Subgroup 2

NONIE IMPACT EVALUATION GUIDANCE

09 January 2008

Foreword

NONIE Impact Evaluation Guidance

A note on the contents of the Subgroup 2 submission

NONIE SG2 (SG2) proposes that the NONIE guidance document include five sections as outlined in the table below. To facilitate reading the SG 2 document it highlights in bold the sections prepared by SG2

Sections of NONIE Guidance	Comments
Section 1: Introduction (to the broad impact evaluation perspective of NONIE)	SG1 has provided an introductory document. SG2 submits that this has a narrow perspective and does not address the comprehensive view of impact evaluation that is espoused by NONIE members. SG2 thus presents an alternate proposal for Section 1 of the NONIE Guidance Document (See Section 1 of the SG2 document). This presents a multi-faceted and contextual character of impact evaluation in development contexts
Section 2: Experimental and Quasi Experimental Designs (with a focus on attribution analysis)	This section is produced by SG1. This is not included in this document.
Section 3: Contribution analysis as an alternative method of causal analysis	This section is included in this SG2 document and presents alternatives to causal attribution methods based on a counterfactual. It provides one example in the current draft. Other cases from development impact evaluation are under preparation. A considerable number of references are under preparation to be later included in the document and in an accompanying website.
Section 4: Other approaches for impact evaluation	This section is included in the SG2 document. A considerable number of references are under preparation to be later included in the document and in an accompanying website. An Annex in the document presents a number of cases that have used some of the approaches presented. More work is needed on these cases as well as on additional ones under preparation.
Section 5: Impact evaluation of new aid instruments and country programs	This is work produced by SG3 This is not included in this SG2 submission.

The key message that provide the framework for the SG2 documents are as follows.

1. **The guidance provided needs to be suitable for the full range of development interventions, (projects, programs and policies), from discrete, homogenous and tightly specified interventions to complex, comprehensive, heterogeneous and emergent interventions.** This will involve drawing on a range of methods and experience from different disciplinary sources.
2. **Methods, techniques and approaches for impact evaluation should match the specific circumstances of the evaluation – its purpose, the nature of the intervention, the questions, the level of existing knowledge, and the resources available.** Methodological appropriateness should be considered the 'gold standard' for impact evaluation.
3. **Rigorous impact evaluation is not just about causal analysis. Impact evaluation involves four tasks: identifying impacts that are valued; gathering evidence of these impacts; causal analysis; and managing the evaluation. Rigorous impact evaluation requires each of these to be done appropriately and effectively.**
4. **Depending on the nature of the intervention, appropriate causal analysis may be causal attribution (involving an explicit counterfactual) or causal contribution (involving explicit attention to the causal packages producing the impacts)**
5. **Causal analysis requires systematic, iterative analytical strategies are needed to assess the causal contribution of interventions to impacts.** Research designs cannot by themselves adequately address all threats to internal validity, which should be systematically identified and investigated, and designs with higher internal validity can have low external validity, or generalizability.
6. **In the NONIE context, impact evaluation has to serve *development.*** The criteria for evaluating impact evaluations and methodologies should be their contribution to improving development.

The documents submitted by SG2 are working drafts which will be refined following review at the NONIE January 2208 meeting and further review by the SG2 Reference Group and by the various agencies of the SG2 members in ensuing mutual alignment with philosophies and directions for impact evaluation.

To enhance guidance and follow-up to understandings to develop from this guidance, SG2 is also preparing a website which will include the references compiled for this work as well as links to other how-do documents and other websites.

Acknowledgments

Subgroup 2 is grateful to its reference group of international evaluators for their commitment to the cause of NONIE and for their very active contributions to the content of this document.

Please see Annex 2 for bios of the Reference Group

Marie-Hélene Adrien:	President and Senior Consultant, Universalia
Paul Balogun	Consultant , Author
Michael Bamberger:	Consultant, Author
Fred Carden:	Director of Evaluation Unit, IDRC Canada
Stewart Donaldson:	Professor and Chair of Psychology, Director of the Institute of Organizational and Program Evaluation Research, and Dean of the School of Behavioural and Organizational Sciences, Claremont Graduate University
Osvaldo Feinstein:	Consultant. Author, Editor
Ted Freeman:	Consultant and Partner Gross Gilroy, Inc.
Sulley Gariba:	Consultant, Executive Director, Institute of Policy Alternatives
Jennifer Greene:	Professor, Educational Psychology, University of Illinois at Urbana-Champaign
Ernie House:	Emeritus Professor, School f Education, University of Colorado
Mel Mark:	Professor of Psychology, Penn State University
John Mayne:	Consultant, Author, Adviser on public sector performance
Masafumi Nagao:	Research Professor, Center for the Study of International Cooperation in Education, Hiroshima University
Michael Quinn Patton:	Consultant. Author, Former President of AEA,
Ray Pawson:	Professor of Social Research Methodology, School of Sociology and Social Policy, University of Leeds
Bob Picciotto:	Visiting Professor, Kings College, London
Patricia Rogers:	Professor in Public Sector Evaluation, CIRCLE (Collaboration for Interdisciplinary Research, Consulting and Learning in Evaluation), Royal Melbourne Institute of Technology
Tom Schwandt:	University Distinguished Teacher/Scholar and Professor of Education, University of Illinois at Urbana-Champaign
Nicoletta Stame:	Professor, University of Rome “La Sapienza”
Bob Williams:	Consultant, Author, member of the Editorial Boards of the American Journal of Evaluation and New Directions in Evaluation.

In particular we wish to thank Patricia Rogers from RMIT, Australia, for her significant assistance with the structuring and writing of major parts of the content; those who contributed to the text and in particular Michael Bamberger, Fred Carden, Osvaldo Feinstein, Jennifer Greene, John Mayne, Ray Pawson, Michael Patton, Bob Picciotto, Nicoletta Stame and Bob Williams - as well as those who committed their time to the case studies (in Section 4) Jocelyn Delarue, Fabrizio Felloni, Divya Nair, Christine Obester, Lee Risby, Patricia Rogers, David Todd and Rob van den Berg.

We also thank Terence Beney from Feedback Research & Analytics, South Africa, for his assistance with some of the research and Tega Shivute, Anish Pradhan, Michelle Sy, and Kutisha Ebron from UNDP for their assistance with research and with the SG2 website resources (wwwNONIE SG2…)

*Subgroup 2**

Tullia Aiazzi (FAO Evaluation Office)

Lucien Back** (UN ECOSOC)

Jocelyn Delarue (AFD)

Fabrizio Felloni (IFAD Evaluation Office)

Divya Nair (GEF Evaluation Office)

Christine Obester (UN OIOS, Inspection and Evaluation Unit)

Zenda Ofir (Co-Coordinator) (Evalnet South Africa and AfrEA)

Sukai Prom-Jackson (Co-Coordinator) (UNDP Evaluation Office)

Lee Alexander Risby (GEF Evaluation Office)

David Todd (GEF Evaluation Office)

* Reviews of the draft document are ongoing in some agencies represented in Sub-group 2.
* * Not associated with the draft document at this stage but will re-join the group in 2008.

Subgroup 2

NONIE IMPACT EVALUATION GUIDANCE

Section 1: Introduction

1. Purpose and Structure of Section 1

The Network of Network on Impact Evaluation (NONIE) seeks to provide guidance for effective, rigorous impact evaluation. Three sub-groups have focused on particular approaches: experimental and quasi-experimental designs; other approaches to rigorous impact evaluation; and approaches suitable for specific types of aid assistance instruments such as Sector-wide Approaches and General Budget Support.

This is the first of three sections produced by sub-group 2. This section (Section 1) provides an overview of impact evaluation – its definition and the context for impact evaluation in development. It sets out four different tasks involved in impact evaluation:

- Identifying impacts that are valued;
- Gathering evidence of these impacts
- Assessing the contribution of the intervention to these impacts
- Managing the impact evaluation

It provides an overview of methods and approaches that are useful in undertaking these different tasks and through this improving the rigour of impact evaluation, and provides some illustrations from evaluations that have used these methods and techniques.

The two other documents of SG3 focus on approaches and methods. Section 3 (which would follow Section 2 prepared by SG1) focuses on causal analysis that may be used when experimental and quasi-experimental designs are inappropriate or not feasible. It provides guidance on when these approaches are suitable and describes how they should be implemented. It presents a case example. Other cases being developed will be included in the future. Section 4 presents a wide range of approaches that are being used for impact evaluation. It is accompanied by a set of cases that illustrate use of the various approaches or methods. These are being refined as part of the continued work of the SG2. NONIE members are being solicited to provide cases that could be used to effectively illustrate the various approaches and methods.

2. Definitions

NONIE uses the DAC definition of impact[1]:

> *…the positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended. These effects can be economic, socio-cultural, institutional, environmental, technological or of other types.*

This definition broadens impact evaluation beyond either simply measuring whether objectives have been achieved or assessing direct effects on intended beneficiaries. It includes the full range of impacts at all levels of the results chain, including ripple effects on families, households and communities, on institutional, technical or social systems, and on the environment. In terms of a simple logic model, there can be multiple intermediate (short and medium term) outcomes over time that eventually lead to impact – some or all of which may be included in an evaluation of impact at a specific moment in time.

This definition emphasises the need for an understanding of the consequences of development interventions in the longer term. The effects that are evident in the short-term may continue to be evident in the long-term, they may increase or lead to other impacts, or they may decrease or even vanish over time.

[1] OECD DAC Glossary of Key Terms in Evaluation and Results-based Management

This definition has some significant implications for impact evaluation:

- Given the range of potential impacts from an intervention, processes are needed for identifying possibly important impacts (including unintended and negative), and negotiating which ones will be addressed in an evaluation.
- Impact evaluation needs to assess the value of the results derived from an intervention. This is not only an empirical question but inherently about values – which impacts are judged as significant (whether positive or negative), what types of processes are valued in themselves (either positive or negative), and what and whose values are used to judge the distribution of costs and benefits of interventions.
- Impact evaluation needs to gather evidence of (or assess the likelihood of) success in, and/or potential for sustaining positive action or effects in the long term. Issues of sustainability should not be neglected in spite of the significant difficulties in its credible evaluation.

3. Impact evaluation in the context of development: challenges and key considerations

Impact evaluation approaches need to remain cognisant of, and responsive to, critical shifts in development and its evaluation, particularly the increased focus on:

1. Non-standardised interventions with multiple and/or emergent components
2. Evidence-based policy and practice
3. Partner country participation in, partnerships and ownership of evaluations
4. Multi-disciplinary approaches to development and to evaluation.

3.1. Non-standardised interventions

The endorsement in 2000 of the Millennium Development Goals by all heads of state, together with other defining events[2] and occurrences has propelled new action which challenges development evaluation to enter new arenas. There is an ongoing shift away from fragmented, top-down and asymmetrical approaches. Increasingly, ideals such as 'harmonisation', 'partnership', 'participation', 'ownership' and 'empowerment' are being emphasized by stakeholders.

Development efforts are also shifting from project to country, regional and global levels; poor countries are contemplating how to build capacities to lead and drive country evaluations; and there are increasing calls for more scrutiny of the consequences of globalisation and rich countries' policies, in particular those that shape the physical world as well as the enabling global and national environments in areas such as private investment, trade and security policies, migration and intellectual property rights (Picciotto, 2007).

Development evaluation has increasingly begun to deal with the challenges of evaluating those complex, comprehensive approaches imperative for development. There is today more than ever a 'continuum' of interventions (Iverson, 2003):

- At one end of the continuum are relatively simple projects characterised by single 'strand' initiatives with explicit objectives, carried out within a relatively short timeframe, often dealing with non-human, non-social environments where interventions can be isolated, manipulated and

[2] Such as the Monterrey Consensus (2002), Rome Declaration on Harmonisation (2003) and the Paris Declaration on Aid Effectiveness (2005)

measured. For example, impact evaluation in the agricultural sector might be able to attribute change in crop yield after introduction of an intervention such as a new technology, agricultural practice or fertiliser; in the health sector, the introduction of bed nets against malaria mosquitoes. For these types of interventions, experimental and quasi-experimental designs may be appropriate for assessing causal contribution, along with attention to the other tasks of impact evaluation

- At the other end of the continuum are comprehensive programmes with an extensive range and scope (increasingly at country, regional or global level), with a variety of activities that cut across sectors, themes and geographic areas, and emergent specific activities. Many of these programmes address aspects proven to be critical for effective development yet difficult to define and measure, such as human security, good governance, political will and capacity, sustainability, and effective institutional systems.

3.2. Emphasis on results-based management and evidence-based policy and practice, including evidence from previous evaluations and research

Impact evaluation has received an enormous amount of attention over the past three years. This trend has been driven by several factors such as the focus on managing and being accountable for results, propagation of the concept of 'evidence-based policy' and an increasingly urgent search for 'what works' - especially given the uneven performance of development initiatives across the world and diminishing global resources amidst competing national and global priorities. Many development interventions appear to leave no trace of sustained positive change after they have been terminated and it is hard to determine the extent to which interventions are making a difference to the situation of the poor.

Greater emphasis on impact evaluation for evidence-based policy-making can create greater risk of manipulation aimed at producing desirable results (positive or negative). Impact evaluations require an honest search for the truth and thus place high demands on the integrity of those commissioning and conducting them. For the sake of honest commitment to development, evaluators and evaluation units need to ensure that impact evaluations are designed and executed in a manner that limits manipulation of processes or results towards any ideological or political agenda. They should also ensure that methodologies are not selected or results recorded in a manner that favours specific types of information above others (for example numbers and statistics above equally useful description) without good reason.

