evidence gathered as part of an impact assessment process.

Summary 3.9 describes some of the key lessons learned in dealing with the critical issue of attribution.

Summary 3.9: Key lessons in dealing with attribution

- Apart from the ethical questions related to the use of control groups in a 'before and after' assessment method, there are a number of operational problems which usually make this approach impractical.
- The method which seeks to compare the situation of project beneficiaries and non-project respondents will usually be more appropriate.
- If non-project respondents are groups that the NGO is planning to work with in the future, then not only are some of the ethical dilemmas associated with this approach lessened, but the data gathered can also be used a baseline for that group.
- If non-project respondents are groups that the NGO does not plan to work with in the future, it seems incumbent on the NGO to make the information available to other appropriate development agencies which may be able to provide support and to consider compensating the respondents in some way.

The importance of cross-checking

Impact assessment is in the end a matter of judgement. For that judgement to be as valid as possible, and to be seen to be as valid as possible, the evidence that is gathered needs to be cross-checked. The search for attribution outlined above is one means of cross-checking information. However, cross-checking or 'triangulation' (a nautical term describing a navigation method which uses three points on the horizon to calculate a position) is important for many other purposes in impact assessment. Although some of the elements of impact assessment will involve discovering incontrovertible facts, much has to do with contrasting sometimes contradictory opinions, judgements, and feelings. Some of these apparently contrasting views may in fact be consistent, if seen from a different perspective, whereas some may really conflict.

The case studies used three main methods of cross-checking: by research method, by researcher or assessor, and by respondent or source of information.

By research method

The next chapter looks in detail at the range of tools and methods used in the case studies. Here I will simply note that the case studies drew upon six main, overlapping families of methods: secondary data review, surveys, individual and group interviews, discussions and workshops (sometimes using participatory tools and methods), case studies, and observation. The larger, more complicated studies drew on all or nearly all of these, whereas the smaller studies used maybe only two main methods.

Studies used different sequencing of methods. For example, ENDA undertook a questionnaire survey before holding focus-group discussions; however, they subsequently felt that the opposite might have been better because the questionnaire could have focused on specific issues raised by the discussion groups. BRAC on the other hand conducted a field study before the survey was undertaken in order to find out what impact BRAC field staff and members of village organisation expected the programme to have, and what they considered likely indicators. This gave valuable guidance for the study and also meant that the researchers tested the draft questionnaire several times to make sure that information on the indicators could be collected. The questionnaire was then followed up by further qualitative and case-study work on women's empowerment.

This combination of qualitative and quantitative methods is an important element of cross-checking.

> *Used* before *a quantitative survey, for example, qualitative enquiry may help with formulation and pre-testing of questionnaires, since a good questionnaire requires comprehensive advance knowledge about the system being studied and rigorous preparation. Qualitative enquiry may also generate hypotheses worthy of investigation, or help to narrow down the questions that more detailed surveys should focus on ... Qualitative enquiry may also be useful* after *a formal survey to follow up interesting leads ... and adds depth and context.* (Moris and Copestake 1993)

It is important to stress the importance of direct observation by those conducting interviews, surveys, or participatory research exercises. This was brought out forcefully in the Ghana study when the research team noticed that a number of houses on the edge of a village had not been included in the map drawn by some of the villagers. When this was remarked upon, villagers revealed that these households had been excluded because they were from a minority ethnic group, and therefore not part of the community. It was only through astute observation, the willingness to cross-check those observations with other information, and by probing the contradiction between the two that new information was revealed.

By researcher or assessor

Another way of cross-checking results is by different researchers undertaking the same or similar research or interviews and comparing results. In the case studies, this was done in some instances during village-level PRA exercises: researchers split up, conducted parallel exercises, and then compared results. However, different researchers will always get different results because of who they are, how they behave, how they ask questions, and so on. This is particularly the case with female and male researchers, and 'insider' or 'outsider' researchers, who are liable to differ in their ability to ask certain questions and have them answered.

In the Pakistan case study, Oxfam programme officers, following in-depth discussions with beneficiary groups, graded the projects and local partners according to agreed criteria. These gradings were done independently by different programme officers and then compared. For instance, one of the criteria that Oxfam wished to assess was the 'participation of the community' in a number of projects run by different partners. The project officers agreed the following scoring system for levels of participation.

0: Beneficiary community has no awareness of project partner's activities.

1: Beneficiary community is aware of project partner and has some information or makes some 'contribution'.