The emphasis on evidence-based policy has also brought a renewed focus on the aggregation of impact evaluation results from previous evaluations and research. Various methods have been developed including meta-analysis, systematic reviews (Shadish 2005) realist synthesis (Pawson, 2007) and the systematic review of studies using mixed methods (Sandelowski et al, 2006). Meta-analysis only synthesises evidence from studies which have generated an effect size (using experimental or quasi-experimental designs) and excludes other evidence; the other approaches include evidence from a range of designs and approaches.

Using these methods, comparable interventions evaluated and analysed across countries and regions, can provide the empirical basis to identify 'robust' performance goals and to help assess the relative effectiveness of alternative intervention designs under different country contexts and settings. These methods can lead to increased emphasis on the rigour of impact evaluations so they can contribute to future knowledge-building as well as meet the information needs of their immediate stakeholders. These methods can also lead to a more selective approach to extensive impact evaluation, where existing knowledge is more systematically reviewed before undertaking a local impact evaluation. In brief, unless an impact evaluation is essential to generate new and useful knowledge that cannot be obtained in another way, it should not be undertaken.

3.3. Partner country participation in, partnerships and ownership of evaluations

Developing countries are also more active and vocal about what works in both development and evaluation in their varying contexts. The demands for national ownership and country-led evaluations, and the use of paradigms and instruments that are valid in representing local knowledge present a most welcomed challenge to development evaluation. Greater understanding of what constitutes reality and valid knowledge among different cultures is an important consideration for enhancing the validity and credibility of impact evaluation. More is needed in this area in working with development partners as we address issues of ownership of evaluation.

3.4. Multi-disciplinary approaches to evaluation

Considerable work has been done to highlight and further develop suitable methods for impact evaluation, yet there is clearly great need for respect and collaboration across disciplines and sub-disciplines in order to innovate and test diverse approaches.

The value of cross-disciplinary approaches to both development and evaluation is now more and more being recognised and promoted (Iverson 2003). For example, a recent cross-disciplinary study shows that current development thinking still makes use of too narrow a range of possible approaches to change (Krznarik, 2007).

Strong pleas have also recently been made for development evaluators to recognise and make full use of the wide spectrum of frameworks and methodologies that have emerged from many different disciplines and provide evaluation with a rich arsenal of possibilities (Kanbur 2002; White 2002; Bamberger and White 2007). For example, in their impact evaluation work evaluators can benefit from approaches developed in different disciplines and sub-disciplines. Among others, neo-institutionalist economists have shown ways to study the impact of institutions as ‘rules of the game’, and interventions such as policies can be considered as attempts to establish specific rules of the game with the expectation (through a ‘theory of change’) of generating certain impacts (Picciotto and Wiesner, 1997).

4. What is involved in an impact evaluation?

4.1. Four components of impact evaluation

Across all different types of impact evaluation there are four common tasks, all of which must be undertaken adequately for rigorous impact evaluation:

1. **Identifying impacts that are valued** – identifying and prioritising impacts that will be included in the evaluation, including intended and unintended, positive and negative, short-term and longer-term, economic, social, environmental for individuals, families, households, communities and organisations.
2. **Gathering evidence of impacts** – retrieving existing data, collecting and creating new data, and addressing challenges in the adequacy and feasibility of measures and indicators, particularly for multi-dimensional and longer-term impacts.
3. **Assessing causal attribution or contribution** – understanding whether the intervention is necessary and sufficient to bring about the impacts of interest, whether it is only successful in particular favourable implementation environments or in conjunction with other interventions, or whether the intervention is one of several paths by which the impacts can be achieved.
4. **Managing the impact evaluation** (whether conducted as an internal or external evaluation) – identifying the intended users, assessing the likely utility of the evaluation and what level of resourcing is warranted, negotiating focus and methods, including what will be seen as credible evidence, and developing reporting methods which meet the needs of different intended users, including the appropriate balance between simple and comprehensive messages about findings and implications.

This implies comprehensive identification of important impacts; systematic and defensible data collection and analysis of evidence of these impacts; sound inferences about the contribution of the intervention to achieving these impacts; and effective management of the evaluation, including transparent reporting of methodology and, where appropriate, formal meta-evaluation. Specific approaches and methods can further improve rigour in each task, as illustrated in the following sections.

4.2. Different types of impact evaluation

While impact evaluation always seeks to understand the impact of an intervention, different situations require different types of impact evaluation in terms of design, method, scale of resourcing and processes for managing the evaluation.

When deciding methodology for impact evaluation, attention needs to be given to

- What is the purpose of the impact evaluation?
- What is the nature of the intervention?
- What level of certainty is required?
- What degree of sensitivity to the effect of different implementation contexts is required?
- What resources are available, including timeframes?
- Who is the audience and / or intended user(s) of the impact evaluation, and what will they consider credible evidence that impacts have been achieved and that the intervention contributed to them?

Given the variety of development scenarios, any one particular impact evaluation approach cannot automatically be the 'preferred' choice. The most appropriate approach and methods for impact evaluation have to be determined by the purpose and intended use of the evaluation, the type of intervention (the implementation environment, level of complexity and the problem to be solved) and the available resources.

As illustration, four scenarios in development impact evaluation are sketched below. They are not the only possible scenarios; others may be developed. They are also not mutually exclusive and may be based on a mixture of some of the four. However, these examples serve to illustrate the need for a variety of approaches and methods to impact evaluation:

Scenario 1: Scaling up a discrete, simple intervention – Generalisibility-driven impact evaluation

In this scenario, impact evaluation is intended to generate knowledge with limited errors to inform decisions about expanding the intervention (going to scale) or recommending the model to others (generalising the model). It focuses on a clearly defined, simple intervention which has previously undergone cycles of formative evaluation and process improvement, where there is a likelihood of a plausible connection between the intervention and the impacts of interest, and some preliminary evidence of impact. Evidence of the impacts can be observed within the timeframe of the evaluation and there are accepted measures and indicators available as well as an explicit and credible counterfactual of what happens in the absence of the intervention. The nature of the intervention is such that it is likely to work in other cultural, economic, and political contexts (for example immunisation against disease).

This situation calls for '*generalisability-driven impact evaluation'* which will have high credibility among a broad audience, including policy makers in other organisations and researchers. The evaluation will need to include attention to the adequacy of implementation in order to distinguish between theory failure and implementation failure, and should attempt to identify if there are contexts which are critically important in achieving the impacts of interest.

In this scenario, experimental or quasi-experimental designs with an explicit counterfactual may be appropriate. A theory-of-change may be developed before the evaluation to identify short-term impacts that can be observed during the life of the evaluation, to identify potential negative impacts that should be included in the overall assessment. Mixed methods may be used to develop more comprehensive evidence of impact and for complementary process evaluation to be able to distinguish between implementation failure and theory failure in the event of lack of intended impacts. In some cases an assessment of the fidelity of implementation of the intervention is important to ensure the stability of the intervention and to avoid errors in decision making.

Scenario 2: Tracking emergent, complex interventions: Developmentally-driven impact evaluation

In this scenario, impact evaluation is intended to inform the ongoing development and potential subsequent scaling up of a complex, emergent intervention. While the initiative aspires to have significant impact in the long term, there are many components to the intervention and thus many uncertainties in the underlying theory of change; many of the details of what to do may have to be worked out along the way; changes in the original design can be expected once the work has begun; and there are substantial risks and problems to be managed as well as opportunities to be seized. If the overall idea and vision works, the impact could be substantial.

This situation calls for '*developmentally-driven impact evaluation'*. The evaluation will need to have processes in place to track how things are developing - including preliminary (short-term) and intermediate (medium-term) outcomes - and the likelihood that these will lead to longer term impacts, for example through appropriate sustainability strategies. It will need to take into account how the cultural, social, and political context affects what emerges, and how changes in context, challenges, opportunities and the environment change what is done. Those involved want to keep the effort targeted on having significant

impact, and document and assess those impacts seriously and rigorously in order to develop an understanding of how and under what circumstances such impacts are obtained, and positive impacts sustained.

In this scenario, experimental designs are not appropriate as the intervention is not discrete and standardised but emergent. A theory of change can be useful not as a fixed blueprint for the intervention but as an evolving description which gradually fills in the details of specific impacts and causal paths to achieve these, and which shows the possible influence of other factors and organisations which may need to be addressed for effective implementation of the intervention. Mixed methods will provide timely quantitative and qualitative data about progress to inform ongoing decisions and actions.

Scenario 3: Ensuring accountability for results: Accountability-driven impact evaluation

In this scenario, impact evaluation is intended as part of the accountability system, in addition to accountability in terms of proper use of resources and appropriate processes. Significant funds have been allocated to an intervention and the evaluation needs to show whether resources have been well spent to achieve intended results and minimise the risk of not sustaining benefits in the long term. Because of the cultural, political, social and/or economic conditions involved, the blueprint of this intervention is unlikely to be a model for others (at least not in its details).

This situation calls for *'accountability-driven impact evaluation'*. The evaluation will focus on whether significant positive impacts or results have been achieved and significant negative impacts avoided. Where positive impacts have not been achieved, the impact evaluation should identify where the expected causal chain has broken and whether the intervention could or should have identified and overcome this obstacle.

In this scenario, a theory of change approach will help to identify potential breaks in the intended causal chain and focus data collection. Systematic analysis of the value added by the intervention will be needed. Given the type of intervention, it is unlikely that a control group or comparison group will be available, and simply using a 'with' and 'without' project comparison to estimate changes due to the intervention will not be adequate. Instead more systematic investigation of the causal contribution of the intervention will be needed for a credible impact evaluation and any associated cost-benefit analysis (economic rate of return). Given the high accountability stakes, considerable attention needs to be given to managing the impact evaluation to ensure the integrity of the data and reporting.

Scenario 4: Learning from comprehensive collaborative interventions: Contribution-driven impact evaluation

In this scenario, impact evaluation is intended to inform multiple organisations that are involved in a complex multi-funder, collaborative, multi-dimensional, and comprehensive intervention with several different projects each working on some part of a larger initiative. The evaluation needs to provide an understanding of whether and how these many parts (different projects / efforts / sites / collaborators) add up to some overall impacts, and in so doing, learn which parts have worked well and which have not worked so well, and why.

This situation calls for *'contribution-driven impact evaluation'*. Instead of undertaking the (most likely impossible) task of determining direct linkages between cause and effect, it will determine those conditions that are necessary, but may not be sufficient for change to occur, and those development changes or impacts to which the intervention has most likely contributed.

In this scenario, considerable attention needs to be paid to the different information needs of the different stakeholders and their potentially different perspectives on what are considered important impacts and credible evidence of them. Again, given the type of intervention, it is unlikely that a control group or comparison group will be available, and simply using a 'with' and 'without' project comparison to estimate

changes due to the intervention will not be adequate. Instead more systematic investigation of the causal contribution of the intervention will be needed.

5. Improving the rigour of impact evaluation

The following section discusses strategies for improving the rigour of impact evaluation in terms of the four components or tasks noted in section 4.1.

Across these different components, three approaches are commonly advocated for improving the quality of impact evaluation:

1. **Theory-based evaluation** – developing an explicit statement about the causal path (or possible causal paths) linking the intervention with intended or observed impacts and then using this to guide the evaluation; can be done before the evaluation, or before the intervention, or retrospectively, or revised iteratively during the intervention or the evaluation
2. **Mixed methods** – using a mix of quantitative and qualitative data sources, types of data, sampling methods, analysis methods in order to balance the limitations of any one method
3. **Participatory approaches** – involving local participants, including intended beneficiaries in the evaluation, not just as sources of data but to develop the evaluation questions, develop evaluative criteria, decide what constitutes credible evidence and contribute to data interpretation and reporting.
4. **Systematic causal analysis** – sometimes using causal attribution (involving an explicit counterfactual) and sometimes using causal contribution (analysing the intervention as an insufficient but necessary part of a causal package that itself is unnecessary but sufficient to cause a result.

The following table describes the key features of these approaches and lists some of the specific methods and techniques.

Table 1: Features of key approaches to impact evaluation

General approach	Key features	Examples of specific methodologies, methods and techniques
1. Theory-based evaluation	An explicit causal chain (or 'theory of change') linking the intervention with specific impacts is articulated and then used to guide the collection of evidence and the analysis of causal contribution by developing hypotheses that can be tested through critical comparisons	• Theory of change • Impact pathway analysis • Realist evaluation
2. Mixed methods	Strengthening the quality of evidence by compensating for the limitations of any one source through a complementary mix of research questions, process for developing research questions, sampling procedures, data collection procedures, type of data, type of data analysis,	Combining sequentially, in parallel or iteratively: • Two types of research questions (with qualitative and quantitative approaches) • The manner in which the research questions are developed (participatory vs pre-planned) • Two types of sampling procedures (e.g. probability and purposive) • Two types of data collection procedures (e.g. focus groups and surveys) • Two types of data (e.g. numerical and textual) • Two types of data analysis (statistical and thematic), and • Two types of conclusions (emic and etic representations, 'objective' and 'subjective' etc.) used
3. Participatory approaches	Engaging a range of stakeholders, including intended beneficiaries, in one or more of the different tasks of impact evaluation – deciding the impacts that are valued; gathering evidence of these; analysing the causal contribution of the intervention; managing the evaluation, including making decisions about design and dissemination	• Systematic Client Consultation (SSC) • Beneficiary Assessment (BA) • Citizen Report Cards (CRCs) and Community Score Cards (CSCs) • Participatory Learning and Action (PLA) family including Rapid Rural Appraisal (RRA), Participatory Rural Appraisal (PRA) and Participatory Poverty Assessment (PPA) • Policy and Social Impact Analysis (PSIA) • Social Assessment (SA) • Self-esteem, associative strength, resourcefulness, action planning and responsibility (SARAR) • Appreciative Inquiry (AI).