2: Project partner consults beneficiaries; some meetings are held; some 'contribution' made.

3: Beneficiary community has some responsibilities at one project stage or there are more regular meetings; some 'contribution' made.

4: There is significant beneficiary involvement at different stages; they have significant responsibilities; some 'contribution' made.

5: The community works in more than one sector; takes initiatives; is able to mobilise resources.

The different graders' assessments were then recorded to make variations in scores clearly visible. There were three graders: A, B, and C (see Table 3.12). Where scorers felt unable to make a judgement, this was indicated by a question mark. For example, the assessors A and C seem to differ in their views on women's participation in project 2; only one scorer has assigned scores to projects 6 and 7; and scorer B was unable to make a judgement on project 3. Thus any independent observer can ask why A and C came to such different opinions or check whether the assessment of a certain project is based on a single person's view.

Table 3.12: Assessment of women and men's participation in seven projects in Pakistan

Project	1		2		3		4		5		6	7
Grader	A	B	A	C	A	B	A	B	A	B	A	A
Women's participation	0	0	3	5	1	?	0	0	2	2.5	3	2
Men's participation	1	?	?	1	1	?	0-3	?	2-3	?	0	2

One of the main tests of rigour in conventional scientific enquiry is that of 'objectivity' and 'confirmability': different observers must be able to agree on a phenomenon, and it must be possible to ensure that research results are not due to the researcher's bias. However, it is increasingly realised both inside and outside the laboratory that the very act of measurement may change the result. It is extremely difficult, if not impossible, for the researcher to remain distant or independent from the process he or she is researching. Therefore, it is important to be clear about the possible bias of individual researchers, and variations between researchers, and to try to diminish these, while recognising that it is impossible to eliminate them entirely. The reliability and credibility of the research is therefore also dependent on the transparency and detail of the record of research methods used. Peers should be able to trace and question the process of research undertaken. For example, the table above from Pakistan indicates how results of qualitative grading can be recorded in a way which makes it clear who was involved in which assessment, when disagreements between assessment occurred and, importantly, when assessors felt unable to make a judgement.

By respondent or source of information

One of the most common forms of cross-checking used in the case studies was by using different informants or sources of information. The first level of cross-checking was within communities, and the second between communities and other informants and stakeholders.

Cross-checking by respondent within communities

At a very basic level, this cross-checking simply involves, as in the Cordes case study, verifying if what had been proposed actually happened and whether it happened in the way that was intended. This validation formed an important step

in their method of assessment. In other studies, the use of well-being ranking and gender analysis pointed to the importance of understanding differences between various groups within communities. Wealth, gender, age, ethnicity, and religion are often important elements in understanding these differences.

In the Ghana study, many women stressed the importance to them of income that they earned from selling produce grown on their own small plots of land. Men however, when asked about household income and the division of labour in the household, did not mention this activity or source of income at all. If the interviews had only involved men (which is not unusual in some surveys) this information would probably not have been revealed at all.

Not only can new information be revealed by finding additional respondents, but differences in priorities can also emerge through cross-checking. Table 3.13 reveals the preferences of different groups in a community in Andhra Pradesh, as discovered by a recent impact assessment undertaken by Action Aid.

Table 3.13: Preferences for change in Andhra Pradesh, India

Priority	Adult men	Adult women	Male youth	All
1	Housing	Land, irrigation, and crops	Land, irrigation, and crops	Housing
2	Land, irrigation, and crops	Housing	Education	Cattle
3	Cattle	Savings and loan	Health	Land, irrigation, and crops
4	Education	Cattle	Wage employment	Savings and loans
5	Savings and loans	Education	Savings and loans	Education
6	Clothing	Clothing	Cattle	Health
7	Leadership		Plantations	Clothing
8			Housing	
9			Self-employment	

Source: Action Aid impact assessment research

This research reveals important differences in priorities between different groups, but it also raises several questions about how group priorities are

determined. Although it is unclear how this happens, the column for 'All' suggests that male priorities tend to win out (housing emerges as first priority, although women rank it second and young men eighth). It also suggests that negotiation does change priorities (cattle emerges as the second priority although none of the groups ranked it higher than third) and that women and young men together still have difficulty in getting their priorities recognised (despite both groups ranking irrigated crops as first priority, it slips to third place in the group listing). Young men possibly have most difficulty in getting their priorities recognised: wage employment does not figure in the group's list.