The following table summarises how these approaches address the four components of impact evaluation

Table 2: Key approaches and their application in the four impact evaluation components

Approach	Identifying impacts that are valued	Generating or gathering evidence of impacts	Assessing causal contribution of intervention	Managing impact evaluation
1. Theory-based evaluation	*Key strength:* May identify a more comprehensive range of impacts valued by different stakeholders by engaging them in the process of developing, reviewing and refining the program theory *Major potential weakness:* May focus only on stated intended impacts.	*Key strength:* Can help to systematically identify the evidence most needed for the evaluation. Analysis can lead to further data collection to test emerging patterns and hypotheses.	*Key strength:* May encourage systematic use of causal analytical methods to develop and test hypotheses *Major potential weakness:* Sometimes not explicitly addressed, and causal contribution is assumed if there is evidence of the expected causal chain.	Not addressed.
2. Mixed methods	*Key strength:* Supports respect for different ways of thinking and knowing, an intentional inclusion of diverse (stakeholder) perspectives and an appreciation for context and complexity *Major potential weakness:* If not implemented effectively, can lead to a lack of coherence or increase dissent between different perspectives	*Key strength*: Triangulating methods and data sources to provide more comprehensive and less biased evidence of impacts *Major potential weakness:* If not correctly and rigorously applied, can provide a misleading veneer of credible evidence	*Key strength:* Can improve the rigour of causal analysis if the weaknesses of one method is addressed by another	Not addressed

3. Participatory approaches	*Key strength:* By engaging a range of stakeholders, a more comprehensive and/or appropriate set of *valued* impacts are likely to be identified *Major potential weakness:* Stakeholders might try to manipulate evaluative criteria to serve specific interests	*Key strength:* Can involve stakeholders in providing evidence or gathering evidence, creating ownership and understanding, and frequently proven to yield more credible findings (refer *i.a.* to Salmen and Kane, 2006) *Major potential weakness:* Stakeholders might try to manipulate information to serve specific interests – rigorous triangulation needed	*Key strength:* Can involve stakeholders in causal analytical approaches, particularly identifying and investigating alternative causal explanations, providing more comprehensive or new insights, and frequently proven to yield more credible findings *Major potential weakness:* Stakeholders might try to manipulate information to serve specific interests - rigorous triangulation needed	*Key strength:* Major focus - deliberately tries to engage a broader group of stakeholders in decisions about the evaluation, including information flow, in order to enhance ownership, utility and use *Major potential weakness:* Can be time-consuming and resource-intensive

Strategies for improving the rigour of impact evaluation in terms of these four components are discussed below.

5.1. Task 1: Identifying impacts of interest

Not simply about stated objectives

Although most development interventions have stated aims and objectives, impact evaluation is not simply a matter of measuring whether these have been achieved. The stated aims might no longer be quite right in terms of what the intervention is intended to achieve, or might not have been quite right from the start. Not all development interventions have adequately incorporated the needs and values of intended beneficiaries in the stated objectives. Impact evaluation also needs to identify important potential unintended impacts, whether positive or negative that should be included.

Impact evaluations need to answer questions related to 'for whom' the impacts have been intended, and how context influences impacts of interest. One of the tasks of impact evaluation is therefore to be clear about who decides what the right aims are and to ensure that the legitimate different perspectives of different stakeholders are given adequate weight. Participatory approaches to impact evaluation have been proven to be effective in identifying the range of perspectives on what are viewed as impacts of interest (both positive impacts that are intended and negative impacts that should be avoided), and in negotiating agreement about which will be included in the evaluation.

Where there are multiple aims, there needs to be agreement about the standards of performance required in the weighting of these – for example, can an intervention be considered a success overall if it fails to meet some of the targets but does well in terms of the main intended outcome?

For development cooperation, judgments on whether it has been relevant or successful may need to be based on whether the assistance managed to contribute to impact at the level of society (or the country) as a whole, rather than on the direct environment of the intervention alone (Van den Berg, 2005).

Including longer-term impacts

In some types of interventions, impacts emerge quickly. In others impact may take much longer, and change over time. The timing of the evaluation is therefore important. Development interventions are usually assumed to contribute to long-term development (with the exception of humanitarian disaster and emergency situations). Focusing on short-term or intermediate outcomes often provide for more useful and immediate information for policy- and decision-making. Attributing long-term impacts to specific actors or interventions is challenging and sometimes near-impossible. Furthermore, as can be demonstrated through system dynamics simulations, for non-linear situations where things cycle or fluctuate, the apparent impact can depend on when in the iteration or fluctuation the measurement is taking place.

On the other hand intermediate outcomes may be misleading, often differing markedly from those achieved in the longer term. Focusing on short-or intermediate term outcomes may therefore camouflage intervention designs unable to have sustained positive effects in the long term due to inadequate attention to a sustainability strategy – for example not embedding the intervention into existing systems or ensuring the necessary capacities and commitments are in place.

Many of the impacts of interest from development interventions will only be evident in the longer-term, such as environmental changes, or changes in social impacts on subsequent generations. Searching for evidence of such impacts too early might mistakenly conclude that they have failed.

Few impact evaluations are conducted over long time frames that will provide direct evidence of long-term impacts, and in any case results are needed before these impacts become evident to inform decisions on continuation, next phases and up-scaling. Impact evaluations therefore need to identify short-term impacts which will indicate whether longer-term impacts are likely to occur. Evidence of negative impacts might also only be apparent in the long-term, so early warning indicators of these can be important to include. A well-articulated programme theory model, that also specifies the time horizons over which different types of outcomes and impacts could reasonably be expected to occur, can help to identify impacts which can and should be explored in an evaluation. Where impact might be expected to cycle of fluctuate, incorporating these into the programme model will guide more appropriate timing and interpretation of data collection.

The sustainability of positive impacts is also likely to be only evident in the longer-term. For example an effective strategy to reduce child malnutrition in a certain population may quite quickly produce impressive impacts, yet fail soon after in the absence of systems, resources and capacities to maintain the work, or follow-up work, after termination of the intervention. Impact evaluations therefore need to identify other impacts that will be observable in the short-term, such as the institutionalisation of practices and the development of organisational capacity, that are likely to contribute to the sustainability of impacts for participants and communities in the longer-term.

Methods to improve rigour

Theory-based approaches to impact evaluation, which identify a chain (or several chains) of expected results, can be helpful in identifying those short-term impacts which are likely to provide an indication of the achievement of longer-term impacts. Most models do not specify the time horizons over which outputs, outcomes and different kinds of impacts are expected to occur. This is one of the reasons (in addition to the

administrative pressures to provide results as quickly as possible) why many so-called 'impact evaluations' are conducted when it is still much too early to be able to detect even many of the outcomes. Articulating time horizons would provide evaluators with a strong case for explaining to clients why completion of the impact evaluation should be delayed, or, if this is not possible, why it should be clear to clients that a report produced in (say) 2009, will not be able to assess impacts but only certain kinds of outcomes. This is important because the first time that the problems of timing of the impact evaluation is mentioned is in the final report, where this sounds like an excuse to avoid recognition of poor project performance.

Participatory approaches, such as the Most Significant Change (MSC) approach, Outcome Mapping (OM) and Beneficiary Assessment (BA) systematically investigate the impacts valued by different stakeholders and can also highlight unintended consequences. Systematic review of previous evaluations of similar interventions can also increase awareness of possible unintended impacts of the intervention. Methods drawn from the systems field can also contribute to rigour, especially with respect to unanticipated and unintended consequences.

5.2. Task 2: Gathering evidence of impacts

Not just measuring what is easy to measure

There is often a considerable difference between what is easy to measure and what is important to measure (or get credible evidence of). The evaluation should answer questions related to how different stakeholders – including those who may not have a 'voice' - have been perceiving and experiencing different impacts from the interventions. Logistic challenges include time lags before impacts are evident (for example in programs that have inter-generational intended impacts), intrusiveness of accurate measures of behaviour or social outcomes, the multi-dimensional nature of many impacts, the cost of gathering data, and privacy concerns about linking data about individuals from different sources.

Theories of change often provide direction for the identification of potential impacts.

Methods to improve rigour

Good impact evaluation draws on existing data where possible, clearly prioritises which impacts warrant additional data collection and develops (and pilot tests) methods for collecting additional data.

Theory-based evaluation approaches also provide guidance in this task by identifying short-term outcomes that may be evident during the life of the evaluation, and that are pre-cursors to the longer-term impacts of interest.

Mixed method approaches can improve the rigour of evidence through compensating for the limitations of any one method. If properly applied, this approach has the potential to increase significantly the quality and rigour of an evaluation through extensive triangulation of methods and sources. It strengthens evidence about what impacts have been achieved in several ways:

- A mix of methods can be used to assess important outcomes or impacts of the intervention being studied. If the results from different methods converge, then inferences about the character and magnitude of these impacts will be stronger. For example, triangulation of standardised indicators of children's educational attainments with results from an analysis of samples of children's academic work yields stronger confidence in the educational impacts observed than either method alone (especially if the methods employed have offsetting biases).
- A mix of methods can be used to assess different facets of complex outcomes or impacts, yielding a broader, richer portrait than can one method alone. For example, standardised indicators of health status could be mixed with onsite observations of practices related to dietary nutrition, water

quality, environmental risks or other contributors to health, jointly yielding a richer understanding of the intervention's impacts on targeted health behaviours.

- One set of methods could be used to assess outcomes or impacts and another set to assess the quality and character of programme implementation, thus enhancing impact evaluation with information about programme experiences.
- Different methods could be used sequentially to develop better questions in the second method, or a targeted sample for assessment or some other measurement enhancement.

There are also other ways to increase rigour in this task, including

- Ensuring that the sampling frame and the sample selection strategies cover the whole of the target intervention and comparison populations. Many sampling frames leave out important sectors of the population (usually the most vulnerable groups or people who have recently moved into the community), while respondent selection procedures often under-represent women, youth or the elderly or ethnic minorities. This is critical because important positive or negative impacts on the vulnerable groups (or other important sectors) are completely ignored because they did not even get included in the sample. This is particularly important (and frequently ignored) where the evaluation uses secondary data sets, as the evaluator often does not have access to information on how the sample was selected.

5.3. Task 3: Assessing causal contribution

Causal attribution or contribution

Some types of interventions lead directly to specific impacts which are unlikely to have occurred in the absence of the intervention, for example the results of vaccination programmes, the use of bednets against malaria or the introduction of a specific agricultural technology. In these cases we can talk about 'causal attribution'.

- there is a linear causality between input > output > impact in order to achieve a clear goal (possibly spelled out in the programme documents), and it is possible to attribute impact to the whole intervention ('treatment');
- the context will be the same, or will be kept constant, or does not affect the results, wherever the intervention is applied;
- the particular intervention works in isolation from other interventions;
- the intervention is implemented exactly as planned and is uniform across all project settings.

Other types of interventions make a causal contribution to specific impacts when there is also a contribution from other factors, including complementary interventions and favourable implementation contexts – for example provision of textbooks which can contribute to improved educational outcomes when it is combined with effective teaching and access to school.

In such cases it is not possible to attribute impact to a single cause and it is necessary to look for multiple, alternative, explanations and contextual factors.

Attribution is particularly difficult in cases where development interventions take place in complicated situations, such as those characterised by multi-site, multi-agency, integrated programmes. For example, an intervention aimed at eliminating child labour will, for sustained success, need to simultaneously withdraw children from work, provide them with appropriate educational opportunities, improve the school environment for such children, create societal awareness, ensure that the perceived 'value' of the children is not diminished, build the capacities and means of families to earn additional incomes and provide a policy and legal environment that stop the perpetrators.

The different ways in which interventions contribute to impacts thus require different designs and methods to undertake causal analysis in impact evaluation. Furthermore, inferring causation clearly becomes increasingly complex across continuums that extend from

- outputs to outcomes to impact, with progressively weakening reliability of the programme theory and hence the certainty of attribution;
- projects (usually characterised by single interventions with explicit objectives and change effects measured in the short term) to comprehensive programmes where activities cut across sectors, themes or geographic areas, with a number of confounding variables, and change effects measured in the longer term; and
- 'simple, single strand interventions' to 'complex adaptive systems' characterised by large numbers of unknown variables and unknown causal connections between variables, by interactions, feedbacks and nonlinear relationships, and high sensitivity to small perturbations.

Methods to improve rigour

Analysis of causal attribution aims to assess the proportion of observed change which can really be attributed to the evaluated intervention.

Rigour in causal attribution analysis involves systematically creating, locating or simulating a counter-factual – an estimate of what would have happened in the absence of the intervention. This might be done through a control group (with random assignment to either receive the intervention/participate in the project or not), a comparison group (matched on observable relevant variables), propensity scores, regression discontinuity designs, or time series analysis. Further detail about these approaches is provided in section 1 of this guidance document.

Analysis of causal contribution aims to demonstrate whether or not the evaluated intervention is one of the causes of observed change. Contribution analysis relies upon chains of logical arguments that are verified through a careful confirmatory analysis.