Groups within communities not only are likely to have differing priorities and views, these may in fact conflict. For example, the CYSD study revealed that women in several communities had initially been discouraged from saving in a group account or joining self-help groups, because their husbands perceived this as a threat to their domestic authority. Those who did join faced hostility from their husbands for extended periods. This hostility has been overcome in some cases, as economic benefits have increased and project staff have deliberately targeted 'awkward' husbands. This is a good example of how cross-checking views and attitudes within communities not only has to do with building consensus or agreeing priorities but can also be about understanding threats or obstacles to achieving impact which need to be taken into account and, if possible, countered.

However, it must be recognised that there are limits to the amount of time that people are willing and able to spend on cross-checking; there may even be a law of diminishing returns. In other words, it is crucial to prioritise what information must be cross-checked, which sources need to be involved, and to set realistic time-frames for doing so.

Cross-checking by comparing information from communities and other informants

Through visits to schools and discussions with teachers, CYSD discovered that although school records had indicated that enrolment was improving, the average attendance level was only 50 per cent of the enrolled students. These students were engaged in household duties such as looking after younger children when parents are working or rearing livestock. Taking the school records at face value, without cross-checking the findings with teachers, could have led to a very misleading impression about what was actually happening. It could also have led to an inaccurate assessment of the quality and impact of schooling if, for example, literacy levels had been calculated according to the numbers of enrolled children.

Tax officials, revenue departments, and forestry officers also proved useful key informants, not least in indicating whether official data and records

could be trusted. In addition, some of these informants had been working in the project areas for much longer than the NGOs in question. Revenue officials in Orissa, who had been working in the area for the past 15 years, were able to describe changing trends in villages which they attributed to CYSD intervention, as well as providing vital information about the proportion of land in the project area owned by outsiders. This helped CYSD understand the importance of distress land sales by tribal communities, and of the encroachment on tribal lands.

Health workers can be key informants with important insight into local communities, not just related to health projects. For example, discussion with residents in Matson, UK, revealed that breathlessness was one of the most commonly cited health problems. A local chemist confirmed that prescriptions for asthma were very common, which in turn led to suspicions that damp in unrefurbished households was one of the possible causes of asthma. In this case, the key informant not only confirmed what people had said, but was also able to give a more precise definition of the problem and point to its possible causes. In terms of the impact assessment, this information helped to understand the potential secondary impacts on people with refurbished houses and those without.

Dealing with contradictions

Nearly all the case studies report the difficulties and challenges of dealing with contradictory information that emerges. In some cases, these contradictions could be explained quite easily. In others, they provided new insights that led to further lines of questioning. As noted in the CYSD study it is important to

study the logic behind such contrast and how best this can be interpreted. If it is not possible for different people to agree on the interpretation of certain findings it may be better to record the different views, rather than attempt to reach a consensus. (CYSD, p.30)

Cross-checking is an important means of exposing bias, specific interests, and power relations. Deliberately seeking out the views of a wide variety of people with different views is an important tool in assessing impact. As Moris and Copestake note,

[this] approach to data is much like that of investigative journalism. Assume bias is present, but allow for it by recognising explicitly a respondents' interests and by cross-checking statements with neutral observers or those holding opposite views ... for example ... [q]uestioning both buyers and sellers, wholesalers and retailers, about the same transaction (e.g. interest rates on

trade credit) often yields an upper and lower figure. By asking who is lying and why, important insights into relative market power may be gained. (Moris and Copestake, pp.48-49)

Summary 3.10 lists some of the main lessons about how to cross-check findings and some of the key issues that emerged from the case studies.

Summary 3.10: Key lessons about cross-checking findings

- Mixing qualitative and quantitative methods is a vital element of cross-checking. You need to consider what sequence of methods makes most sense.
- Direct observation is an important means of checking if there are discrepancies between what people say (or do not say) and their actions.
- Using different 'assessors' or stakeholders to review the same issue can reveal areas of consensus and difference. Agreeing at the outset which criteria to explore, or using a single reporting format can simplify comparison at a later date.
- Making the methods by which information was collected clear and transparent allows third parties or peers to cross-check the process of enquiry.
- Differences in power and status within communities — especially relating to gender, wealth, age, and ethnicity – make cross-checking between groups even more important. Using well-being ranking and gender analysis can be an important part of this.
- Key informants who know a particular region or group of people very well — health workers, local officials, teachers, and so on — can be an excellent source of information and can also verify information from other sources.

Summing up: the problem of aggregation and synthesis

This chapter has explored how the case studies went about designing their impact assessment processes, what preparatory steps were taken, how they coped with baseline data and indicators (or the lack of them), as well as how they tried to deal with the difficulties of attributing change to a given intervention. I have also described how the studies approached the need to cross-check findings and sometimes to record differences in views and

opinions rather than attempting always to reach a consensus. This leads me to one of the trickiest problems that impact assessment processes have to take into account: how to summarise and present findings in a useful and relevant form, or forms, which do not lose the richness, diversity, and complexity of the story. It is important to address the question of synthesis at the design stage of an impact assessment, so that data collection and analysis can take its form into account.

Feedback

One of the ways to ensure that any synthesis of findings does not miss or misrepresent important views is to cross-check them with different stakeholders, through field presentations or workshops, and by circulating draft findings and reports for comment. This has happened or is planned in nearly all the case studies and has involved the following forms of sharing findings.

Community feedback: Findings can be communicated through verbal presentations and workshops, although it is unclear whether any of the case-study participants will be translating reports or findings into appropriate local languages.

Presentations and workshops with staff: This approach allows both the research team and local staff to review findings and learn more about methods of impact assessment.

Multi-stakeholder workshops: These can involve beneficiaries, staff, donors, other NGOs, bi- and multilateral agencies, government, and academics.

All these methods have been important in cross-checking findings, generating new insights, communicating results, and in helping different stakeholders understand differing opinions and views. Perhaps the most striking example of the importance of eliciting feedback comes from the Proshika case study. After conducting a large household survey, the results were separately communicated to programme partners from village organisations, middle managers, and senior managers. First, each group was asked to present what they expected the results to be against the key indicators of the study; then the actual results were presented and the divergence analysed. As a result, the findings for five of the 13 indicators were seriously questioned, which led to the issues being further explored. Follow-up qualitative and participatory research not only confirmed the major differences which the feedback had identified, but also revealed why the questionnaire process had produced the results in question.

However, workshops, in particular those involving many stakeholders, can also spark off conflict which it may, or may not, be wise to provoke. For example, the meeting of a wide range of stakeholders in Ikafe at the end of the field work resulted in heated debate. During the review, the disquiet of some of the representatives of the host population lead to threats of violence against Oxfam staff; indeed, some bullets were fired at the Oxfam compound just after the review. Although most people involved believe that this was the result of the 'normal' insecurity in the area it indicates that in situations where high levels of tension exist, bringing all parties together may be risky.

The report

Most impact assessments produce a report (or reports). Its structure can help summarise information in ways that do not sacrifice too much and which can also facilitate feedback. First, it is important than any intervention is properly contextualised. As I noted in Chapter 2, impact assessment essentially analyses how a given set of actions combines with the context to produce change. The report should therefore explain how the political, social, economic, and environmental climate influences the outcomes of a given project and how people perceive its benefits. Second, the report should distinguish between data, analysis, and recommendations, and clarify whose analysis or recommendations they are. This allows readers to agree with the data or analysis while contesting the recommendations or vice-versa. Third, as explained in the previous section, it is important that the report retains and presents any important differences in opinion or analysis from different stakeholders or even within the assessment team. Fourth, the report should present disaggregated data for relevant groups, even if this shows no differences between them. Disaggregation by gender, age, and well-being — although it was not done in most of the case studies — should almost certainly be done in every case. Other dimensions of difference such as ethnicity, clan, religion, or sexual orientation may also be important in some contexts.

Last, the use of case studies and people's own words can not only breathe life into the analysis, but also exemplify more forcefully what change means to individuals and communities, and what it feels like to experience it from day to day. It should be remembered that this may need to be interpreted for those who not familiar with the context.

Other means of communication

Although none of the case studies used video, radio, cassettes, or comic books to communicate the results of the research or to complement the main report, these methods are increasingly being used in the development field (Braden 1997).

Recent Oxfam experience in the UK and East Africa suggests that video in particular has real potential not only to enrich dry reports, but also for individuals and groups to represent themselves and the changes they have gone through, as well as communicating their feelings to others. With adequate and relatively cheap training, video has the added advantage of being a tool that communities can use to track and record change over time in a quite simple and cost-effective manner. It should be noted that synthesising several hours of video tapes is certainly no less, and possibly a little more, difficult than several hundred pages of data. However, groups and communities which synthesise and edit the material themselves, have added control over how they would like to portray themselves and their achievements.

In large organisations, the problems of aggregation are even more acute. I explore this in more depth in Chapter 7.