Rigour in causal contribution analysis involves systematically identifying and investigating alternative explanations for observed impacts. This includes being able to rule out implementation failure as an explanation of lack of results, and developing testable hypotheses and predictions to identify the conditions under which interventions contribute to specific impacts. This might be done through contribution analysis, realist evaluation, impact pathways analysis, as well as system dynamics and simulation based methods from complex adaptive systems (e.g. agent based modelling). Further detail about these approaches is provided in Section 4 of this guidance document.

Many sciences do not work with explicit counterfactuals. For example, in 1919 a solar eclipse was used to verify the prediction of the General Theory of Relativity that light would be bent by gravity. Observations confirmed that stars that were supposed to be invisible according to Newtonian physics, because they were eclipsed by the sun, were in actual fact visible, because their light rays were bent by the gravity of the sun. Thus it turned out that Newtonian physics was "counter to the facts" and the General Relativity Theory was in conformity with the facts. Nevertheless, the natural sciences do not tend to describe this as a "counterfactual" testing, but as a prediction which can be falsified. The counterfactual is "implied" or "virtual", and can be said to consist of the statement "if Newtonian physics were the case, then stars x, y and z would not have been visible during the observation in 1919".

Similarly, the historical sciences never pose a formal counterfactual, but are full of implied or virtual counterfactuals. History consists mainly of establishing the "factual", not the "counterfactual", and rarely speculates on what could have been the case but did not happen. However, it has also been argued that history focuses on explaining why a certain course of events was inevitable or why certain factors were necessary conditions for future events to unfold. This has led to a recent development in historiography of

counterfactual history, which poses the "what if" question on key moments in human history. Criticism of this new development by established historians has been that this was no more than frivolous entertainment, which has led the counterfactual historians to reply that regular history is by definition full of implied and virtual counterfactuals. To describe a key moment in history means after all that if that moment would not have happened as it did, history would have taken another (counterfactual) course.

5.4. Task 4: Managing impact evaluation

Not just a technical exercise

Whether conducted internally or externally, evaluations need to be properly managed to ensure that they balance the requirements of utility, accuracy, feasibility and propriety in an often political environment. Given the increasing focus on impact evaluation as part of evidence-based decision-making, they are well positioned to have a powerful influence on policies, strategies and budgets for development. In turn they are attractive targets for political or ideological influence, and different motivations and demands for information can shape the nature of impact evaluations and findings.

Most impact evaluations must be designed, implemented, analysed, disseminated and used under budget, time and data constraints and while having to diverse and often competing political interests. Given these constraints, the management of a real-world evaluation is much more complicated than selecting among the alternative textbook design. Many evaluations fail because the stakeholders were not involved, or the findings were not used because they did not address the priorities of the stakeholders. Others fail because of administrative or political difficulties in getting access to the required data, being able to meet with all of the individuals and groups that should be interviewed, or being able to ask all of the kinds of questions that the evaluator feels are necessary. Many other evaluations fail because the sampling frame, often based on existing administrative data, omits important sectors of the target population - often without anyone being aware of this. In other cases the budget was insufficient, or was too unpredictable to permit an adequate evaluation to be conducted.

While many of these constraints are presented in the final evaluation report as being completely beyond the control of the evaluator, in fact their effects could often have been reduced by more effective management of the evaluation. For example, a more thorough scoping analysis could have revealed many of these problems and the client could then have been made aware of the likely limitations on the methodological rigor of the findings. The client and evaluator could then strategise to either seek ways to increase the budget or extend the time, or agree to limit the scope of the evaluation and what it promises to deliver. If clients understand that the current design will not hold up under the scrutiny of the projects critics, they can often find ways to help address some of the constraints.

Methods to improve rigour

For the sake of honest commitment to development, evaluators and evaluation units should ensure that impact evaluations are designed and executed in a manner that limits manipulation of processes or results towards any ideological or political agenda. They should also ensure there are realistic expectations of what can be achieved by a single evaluation within time and resource constraints, and that findings from the evaluation are presented in ways that are accessible to the intended users. This includes finding a balance between simple, clear messages and properly acknowledging the complexities and limitations of the findings.

International evaluation standards (such as the OECD DAC / UNEG Norms and Standards and/or the standards and guidelines developed by national or regional evaluation associations), should be applied where appropriate.

6. When to do an impact evaluation

Box 1: Examples of types of evaluation in the life of an intervention

- Ex-ante evaluation
- Built-in / self-evaluation
- Interim / formative / mid-term (including implementation or process) evaluation
- Inter-phase evaluation
- Summative / ex-post / outcomes evaluation
- Impact evaluation (which usually includes outcomes)

6.1. Impact evaluation is one among many different types of evaluation

Many managers and policymakers tend to assume that impact evaluation is synonymous with any kind of evaluation. It is thus often the case that they will request an 'impact evaluation' when the real need is for a quite different kind of evaluation (for example to provide feedback on the implementation process, or to assess accessibility of programme services to vulnerable groups). Ensuring clarity in the information needed and for what purpose is an important prerequisite to defining the type of evaluation to be conducted. Section 2 highlights the fact that impact evaluation is focused on results and changes at all levels of results chain. Impact evaluation draws on, and complements rather than replaces other types of monitoring and evaluation activities (Box 1). It does not answer all types of evaluation questions and can usually not be done without knowledge determined through prior evaluations, for example of the relevance and quality of the intervention design and the design processes, and the (cost)effectiveness and efficiency of the institutional and management systems and of implementation processes. It should therefore be seen as one in a cycle of potentially useful evaluations in the lifetime of an intervention.

6.2. Impact evaluations should be undertaken only under certain circumstances

An impact evaluation should ideally be conducted when an assessment shows that political, technical, resource and other practical considerations are adequate.

- The evaluation has a clearly defined purpose and agreed upon intended use, appropriate to its timing and with support of influential stakeholders.
- There is a clear demand for information that can *only* be satisfied through an impact evaluation, and the cost of *not* having this information will be too high.
- There is clarity about the evaluation design, which depends on the use of the evaluation, the nature of the intervention (e.g. discrete and stable or broad and emergent) and the state of existing knowledge about it. The evaluation design has to be clearly described and well justified after due consideration of alternatives and constraints.
- The evaluation design has a chance to be credibly executed given the nature and context of the intervention, the data and information needs and the availability of adequate resources and expertise to conduct the evaluation.

Impact evaluations may not be appropriate when:

- Other valuable forms of evaluation will yield more useful information to support decisions to be made or serve other purposes. Before embarking on an impact evaluation its value should therefore be assessed against the full spectrum of evaluation types and measured against ongoing development priorities;
- It moves too much resources and attention away from the need to develop and use a rich spectrum of evaluation approaches and capacities;
- Political, technical, practical or resource considerations are likely to prevent a credible, rigorous, useful evaluation;

- When there are signs that the evaluation will not be used (or may be misused, for example for political reasons).

7. Design options for causal analysis

In terms of causal analysis, the two broad options are:

- **Analysis of causal attribution**, using experimental or quasi-experimental designs, incorporating an explicit counter-factual (described in more detail in section one of this document)
- **Analysis of causal contribution**, using iterative theory-building and testing using critical comparisons and predictions (described in more detail in section two of this document)

Causal attribution is appropriate when it is possible to create, locate or simulate a counterfactual.

> *The basic idea of counterfactual theories of causation is that the meaning of a singular causal claim of the form "Event c caused event e" can be explained in terms of counterfactual conditionals of the form "If c had not occurred, e would not have occurred". (Menzies, 2001)*

The logic of counterfactuals as conditional statements asserts:

> If x, then y (x is a sufficient condition to produce y)
>
> If not x, then not y (x is a necessary condition to produce y).

Counterfactual formulations work well with discrete interventions and discrete results. Counterfactual formulations specify strong and direct causal links between discrete causes and effects. In the real world, these conditionals become probabilities rather than certainties.

In a development context, counterfactual conditionals work well for discrete interventions like immunisations, introduction of textbooks into a school, or a standardised microfinance program. Indeed, such formulations nicely capture counterfactual logic, which is why they are typically used as examples of "rigorous" research and evaluation. In essence, counterfactuals work as a form of logic where the world is governed by specific if-then generalisations (including a chain of such if-then statements). This is what is commonly described philosophically and logically as linear thinking. For these types of interventions, analysis of causal attribution is likely to be appropriate, providing there is a high degree of control of the presence or absence of x, and the direct measurement of y to test: if x, then y, and if not x, then not y.

For complex, dynamic, and emergent interventions, it is often not possible to posit a discrete "x" or a singular "y," so the conditional counterfactual, if x, then y, lacks specificity and therefore lacks meaning. This is the case, for example, with comprehensive development initiatives which involve multiple projects and many actors over several years.

Not even a chain of causal connections can be specified (e.g. a complicated logic model) because the many and diverse elements of the intervention (community development) and the multiple actors are interdependent in ways that cannot be disentangled or specified. It is possible to map and trace these interrelationships and their consequences, but these maps do not reduce to if-then counterfactual conditional statements..

In development, the counterfactual control condition (not x) does not exist in a meaningful way for complex, multidimensional interventions and for interventions that occur over a somewhat lengthy period of time (three or more years). A steady-state, static control counterfactual does not exist, and cannot exist, because the world will have changed substantially during that time. In such cases, analysis of causal contribution is more appropriate (Patton, 2008, Pawson, 2008).

Subgroup 2

NONIE IMPACT EVALUATION GUIDANCE

Section 2*: Experimental and Quasi-Experimental Approaches to Impact Evaluation

*Please refer to Subgroup 1 document.

Subgroup 2

NONIE IMPACT EVALUATION GUIDANCE

Section 3: Analysis of Causal Contribution Approaches to Impact Evaluation

8. The need for approaches for analysing causal contribution

Causal attribution is based on being able to create, locate or simulate a counter-factual – an estimate of what would have happened in the absence of the intervention. When this is not possible or appropriate, analysis of causal contribution is needed instead. Where interventions are complicated (consisting of many components, all of which are needed to produce the impacts) or complex (evolving and emergent), it is not possible or appropriate to develop an explicit counter-factual.

Aspect	**Simple intervention**	**Complicated or complex intervention**	**Challenges for impact evaluation**
Simultaneous causal strands	Single causal strand. Intervention is sufficient to produce the impacts	Multiple simultaneous causal strands required to produce the impacts	As effective programs may need to optimise several causal paths, not just one; evaluation should both document and support this. Program components may not be effective unless they are in a favourable context. Replication of an effective program may depend on understanding the context that supports it.
Alternative causal strands	Universal mechanism. Intervention is necessary to produce the impacts	Different causal mechanisms operating in different contexts	The use of an explicit counter-factual may be inappropriate when there are alternative ways to achieve the outcome.
Non-linearity and disproportionate outcomes	Linear causality, proportional impact	Recursive, with feedback loops	A small initial effect may lead to a large ultimate effect through a reinforcing loop or critical tipping point. Evidence of impacts may not be readily evident.
Emergent outcomes	Pre-identified outcomes	Emergent outcomes	Specific measures may not be able to be developed in advance, making pre- and post-comparisons difficult and the development of a counter-factual impossible.

9. Analysing causal contribution

9.1. The analysis of causal contribution features:

- Use of theory-based evaluation in an iterative process of building, testing and refining the causal model
- Attention to rigour through critical analysis of evidence not through a particular research design

Contribution analysis (Mayne, 2001; Mackay et al, 2002; Leeuw, 2003; Mayne and Rist, 2006, EuropeAid, 2006) involves a full description of the external intervention(s) coupled to an assessment of changes in development status which appears to have been associated with it. The intent is to show the extent to which the observed (detected) changes can be attributed to the interventions being evaluated. The approach works under the assumption that it is often more appropriate to use impact evaluations to identify an intervention's *causal contribution* to observed changes, rather than attempt to establish causal attribution. Contribution analysis is a tool that helps tell a well-founded, credible story about how an intervention is performing in terms of its intended results.

This methodology sets out to test the theory of change behind an intervention, paying attention to alternative explanations for the observed results. The TOC sets out why it is believed that the intervention will lead to a contribution to the intended results, including impacts. The analysis tests this theory against logic and the evidence available on the various assumptions behind the theory of change, and examines other influencing factors as possible alternative explanations. The analysis either confirms the postulated theory of change, or suggests revisions in the theory where the reality appears otherwise. Causal contribution is thus inferred from the following evidence:

1. There is a reasoned theory of change for the intervention: it makes sense, it is plausible, and is agreed by key players.
2. The activities of the intervention were implemented.
3. The theory of change—or key elements thereof— is verified by evidence: the chain of expected results occurred.
4. Other influencing factors have been assessed and either shown not to have made a significant contribution, or their relative role in contributing to the desired result has been recognised.

The analysis is best done iteratively, building up over time a more robust assessment of causal contribution. The overall aim is to reduce the uncertainty about the contribution the intervention is making to the observed results through an increased understanding of why the observed results have occurred (or not) and the roles played by the intervention and other factors. At level of impact this is the most challenging, and a 'contribution story' has to be developed for each major strategy that is part of an intervention, at different levels of analysis. They would be linked, as each would treat the other strategies as influencing factors.

Mayne's suggestions for contribution analysis (2001) is one example that sets out a series of iterative steps for doing an analysis of causal contribution, using a theory-based evaluation approach:

Step 1: Set out the attribution problem to be addressed
Step 2: Develop the postulated theory of change and the risks to it
Step 3: Gather the existing evidence on the theory of change
Step 4: Assemble and assess the contribution story, and challenges to it
Step 5: Seek out additional evidence
Step 6: Revise and strengthen the contribution story

Contribution analysis can include systematic identification and investigation of variations in implementation and related differences in impacts, identification and investigation of differences in observed impacts, and identification and investigation of possible alternative explanations. For example, if using a before and after design without a control group or comparison group, alternative possible explanations for observed impacts, such as maturation or learning from the pre-test, can be investigated to see if they can be ruled out as causal explanations.

An understanding of context, 'why' things are as they are and as far as possible, circumstances that may lead to success can be obtained by drawing among others from TOC and/or realist evaluation approaches. Using mixed methods for triangulation of sources, methods and theories is also an important means to strengthen evidence inferring causality.

Examples of impact evaluations using contribution analysis include the analysis of the AusAid funded Fiji Education Sector Program (FESP) (Kotvoys, 2006).

9.2. Other approaches to impact evaluation using analysis of causal contribution

Impact pathway evaluation

This is a further development of TBE based on a GTZ Impact Model (Kuby, 1999), which shows an 'attribution gap' between the direct benefits (which can be demonstrated through project level monitoring and evaluation) and the indirect, longer term development results (observed changes) of the intervention. Impact pathway evaluation represents a set of hypotheses about what needs to happen for the intervention outputs to be transformed, over time, into impact on highly aggregated development indicators. This 'bridge' is constructed during the planning stage of the intervention, and helps stakeholders to develop a common understanding of what it is trying to achieve as well as identify likely scaling-up pathways. This is then investigated empirically through testable hypotheses and predictions. (Douthwaite. and Olanrewaju, 2002)/

Examples of development impact evaluations using impact pathways for evaluation include the Challenge Program on Water and Food (CPWF) (Douthwaite et al, 2006), and integrated pest management (Douthwaite et al, 2003b).

9.3. Realist evaluation

Realist evaluation (Pawson and Tilley 1997; Henry et al, 1998; Pawson, 2002)is a particular type of TBE that pays particular attention to influences from context (implementation environments and participant characteristics) and to the likelihood that different contexts will exert different influences on the same theory of change, rendering it differentially enacted and effective in different contexts. In the evaluation of complex multi-site interventions, additional challenges include (i) the different contexts may yield different reactions to the same intervention, and (ii) putting in place alternative mechanisms may produce different results. . Realist evaluation does not ask "What works?" or "Does this intervention work?" but ask instead "*What works for whom in what circumstances and in what respects, and how*?"

Examples of development impact evaluations using realist evaluation include the evaluation of the training partnership between the World Bank and the University of Sao Paulo (Marra, 2004).

Subgroup 2

NONIE IMPACT EVALUATION GUIDANCE

Section 4: Other Approaches for Impact Evaluation

10. Overview

The previous section focused on approaches for analyzing the causal contribution of interventions. This section lists other approaches which can improve the rigour of impact evaluation. Some contribute to a more comprehensive or considered scope of impacts that are included, some generate more comprehensive evidence of impacts, or support the uptake of findings.

General approach	Key features	Specific methods
Systemic evaluation	Application of various methods from systems which all focus on: inter-relationships between elements of a situation; an investigation of the different perspectives through which a situation can be evaluated; and assessing the implications of the boundaries that are drawn around a situation and any inquiry into that situation.	• Systems dynamics • Viable Systems (Cybernetics • Soft Systems • Critical Systems • Cultural Historical Activity Theory • Complex Adaptive Systems
Participatory approaches	Engaging a range of stakeholders, including intended beneficiaries, in one or more of the different tasks of impact evaluation – deciding the impacts that are valued; gathering evidence of these; analyzing the causal contribution of the intervention; managing the evaluation, including making decisions about design and dissemination	• Beneficiary Assessment (BA); • Participatory Learning and Action (PLA) family including Rapid Rural Appraisal (RRA), Participatory Rural Appraisal (PRA) and Participatory Poverty Assessment (PPA); • Policy and Social Impact Analysis (PSIA); • Social Assessment (SA); Systematic • Client Consultation (SSC) • Self-esteem, associative strength, resourcefulness, action planning and responsibility (SARAR); • Objectives-Oriented Project Planning (ZOPP); • Appreciative Inquiry (AI) • Citizen Report Cards (CRCs) and Community Score Cards (CSCs).
Other approaches		• Outcome Mapping • Success Case • Most Significant Change • Method for Impact Assessment of Projects and Programs (MAPP) • Participatory Impact Pathway Assessment

Table showing how the various approaches address the four key tasks of impact evaluation

Approach	Identifying impacts that are valued	Generating or gathering evidence of impacts	Assessing causal contribution of intervention	Managing impact evaluation
Systemic evaluation			Major focus – supports the development of causal models which address non-linear and non-simple causality	Some approaches pay attention to who is included in the evaluation and how decisions are made
Participatory approaches	Major focus – by engaging a range of stakeholders, a more comprehensive set of valued impacts are likely to be identified	Can involve stakeholders in providing evidence or gathering evidence	Can involve stakeholders in causal analytical approaches, particularly identifying and investigating alternative causal explanations	Major focus – deliberately tries to engage a broader group of stakeholders in decisions about the evaluation, including information flow

10.1. Systemic evaluation

Systemic evaluation has been increasingly used in recent years. These approaches use a range of methods and methodologies developed over the past 50 years within the systems field. Systemic evaluation approaches are primarily focused on generating deep insights into complex situations and resolving issues that flow from these situations. The approaches share three features (Williams et al, 2007): a focus on inter-relationships between elements of a situation; an investigation of the different perspectives through which a situation can be evaluated; and assessing the implications of the boundaries that are drawn around a situation and any inquiry into that situation.

Systemic evaluation provides powerful tools for identifying impacts of interest, including exploring unanticipated consequences and the implications of acknowledging (and especially not acknowledging) the range of motivations that reveal different notions of worth within interventions. Systemic evaluation is especially useful for framing an evaluation - that is, deciding criteria of worth, the scale of an evaluation and its overall scope. It also provides useful tools for gathering evidence of impacts and understanding the causal contribution of the intervention. In particular it generates insights into appropriate management of the impact evaluation, including understanding factors which might inhibit rigorous impact evaluation.

Different types of systemic evaluation are particularly useful for particular kinds of evaluation questions, as shown below:

General systems principles:

- What is the nature of the inter-relationships within an intervention? What is the structure of these inter-relationships, what are the processes between them, what are the patterns that emerge from those processes, with what are the results? Why does this matter? To whom? In what context?

- What are the different ways in which an intervention can be understood? How are these different understandings going to affect the way in which people judge "success"? How will it affect their behaviour - especially when things go wrong from their perspective? With what result and significance?
- Who or what is being excluded, marginalised or made a victim, by the way in which this intervention is being viewed or is operating? What does this say about what is "valued", by whom, in this situation? What are the consequences of these decisions?

From system dynamics methodologies:

- How does "delay" impact on the performance of the intervention?
- How do patterns of feedback affect the behaviour of an intervention?
- What controls the way in which resources flow through a intervention? How does this affect performance?

From Viable Systems (Cybernetics) methodologies:

- What are the operational, co-ordination, management, strategy and governance needs of the intervention necessary to deliver on its purpose?
- What information is needed at each level of an intervention to achieve the purpose?
- How does information flow through a intervention?
- Is the right information available at the most appropriate level of an intervention's hierarchy of tasks?

From Soft Systems methodologies:

- What are the different ways in which the intervention can be viewed?
- How does each of these ways express and give meaning to what people working within the intervention value?
- How are these perspectives reflected in people's motivations and behaviours?
- How does each of these ways affect a program's performance?

From Critical Systems methodologies:

- What are the implications of how we choose to measure the "success" of a program? Did these choices really measure the success or just something that was easily measured? If just the measures were achieved, then would it be deemed successful? To whom?
- Who was trusted to take the key decisions in the intervention and what were the implications of that choice? To what extent were they in control of the intervention or predominantly influenced by the environment that lies outside the intervention (i.e. can they really be held accountable)?
- Why were the "experts" considered to have the necessary expertise - and what was the consequence of that decision?

- Who or what was marginalised by the intervention, or by the way the intervention was conceived and framed? What are the practical and ethical implications of this - and what does that say about the values implicit in the intervention that may form the criteria for an evaluation?

By replacing "intervention" with "evaluation", critical systems methodologies provide a powerful alternative to traditional evaluability assessments

From Cultural-Historical Activity Theory methodologies:

- What tools, rules and roles were brought to bear on each purpose that motivated participants? With what consequence on whom in what context/histories/environment?
- How well did participants handle contradictions between (and within) these rules, roles, tools purposes and contexts/histories/environments? With what consequence on whom?

From Complex Adaptive Systems methodologies:

- Does the intervention display simple, complicated, complex or chaotic behaviours? From whose perspective? With what consequences?
- What patterns are observable within the intervention? What have been the likely generators of those patterns?
- What deep behavioural rules are operating that determine these patterns?
- What levels of agreement or differences are there in the situation? From whose perspective? What are the implications of that?
- What levels of turbulence or certainty are there in the programme or its environment? From whose perspective?
- How do these degrees of certainty and agreement interact? With what consequence?

10.2. Approaches based on stakeholder participation

These approaches are built on the principle that stakeholders should be involved in all stages of evaluation, including determining objectives and impacts, identifying and selecting indicators, and participating in data collection and analysis. It was developed out of disillusionment with ‘external’ and ‘scientific’ evaluation (Hulme, 2000; Iverson, 2003). It has taken root since development initiatives, primarily in the World Bank, experienced setbacks arising from the failure to understand and take into account social and cultural factors (Kottak, 1985). This has been accompanied by the emergence of a more holistic view of development that includes the social and cultural, a recognition that a variety of stakeholders in civil society have a role to play, a more aggressive exploration and a more accessible presentation of the philosophy of knowledge, the strengthening of qualitative research and the ability to analyse it, and the development of new approaches, including participatory approaches, and what might be called ‘designer’ toolkits, each aimed at a specific issue (Salmen and Kane, 2006).

Methodologies commonly included under this umbrella include Beneficiary Assessment (BA); the Participatory Learning and Action (PLA) family including Rapid Rural Appraisal (RRA), Participatory Rural Appraisal (PRA) and Participatory Poverty Assessment (PPA); Policy and Social Impact Analysis (PSIA); Social Assessment (SA); Systematic Client Consultation (SSC); Self-esteem, associative strength, resourcefulness, action planning and responsibility (SARAR); Objectives-Oriented Project Planning (ZOPP); Appreciative Inquiry (AI), and Citizen Report Cards (CRCs) and Community Score Cards (CSCs).

The characteristics, differences, utility and credibility of these methodologies are comprehensively described in a recent publication by Salmen and Kane (2006). They argue - and give convincing examples - that where executed according to accepted standards, using a mix of qualitative and quantitative data as well as rigorous triangulation, better results are obtained than through other methods. Among others a World Bank paper (World Bank, 2005) on the Bank's engagement with civil society quotes several studies that have found a large and significant difference in the levels of success between projects conducted without civic engagement, and those that included elements of true participation.

10.3. Outcome Mapping

Outcome Mapping (OM) (IDRC, 2001) is a methodology that focuses on outcomes as behavioural change[3]. [4]The outcomes can be logically linked to an intervention's activities, although they may not be necessarily directly caused by them. These changes are aimed at contributing to specific aspects of human and ecological well-being by providing partners with new tools, techniques and resources to contribute to the development process. 'Boundary partners' are individuals, groups and organisations with whom the intervention interacts directly and with whom the intervention anticipates opportunities for influence; most activities will involve multiple outcomes because they have multiple boundary partners.

10.4. Success Case Method

The Success Case Method (Brinkerhoff, 2003,)is a widely adopted example of mixed method, drawing from several established traditions including theory based evaluation, organisational development, appreciative inquiry, narrative analysis and quantitative statistical analysis of impact. It has been expanded in scope by those who combine it with realist methodologies (e.g. Dart) and soft systems methodologies (e.g. Williams). It also shares much in common with the Positive Deviance approach that has been applied to health interventions in many developing countries.

The Success Case Method identifies individual cases that have been particularly successful (and unsuccessful) and uses case study analytical methods to develop credible arguments about the contribution of the intervention to these.

10.5. Most Significant Change

The most significant change (MSC) technique (Davies and Dart, 2005) is a form of participatory monitoring and evaluation[5]. It is participatory because many intervention stakeholders are involved both in deciding the types of change to be recorded, and in analysing the data. It is a form of monitoring because it occurs throughout the intervention cycle and provides information to help people manage the intervention. It contributes to impact evaluation in part because it provides data on impact and outcomes that can be used to help assess the performance of the intervention as a whole – but largely through providing a tool for identifying and rating the impacts that are valued by different stakeholders.

[3] Defining outcomes as changes in the behaviour, relationships, activities or actions of the people

[4] OM has an active listserv and website also of about 750 (www.outcomemapping.ca plus significant material on OM at www.idrc.ca/evaluation

[5] There is currently an active listserv of around 750 persons experimenting with this approach

10.6. MAPP

The 'Method for Impact Assessment of Projects and Programs' (MAPP) (Späth 2004) is a methodological framework for combining a qualitative approach with participatory assessment instruments, including a quantification step. It orients itself towards principles and procedures of PRA methodology, including triangulation, optimal ignorance[6] and communal learning[7]. A major element of this methodology is the conduct of workshops with representatives of relevant stakeholders. Perceived key processes are jointly reflected in structured group discussions in which at least six interlinked and logically connected steps are accomplished: (i) lifeline; (ii) trend analysis; (iii) activity list; (iv) influence matrix; (v) transect – or data cross checking; and (vi) development and impact profile.

Further details about these various methods are included in a website being develop to accompany SG2 guidance.

Annex 1 presents a number of cases that address some of the approaches and methods that are outlined in Sections 3 and 4.

[6] The capability to select relevant data and to avoid information overkill

[7] The findings of an assessment are the result of a communication process among relevant groups

Key References

Baker JL. 2000. *Evaluating the impact of development projects on poverty: A handbook for practitioners.* The World Bank, Washington DC.

Bamberger M., Rugh J and Mabry L. 2006. *RealWorld Evaluation: Working under budget, time, data and political constraints.* Sage Publications. 468pp.

Bamberger M. 2007. *A framework for assessing validity and utilisation of evaluations.* Presentation at the Conference of the American Evaluation Association, Baltimore, USA.

Bamberger M and White H. 2007. *Using strong evaluation designs in developing countries: Experiences and challenges.* Journal of Multidisciplinary Evaluation 4(8): 58-73

Bauer H et al. Springer Heinze. 2002. *Establishing plausibility in impact assessment.* The Working Group on Assessing the Impact of Agricultural Research on Development. Paper presented at the International Conference on the Impacts of Research and Development, San Jose, Costa Rica, 4-8 February.

Brinkerhoff, R. 2003, *The Success Case Method* Berrett Koehler : San Francisco

Bryman A. 2007. *Barriers to integrating quantitative and qualitative research.* Journal of Mixed Methods Research 1(1): 8-22.

Burke JR and Onwuegbuzie AJ. 2004. *Mixed methods research: A research paradigm whose time has come*. Educational Researcher 33(7): 14-26.

Centre for Global Development. 2006. *When will we ever learn? Improving lives through impact evaluation.* Report by the Evaluation Gap Working Group, CGD, Washington DC.

Chambers R. 1997. *Whose Reality Counts? Putting the Last First.* London: Intermediate Technology Publications.

Chatterji M. 2007. *Grades of evidence: Variability in quality of findings in effectiveness studies of complex field interventions.* American Journal of Evaluation 28(3): 239-255.

Chen HT. 2006. *A theory-driven evaluation perspective on mixed methods research.* Research in the Schools 13(1): 75-83.

Claremont Graduate University Stauffer Symposium. 2006. *What constitutes credible evidence in evaluation and applied research?* With M Scriven, J Greene, T Schwandt, M Mark, G Henry and others. 19 August. Video streaming at http://www.cgu.edu/pages/4085.asp

Cook, T. D. 2000. The false choice between theory-based evaluation and experimentation. *New Directions in Evaluation: Challenges and Opportunities in Program Theory Evaluation* 87: 27-34.

Creswell JW et al. 2006. *How interpretive qualitative research extends mixed methods research.* Research in the Schools, 13(1): 1-11.

Davidson, E. J. 2006 The RCTs-Only Doctrine: Brakes on the Acquisition of Knowledge? *Journal of MultiDisciplinary Evaluation,* Number 6 (November 2006): ii-v. Downloaded 31 December 2007 http://survey.ate.wmich.edu/jmde/index.php/jmde_1/article/view/35/45

Davies R. 2005. *Scale, complexity and the representation of theories of change.* Evaluation 11(2): 133-149.

Davies R and Dart J. 2005. *The 'Most Significant Change' Technique.* Available at http://www.mande.co.uk/docs/MSCGuide.pdf.

Davis P. 2005. *The limits of realist evaluation: surfacing and exploring assumptions in assessing the best value performance regime.* Evaluation 11(3): 275-295.

Douthwaite B. and Olanrewaju A. 2002. *Impact Pathway Evaluation. An approach for achieving and attributing impact in African agriculture.* Paper presented at the CGIAR Social Science Conference in Cali, Colombia, 11-14 September.

Douthwaite, B , Alvarez, S., Cook, S., Davies, R., George, P., Howell, J. 2006 *'The Use and Potential of Impact Pathways in the Challenge Program on Water and Food (CPWF)'* Discussion Paper. Downloaded 6 January 2008 from http://boru.pbwiki.com/f/Impact-pathways-discussion-paper-CPWF_2.doc/

Douthwaite, B., Delve, R., Ekboir, J., and Twomlow, S. (2003a) 'Contending with Complexity: The Role of Evaluation in Implementing Sustainable Natural Resource Management', International Journal of Agricultural Sustainability, 1 (1): 51-66.

Douthwaite, B., Kubyb, T., van de Fliertc, E., and Schulzd, S. (2003b) 'Impact pathway evaluation: an approach for achieving and attributing impact in complex systems' Agricultural Systems 78:243–265.

Earl E., Carden F and Smutylo T. 2001. *Outcome Mapping: Building Learning and Reflection into Development Programs.* Published by the International Development Research Centre (IDRC), Ottawa, Canada.

EuropeAid 2006 'Cause and effect analysis'. Methodological guidelines. http://ec.europa.eu/europeaid/evaluation/methodology/methods/mth_att_en.htm

Galster et al. 2004. *Measuring the impacts of community development initiatives: a new application of the adjusted interrupted time-series method.* Evaluation Research 38(6): 502-538.

GEF. 2006. *Impact Evaluation: Final report on a proposed approach*. Working Document prepared by Foundations of Success. Global Environment Facility, Washington DC, USA.

GEF. 2007. *Methodological challenges in impact evaluation: The case of the Global Environment Facility (GEF).* Discussion paper prepared by Vaessen J and Todd D.

Glass, G. 1997 , Interrupted Time Series Quasi-Experiments' in Jaeger, R. M, 1997 *Complementary methods for research in education*. 2nd Edition. Pp.589-608. Washington D. C.: American Educational Research Association.

Green LW and Glasgow RE. 2006. *Evaluating the relevance, generalisation and applicability of research.* Evaluation and the Health Professions. 29(1): 126-153.

Greene JC et al. 2001. *The merits of mixing methods in evaluation*. Evaluation 7(1): 25-44.

Greene JC and Caracelli VJ (eds.). 1997. *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms,* New Directions for Evaluation 74. San Francisco: Jossey Bass.

Hammersley M. 2005. *Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers' case for research-based policy making and practice.* Evidence and Policy 1(1): 85-100.

Hansen HF. 2005. *Choosing evaluation models: A discussion on evaluation design.* Evaluation Vol 11(4): 447-462

Harris J. 2002. *The case for cross-disciplinary approaches in international development*. World Development 30(3): 487-496.

Henry GT; and Pawson R and Tilley N. 2005. In Mathison S (Ed). *Encyclopaedia of Evaluation.* Thousand Oaks CA: Sage. 481pp

Henry GT, Julnes G and Mark, MM. (Eds). 1998. *Realist evaluation: An emerging theory in support of practice.* New Directions for Evaluation 78.

Herweg K and Steiner K. 2002. *Impact monitoring and assessment.* Bern, Centre for Development and Environment (CDE, Switzerland). Vol 1 and 2. Quoted in Späth B, 2004.

Holland J and Blackburn J. 1998. *Whose Voice? Participatory Research and Policy Change*. London: Intermediate Technology Publications

House ER. 2006. *Blowback.* Consequences of evaluation for evaluation. Keynote address at Conference of the American Evaluation Association, Portland, USA .

Howe KR. 2004. *A critique of experimentalism.* Qualitative Inquiry 10(1): 42-61.

Hulme D. 2000. *Impact assessment methodologies for microfinance: Theory, experience and better practice.* In World Development 28(1):79-98.

Iverson A. 2003. *Attribution and aid evaluation in international development: A literature review.* Published by the International Development Research Centre (IDRC), Ottawa, Canada.

Kanbur R. 2002. *Economics, social science and development*. World Development 30(3): 477-486.

Kottak CP. 1985. *When People Don't Come First: Some Sociological Lessons from Completed Projects.* In *Putting People First*, 2nd Edition, Michael Cernea (Ed.). New York/London: Oxford University Press.

Kotvoys, F 2006 'Contribution Analysis – A New Approach to Evaluation in International Development'. Paper presented at the 2006 International Conference of the Australasian Evaluation Society, Darwin. Downloaded 6 january 2008 from http://www.aes.asn.au/conferences/2006/papers/022%20Fiona%20Kotvojs.pdf.

Krznaric R. 2007. *How change happens. Interdisciplinary Perspectives for Human Development,* OXFAM Research Report.

Kuby T. 1999. *Innovation as a social process: What does this mean for impact assessment in agricultural research?* Paper presented at a CIAT workshop, Costa Rica, September 1999.

Leeuw FL. 2003. *Reconstructing program theories: methods available and problems to be solved.* American Journal of Evaluation, 24(1):5-20.

Mackay R, Horton D, Dupleich L and Andersen A. (2002). *Evaluating Organizational Capacity Development.* Canadian Journal of Program Evaluation, 17(2): 121-150.

Mark MM, Henry GT and Julnes G. 2000. *Evaluation: An integrated framework for understanding, guiding and improving policies and programs*. San Francisco: Jossey-Bass.

Mark MR. 2001. *What works and how can we tell?* Evaluation Seminar 2. Victoria Department of Natural Resources and Environment.

Marra, *M, 2004, 'Knowledge partnerships for development: what challenges for evaluation?' Evaluation and Program Planning* **27** (2004) 151–160 .

Matheson S (Ed). Encyclopedia of Evaluation. Thousand Oaks CA:Sage

Mayne J. 2001. *Addressing attribution through contribution analysis: Using performance measures sensibly.* Canadian Journal of Program Evaluation, 16(1): 1-24.

Mayne J. (2005). *Ensuring Quality for Evaluation: lessons from auditors*. Canadian Journal of Program Evaluation, 20(1): 37-64.

Mayne J and Rist RC. 2006. *Studies are not enough: the necessary transformation of evaluation.* Canadian Journal of Program Evaluation, 21(3): 93-120.

Menzies, P, 2001 'Counter factual theories of causation. *Stanford Encyclopedia of Philosophy* . Downloaded from http://plato.stanford.edu/entries/causation-counterfactual/ 7 January 2008.

Mitchell M and Egudo M. 2003. *A review of narrative methodology.* Published by DSTO Systems Sciences Laboratory, Department of Defence, Government of Australia.

Mohr LB. 1999. *The qualitative method of impact analysis*. American Journal of Evaluation 20(1):69-84.

Onwuegbuzie A and Johnson RB. 2006. *The validity issue in mixed research.* Research in the Schools 13(1) 2006:48-63.

Patton MQ. 2008. *Utilization-focused Evaluation.* Fourth Edition. Thousand Oaks, CA. Sage.

Pawson R and Tilley N. 1997. *Realistic Evaluation*. Thousand Oaks, CA: Sage. 235pp.

Pawson R. 2002. *Evidence-based Policy: In search of a method*. Evaluation 8(2): 157-181.

Pawson R. 2006. *Evidence-based policy. A realist perspective*. London: Sage. 196pp.

Pawson R. 2008. *Simple principles for the evaluation of complex programmes.* From Killoran A et al. *Public Health Evidence: Tackling health inequalities.* Oxford University Press. Forthcoming

Picciotto, R. and Wiesner, E. (eds.), *Evaluation and Development: The. institutional dimension.* New Brunswick, Transaction, 1997

Picciotto R. 2007. *The new environment for development evaluation.* American Journal of Evaluation 28(4): 509-521.

Rog D, Julnes G, Shadish W, Levin H and Datta, L-E. 2007. *Learning to promote quality over ideology for methodology.* Panel presentation at the 21st Annual Conference of the American Evaluation Association, Baltimore, USA.

Rogers PJ 2008 'Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions' *Evaluation* Vol 14(1): 29 – 48

Salmen L and Kane E. 2006. *Bridging Diversity. Participatory Learning for Responsive Development.* The World Bank, Washington DC, pp144.

Sandelowski et al. 2006. *Defining and designing mixed research synthesis studies*. Research in the Schools 13(1): 29-40.

Schorr LB. *Determining "what works" in social programs and social policies: Towards a more inclusive knowledge base"*. Published by The Brookings Institute.

Shadish W et al. 2005. *Evidence-based decision-making: Enhancing systematic reviews of program evaluation results in Europe.* Evaluation 11(1): 95-109.

Späth B. 2004. *Current state of the art in impact assessment: With a special view on small enterprise development.* Report for SDC.

Tang KC et al. *Evidence-based health promotion: recollections, reflections and considerations*. Journal of Epidemiology and Community Health 57: 11-16.

Tashakkori A and Teddlie C. 2003. *Handbook of mixed methods in social and behavioral research.* Thousand Oaks, CA: Sage.

Tashakkori A and Creswell JW. 2007. *Editorial: The new era of mixed methods*. Journal of Mixed Methods 1(1): 4.

Teddlie C and Tashakkori A. 2006. *A general typology of research designs featuring mixed methods.* Research in the Schools 13(1): 12-28.

Tetlock PE. (2005). *Expert Political Judgment: How good is it? How can we know?* Princeton (complete) University Press

Tukey JW. 1962. *The future of data analysis.* Annals of Mathematical Statistics 33(1): 1-67.

Turner, M. 2001. *Cognitive Dimensions of Social Sciences*, Oxford, Oxford University Press

Van den Berg R. 2006. *Results evaluation and impact assessment in development cooperation.* Evaluation 11(1): 27-36.

Weiss CH. 1997. *Theory-based evaluation: past, present, and future.* New Directions for Evaluation 7: 41-55.

Weiss CH. 2002. *"What to do until the random assigner comes".* In Mosteller & Boruch. Evidence Matters….

White H. 2002. *Combining quantitative and qualitative approaches in poverty analysis.* World Development 30(3): 511-522.

Williams B, Rogers PJ, McDonald B and Davidson EJ. 2006. *Draft Checklist for Impact Evaluation.*

World Bank paper. 2005. *Issues and Options for improving engagement between the World Bank and Civil Society Organisations.* Quoted in Salmen and Kane, 2006, p5.

World Health Organisation. 2004. *Promoting mental health.* Summary Report.

Yin RK. 2006. *Mixed methods research: Are the methods genuinely integrated or merely parallel?* Research in the Schools 13(1): 41-47.

ANNEX I:

Cases of Impact Evaluation Using the Approaches Highlighted

Introduction

This section presents impact evaluations conducted using the approaches and methods described in the section above. A summary is provided of how the evaluation addressed the 4 tasks of impact evaluation outlined under section X. These points are also elaborated as part of the case description.

The SG2 will continue to refine the cases. A considerable number of other cases being developed will also be included in the future versions of the document and subsequent to a review of the cases presented here..

Case 1: Ex-Post Impact Study of the Noakhali Rural Development Project in Bangladesh[8]

1. Summary

In terms of the 4 tasks of impact evaluation, this case study has the following features:

Identifying impacts that are valued – the evaluation examined the intended and unintended socio-economic impacts, with particular attention to the impact for women and to the sustainability and sustainment of these impacts

Gathering evidence of these impacts – the evaluation drew on a wide range of existing evidence and also used mixed methods to generate additional evidence; because the evaluation was conducted 9 years after the project had ended, it was possible to directly investigate the extent to which impacts had been sustained.

Assessing causal contribution – careful attention was paid to differential impacts in different contexts in order to interpret the significance of before/after and with/without comparisons; the intervention was only successful in contexts which provided the other necessary 'ingredients' for success

Managing the impact evaluation - the evaluation had significant resources and was preceded by considerable planning and review of existing evidence.

2. Summary of the intervention; its main characteristics

The Noakhali Rural Development Project (NRDP) was an Integrated Rural Development Projects (IRDP) in Bangladesh, funded for DKK 389 million by DANIDA. It was implemented in two phases over a period of 14 years, 1978-92, in the greater Noakhali district, one of the poorest regions of Bangladesh, which had a population of approximately 4 million. More than 60 long-term expatriate advisers – most of them Danish – worked 2-3 years each on the project together with a Bangladeshi staff of up to 1,000 (at the peak).

During NRDP-I the project comprised activities in 14 different areas grouped under four headings:

- Infrastructure (roads, canals, market places, public facilities)

[8] This case study is drawn from the 2002 report published by the Ministry of Foreign Affairs, Denmark.

- Agriculture (credit, cooperatives, irrigation, extension, marketing)
- Other productive activities (livestock, fish ponds, cottage industries)
- Social sector (health & family planning, education).

The overarching objective of NRDP-I was to "promote economic growth and social progress in particular aiming at the poorer sections of the population" (Danida, 1977(I):12). The poorer sections were to be reached in particular through the creation of temporary employment in construction activities (infrastructure) and engaging them in income generating activities (other productive activities). There was also an aim to create more employment in agriculture for landless labourers through intensification. Almost all the major activities started under NRDP-I continued under NRDP-II, albeit with some modifications and additions. The overarching objective was kept with one notable addition: "To promote economic growth and social progress in particular aiming at the poorer segments of the population including women" (Danida, 1984(1):i). A special focus on women was thus included, based on the experience that so far most of the benefits of the project had accrued to men.

3. The purpose, intended use and key evaluation questions

This ex-post impact study of the Noakhali Rural Development Project (NRDP) was carried out nine years after the project was terminated. At the time of implementation NRDP was one of the largest projects funded by Danida, and it was considered an excellent example of integrated rural development, which was a common type of support during the seventies and eighties. In spite of the potential lessons to be learned from the project, it was not evaluated upon completion in 1992. This fact and an interest in the sustainability factor in Danish development assistance led to the commission of the study. What type of impact could still be traced in Noakhali nine years after Danida terminated its support to the project?

While the study dealt with aspects of the project implementation, its main focus was on the project's socioeconomic impact in the Noakhali region. The study aimed to identify the intended as well as unintended impact of the project, in particular whether it had stimulated economic growth and social development and improved the livelihoods of the poor, including women, such as the project had set out to do.

The evaluation focused on the following questions:

- What has been the short- and long-term – intended as well as unintended – impactof the project?
- Has the project stimulated economic growth and social development in the area?
- Has the project contributed to improving the livelihoods of the poorest section ofthe population, including women?
- Have the institutional and capacity building activities engendered or reinforced by the project produced sustainable results?

4. Concise description of the evaluation, focusing on the approach, the rationale for the choice of approach and methods - linked to the four key tasks described in this document

Identifying impacts of interest

This study focuses on the impact of NRDP, in particular the long-term impact (i.e. nine years after). But impact cannot be understood in isolation from implementation and hence the study analyses various elements and problems in the way the project was designed and executed. Impact can also not be understood isolated from the context, both the natural/physical and in particular the societal – social, cultural, economic, political – context. In comparison with ordinary evaluations this study puts a lot more emphasis on understanding the national and in particular the local context.

Gathering evidence of impacts

One of the distinguishing features of this impact study, compared to normal evaluations, is the order and kind of fieldwork. The fieldwork lasted four months and involved a team of eight researchers (three European and five Bangladeshi) and 15 assistants. The researchers spent 11/2-31/2 months in the field, the assistants 2-4 months.

The following is a list of the methods used:

- Documentary study (project documents, research reports etc.)
- Archival work (in the Danish embassy, Dhaka)
- Questionnaire survey with former advisers and Danida staff members
- Stakeholder interviews (Danida staff, former advisers, Bangladeshi staff etc.)
- Quantitative analysis of project monitoring data
- Key informant interviews
- Compilation and analysis of material about context (statistics, articles, reports etc.)
- Institutional mapping (particularly NGOs in the area)
- Representative surveys of project components
- Assessment of buildings, roads and irrigation canals (function, maintenance etc.)
- Questionnaire-based interviews with beneficiaries and non-beneficiaries
- Extensive and intensive village studies (surveys, interviews etc.)
- Observation
- Focus group interviews
- In-depth interviews (issue-based and life stories).

A lot of effort was spent on studying the documentary evidence about the project. In the history of Danish development cooperation no other project has been subject to so many studies and reports, not to speak of the vast number of newspaper articles. Most important for the impact study have been the appraisal reports and the evaluations plus the final project completion report. But in addition to this there exists an enormous amount of reports on all aspects of the project. A catalogue from 1993 lists more than 1,500 reports produced by and for the NRDP (NRDP, 1993)! Both the project and the local context were, moreover, intensively studied in a research project carried out in cooperation between the Centre for Development Research (CDR) and Bangladesh Institute of Development Studies (BIDS).

A special effort was made to solicit the views of a number of key actors (or stakeholders) in the project and other key informants. This included numerous former NRDP and BRDB officers, expatriate former advisers as well as former key Danida staff, both based in the Danish Embassy in Dhaka and in the Ministry of Foreign Affairs in Copenhagen. They were asked about their views on strengths and weaknesses of the project and the components they know best, about their own involvement and about their judgement

regarding likely impact. A questionnaire survey was carried out among the around 60 former expatriate long-term advisers and 25 former key staff members in the Danish embassy, Danida and other key informants. In both cases about half returned the filled-in questionnaires. This was followed up by a number of individual interviews.

The main method in four of the five component studies was surveys with interviews, based on standardised questionnaires, with a random – or at least reasonably representative – sample of beneficiaries (of course combined with documentary evidence, key informant interviews etc.). A great deal of effort was taken in ensuring that the survey samples are reasonably representative.

The infrastructure component was studied by partly different methods, since in this case the beneficiaries were less well defined. It was decided to make a survey of all the buildings that were constructed during the first phase of the project in order to assess their current use, maintenance standard and benefits. In this phase the emphasis was on construction; in the second phase it shifted to maintenance. Moreover, a number of roads were selected for study, both of their current maintenance standard, their use etc., but also the employment the road construction and maintenance generated, particularly for groups of destitute women. The study also attempted to assess socio-economic impact of the roads on different groups (poor/better-off, men/women etc.).

Assessing causal contribution

The impact of a development intervention is a result of the interplay of the intervention and the context. It is the matching of what the project has to offer and people's needs and capabilities that produces the outcome and impact. Moreover, the development processes engendered unfold in a setting, which is often characterised by inequalities, structural constraints and power relations. This certainly has been the case in Noakhali. As a consequence there will be differential impact, varying between individuals and according to gender, socio-economic group and political leverage.

In addition to the documentary studies, interviews and questionnaire survey the actual fieldwork has employed a range of both quantitative and qualitative methods. The approach can be characterised as a contextualised, tailor-made ex-post impact study. There is considerable emphasis on uncovering elements of the societal context in which the project has been implemented. This covers both the national context and the local context. The approach is tailor-made in the sense that it will be made to fit the study design outlined above and apply an appropriate mix of methods.

An element in the method is the incorporation in the study of both before/after and with/without perspectives. These, however, are not seen as the ultimate test of impact (success or failure), but interpreted cautiously, bearing in mind that the area's development has also been influenced by a range of other factors (market forces, changing government policies, other development interventions etc.), both during the 14 years the project was implemented and during the nine years that have lapsed since its termination.

Considerable weight was accorded to studying what has happened in the villages that have previously been studied and for which some comparable data exist. Four villages were studied intensively in 1979 and briefly restudied in 1988 and 1994. These studies – together with a thorough restudy in the year 2001 – provide a unique opportunity to compare the situation before, during and after the project. Moreover, 10 villages were monitored under the project's 'Village-wise Impact Monitoring System' in the years 1988-90, some of these being 'with' (+NRDP) and some (largely) 'without' (-NRDP) the project. Analysis of the monitoring data combined with a restudy of a sample of these villages illuminates the impact of the project in relation to other factors. It was decided to study a total of 16 villages, 3 intensively (all +NRDP, about 3 weeks each), 12 extensively (9 +NRDP, 3 –NRDP, 3-5 days each). As a matter of principle, this part of the study looks at impact in terms of the project as a whole. It brings in focus the project benefits as perceived by different groups and individuals and tries to study how the project has impinged on economic and social processes of development and change. At the same time it provides a picture of the considerable variety found in the local context.

In the evaluation of the Mass Education Program, the problem of attribution was dealt with as carefully as possible. Firstly, a parallel comparison has been made between the MEP beneficiaries on the one hand and non-beneficiaries on the other, in order to identify (if any) the changes directly or indirectly related to the programme. Such comparison was vital due to the absence of any reliable and comparable baseline data. Secondly, specific queries were made in relation to the impact of the programme as perceived by the beneficiaries and other stakeholders of MEP, assuming that they would be able to perceive the impact of the programme intervention on their own lives in a way that would not be possible for others. And finally, views of non-beneficiaries and non-stakeholders were sought in order to have less affected opinion from people who do not have any valid reason for either understating or overstating the impact of MEP. It was through such a cautious approach that the question of attribution was addressed. Arguably, elements of subjectivity may still have remained in the conclusions and assumptions, but that is unavoidable in a study that seeks to uncover the impact of an education project.

Managing the impact evaluation

The impact study was commissioned by Danida and carried out by Centre for Development Research, who also co-funded the study as a component of its Aid Impact Research Programme. The research team comprised independent researchers from Bangladesh, Denmark and the UK. A reference group of nine persons (former advisers, Danida officers and researchers) followed the study from the beginning to the end. It discussed the approach paper in an initial meeting and the draft reports in a final meeting. In between it received three progress reports from the team leader and took up discussions by e-mail correspondence. The study was prepared during the year 2000 and fieldwork carried out in the period January-May 2001. The study consists of a main report and 7 topical reports.

The first step in establishing a study design was the elaboration of an approach paper (study outline) by the team leader. This was followed by a two weeks' reconnaissance visit to Dhaka and the greater Noakhali area. During this visit Bangladeshi researchers and assistants were recruited to the team, and more detailed plans for the subsequent fieldwork were drafted. Moreover, a background paper by Hasnat Abdul Hye, former Director General of BRDB and Secretary, Ministry of Local Government, was commissioned. The paper was later brought out under the title 'Comparative Perspectives on the Noakhali Rural Development Project' (Hye, 2000).

The fieldwork was preceded by a two days methodology-cum-planning workshop in Dhaka. The actual fieldwork lasted four months – from mid-January to mid-May 2001. The study team comprised 23 persons, five Bangladeshi researchers (two men, three women), three European researchers (two men, one woman), six research assistants (all men) and nine field assistants (including two women, all from Bangladesh). The researchers spent 11/2-31/2 months in the field, the assistants 2-4 months. Most of the time the team worked 60-70 hours a week. So it takes a good deal of resources to accomplish such a big and complex impact study.

Formatted: Bullets and Numbering

1. Case 2: Mixed method impact evaluation of IFAD projects in Gambia, Ghana and Morocco

1. Summary

In terms of the 4 tasks of impact evaluation, this case study has the following features:

Identifying impacts that are valued – the evaluation included intended and unintended impacts, and examined the magnitude, coverage and targeting of changes

Gathering evidence of these impacts –the evaluation used mixed methods to gather evidence of impacts and the quality of processes with cross-checking between sources

Assessing causal contribution –no baseline data were available. Instead a comparison group was constructed, and analysis of other contributing factors was made to ensure appropriate comparisons

Managing the impact evaluation – the evaluation was undertaken within significant resource constraints, using an interdisciplinary team.

2. Introduction and Background

Evaluations of rural development projects and country programmes are routinely conducted by the Office of Evaluation of IFAD. The ultimate objectives of these evaluations is to (i) set a basis for accountability by assessing the development results and (ii) contribute to learning and improvement of design and implementation by providing lessons learned and practical recommendations. These evaluations follow a standardised methodology and a set of evaluation questions including the following: (i) project performance (relevance, effectiveness, and efficiency), (ii) project impact, (iii) overarching factors (sustainability, innovation and replication) and (iv) the performance of the partners. As can be seen, impact is but one the key evaluation questions and the resources allocated to the evaluation (budget, specialists and time) have to be shared for the entirety of the evaluation question.

As such, these evaluations are to be conducted under resource constraints. In addition, very limited data are available on socio-economic changes taking place in the project area that can be ascribed to an impact definition. IFAD adopts an impact definition which is similar to the DAC standard.[9] The key feature of IFAD evaluation is that they are conducted just before or immediately after project conclusion: the effects can be observed after 4-7 years of operations and the future evolution can be estimated through an educated guess on sustainability perspectives. Several impact domains are considered including: (i) household income and assets, (ii) human capital, (iii) social capital, (iv) food security, (v) environment and (vi) institutions.

3. Sequencing of the process and choice of methods

This short case study is based on evaluations conducted in Gambia, Ghana and Morocco between 2004 and 2006. As explained above, evaluations had multiple questions to answer and impact assessment was but one of them. Moreover, impact domains were quite diverse. This meant that some questions and domains required quantitative evidence (for example in the case of household income and assets) while a more qualitative assessment would be in order for other domains (for example social capital). In many instances, however, more than one method would have to be used to answer the same questions, in order to cross-

[9] DAC definition: Positive and negative, primary and secondary long-term effects to which a project has contributed, directly or indirectly"

check the validity of findings, identify discrepancy and formulate hypotheses on the explanation of apparent inconsistencies.

As the final objective of the evaluation was not only to assess results but also provide future intervention designers with adequate knowledge and insights, the evaluation design could not be confined to addressing a dichotomy between "significant impact has been observed" and "no significant impact has been observed". Findings would need to be rich enough and grounded in field experience in order to provide a plausible explanation that would lead, when suitable, to a solution to identified problems and to recommendations to improve the design and the execution of the operations.

Countries and projects considered in this case study were ostensibly diverse. In all cases however, the first step in the evaluation consisted of a desk review of the project documentation. This allowed the evaluation team to understand or reconstruct the intervention theory (often implicit) and the logical framework. The intervention theory or logical framework would help identify a set of hypotheses on changes that may be observed in the field as well as on intermediary steps that would lead to those changes.

In particular, our preliminary desk analysis highlighted that our results assessment would have to be supplemented with some analysis of implementation performance. The latter would include some insight in the business processes (for example the management and resource allocation made by the project implementation unit) and the quality of service rendered (for example the topics and the communication quality of an extension service or the construction quality of a feeder road or of a drinking water scheme).

The second step was to conduct a preparatory mission. Among other purposes, this mission was instrumental in fine tuning our hypotheses on project results and designing the methods and instruments. Given the special emphasis of the IFAD interventions on the rural poor, impact evaluation would need to shed light, to the extent possible, on the following dimensions of impact: (i) magnitude of changes, (ii) coverage (that is the number of persons or households served by the projects and (iii) targeting, that is gauging the distribution of project benefits according to social, ethnic or gender grouping.

As pointed out before, a major concern was the absence of a baseline survey which could be used as a reference for impact assessment. This required reconstructing the "before project" situation. By the same token, it was clear that the observed results could not simply attributed to the evaluated interventions. In addition to exogenous factors such as weather changes, other important factors were at play, for example changes in government strategies and policies (such as the increased support to grassroots associations by Moroccan public agencies) or operations supported by other development organisations in the same or in adjacent zones. This meant that the evaluated interventions would interplay with existing dynamics and interact with other interventions. Understanding synergies or conflicts between parallel dynamics could not be done simply through inferential statistic instruments but required interaction with a wider range of stakeholders.

The third step in the process was the fielding of a data collection survey (after pre-testing the instruments) which would help the evaluation cope with the dearth of impact data. The selected techniques for data collection included: (i) a quantitative survey with a range of 200 – 300 households (including both project and control groups) and (ii) a more reduced set of focus group discussion with groups of project users and "control groups" stratified based on the economic activities in which they had engaged and the area where they were leaving.

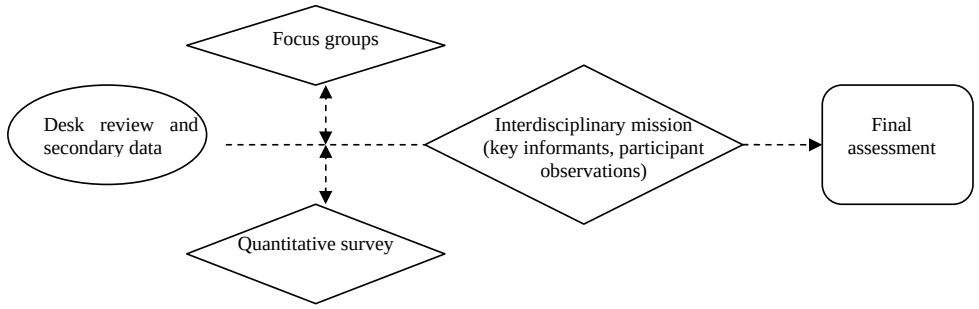
In the quantitative survey standardised questionnaires were administered to final project users (mostly farmers or herders) as well as to non-project groups (control observations) on the situation before (recall methods) and after the project. Recall methods were adopted to make up for the absence of a baseline.

In the course of focus group interviews, open-ended discussion guidelines were adopted and results were mostly of qualitative nature. Some of the focus group facilitators had also been involved in the quantitative survey and could refer the discussion to observations previously made. After the completion of data collection and analysis, a first cross-checking of results could be made between the results of quantitative and qualitative analysis.

As a fourth step, an interdisciplinary evaluation team would be fielded. Results from the preliminary data collection exercise were made available to the evaluation team. The data collection coordinator was a member of the evaluation team or in a position to advise its members. The evaluation would conduct field visits and conduct a further validate survey and focus group data through participant observations and interviews with key informants (and further focus group discussions if necessary). The team would also spend adequate time with project management units in order to gather a better insight of implementation and business processes.

The final impact assessment would be made by means of triangulation of evidence captured from the (scarce) existing documentation, the preliminary data collection exercise and the main interdisciplinary mission (Figure 1).

Figure 1



4. Constraints in data gathering and analysis

Threats to the validity of recall methods. According to the available literature sources[10] and our own experience, the reliability of recall methods may be questionable for monetary indicators (e.g. income) but higher for easier-to-remember facts (e.g. household appliances, approximate herd size). Focus group discussions helped identify possible sources of bias in the quantitative survey and ways to address them.

Finding "equivalent" samples for with and without-project observations. One of the challenges was to extract a control sample that would be "similar" in the salient characteristics to the project sample. In other words, problems of sampling bias and endogeneity should have been controlled for (e.g. more entrepreneurial people are more likely to participate in a rural finance intervention). In sampling control observations serious attempts were made to match project and non-project households based on similarity of main economic activities, agro-ecological environment, household size and resource endowment. In some instances, household that had just started to be served by the projects ("new entries") were considered as control groups, on the grounds that they would broadly satisfy the same eligibility criteria at entry as "older" project clients. However, no statistical technique (e.g. instrumental variables, Heckman's procedure or propensity score) was adopted to test for sampling bias, due to limited time and resources.

Coping with linguistic gaps. Given the broad scope of the evaluations (section A), a team of international sector specialists was required. However, international experts were not necessarily the best suited for data collection analysis which calls for fluency in the local vernacular, knowledge of local practices and skills to obtain the most possible information within a limited time frame. Staggering the process in several phases

[10] Typical problems with recall methods are that of: (i) incorrect recalling and (ii) telescoping, i.e. projecting backward or forward an event: for example the purchase of a durable good which took place 7 years ago (before the project started) could be projected to four years ago, during project implementation. See, for example, Bamberger, M., Rugh, J., Church, M. and Fort L. (2004): Shoestring Evaluation: Designing Impact Evaluations under Budget, Time and Data Constraints, *American Journal of Evaluation, 3 2004*; vol. 25: pp. 5 - 37

was a viable solution. The preliminary data collection exercise was conducted by a team of local specialists, with university students or local teachers or literate nurses serving as enumerators.

5. Main value added of mixed methods and opportunities for improvement

The choice of methods was made taking into account the objectives of the evaluations and the resource constraints (time, budget and expertise) in conducting the exercise. The combination of multiple methods allowed us to cross-check the evidence and understand, for example, when survey questions were likely to be misinterpreted or generate over or under-reporting. On the other hand, quantitative evidence allowed us to shed light on the prevalence of certain phenomena highlighted during the focus group discussion. Finally the interactions with key informants and project managers and staff helped us better understand the reasons for under or over-achievements and come up with more practical recommendations.

The findings, together with the main conclusions and recommendations in the report were adopted in order to design new projects or a new country strategy. Also there was interest from the concerned project implementation agencies to adopt the format of the survey in order to conduct future impact assessments on their own.

Due to time constraints, only inferential analysis was conducted on the quantitative survey data. A fully-fledged econometric analysis would have been desirable. By the same token, further analysis of focus group discussion outcomes would be desirable in principle.

6. A few highlights on the management of the process.

The overall process design, as well as the choice of methods and the design of the data collection instruments was made by the Lead Evaluator in the Office of Evaluation of IFAD, in consultation with international sectoral specialists and the local survey coordinator.

The pre-mission data collection exercise was coordinated by a local rural sociologist, with the help of a statistician for the design of the sampling framework and data analysis.

Time required conducting the survey and focus groups:

- Develop draft questionnaire and sampling frame, identify enumerators: 3 weeks
- Conduct a quick trip on the ground, contact project authorities and pre-test questionnaires: 3 days
- Train enumerators' and coders' team: 3 days
- Survey administering: depending on the length of the questionnaire, on average an enumerator will be able to fill no more than 3-5 questionnaires per day. In addition time needs to be allowed for travel, rest. With a team of 6 enumerators, in 9-10 working days up to 200 questionnaires can be filled in, in the absence of major transportation problems.
- Data coding: it may vary depending on the length and complexity of the questionnaire. It is safe to assume 5-7 days.
- Time for conducting focus groups discussions: 7 days based on the hypothesis that around 10 FGD would be conducted by 2 teams.
- Data analysis. Depending on the analysis requirement it will require 1-2 weeks only to generate the tables and summary of focus group discussions
- Drafting survey report: 2 weeks

Note: as some of the above tasks can be conducted simultaneously, the total time for conducting a preliminary data collection exercise may be lower than the sum of its parts.

Case 3: Systemic Impact Evaluation: Agricultural Development Projects in Guinea

1. Summary

In terms of the 4 tasks of impact evaluation, this case study has the following features:

Identifying impacts that are valued – the evauation focused on impact in terms of poverty alleviation ; the distribution of benefits was of particular interest, not just the mean effect.

Gathering evidence of these impacts –.all data gathering was conducted after the intervention had been completed. ; mixed methods were used, including attention to describing the different implementation contexts.

Assessing causal contribution – this is the major focus of the case study. A counter-factual was created by creating a comparison group, taking into account the endogenous and exogenous factors affecting impacts. Modelling was used to develop an estimate of the impact.

Managing the impact evaluation – the evaluation was undertaken as part of a PhD and appears to have little stakeholder engagement and subsequent use.

This impact evaluation concerned two types of agricultural projects based in the kpèlè region, in Guinea. The first one[11] was the Guinean Oil Palms and Rubber Company (SOGUIPAH). It was founded in 1987 by the Guinean government to take charge of developing palm oil and rubber production at the national level. With the support of several donors, SOGUIPAH quickly set up a program of industrial plantations[12] by negotiating the ownership of 22,830 ha with villagers. In addition, several successive programs were implemented between 1989 and 1998 with SOGUIPAH to establish contractual plantations[13] on farmers' own land and at the request of the farmers (1552 ha of palm trees and 1396 ha of rubber trees) and to improve 1093 ha of lowland areas for irrigated rice production.

The impact evaluation took place in a context of policy debates between different rural stakeholders at a regional level: two seminars had been held in 2002 and 2003 between the farmers' syndicates, the state administration, private sector and development partners (donors, NGOs) to discuss a regional strategy for agricultural development. These two seminars revealed that there was little evidence of what should be done to alleviate rural poverty, despite a long history of development projects. The impact of these projects on farmers' income seemed to be particularly relevant to assess, notably in order to compare the projects efficiency.

This question was investigated through a PhD thesis which was entirely managed by the AGROPARISTECH[14]. It was financed by AFD, one of the main donor in the rural sector in Guinea.

This thesis proposed a new method, the systemic impact evaluation, aiming at quantifying impact using a qualitative approach. It enables to understand both the process through which impact materializes and to rigorously quantify the impact of agricultural development projects on the farmers' income, using a counterfactual. The analysis is notably based on the comprehension of the agrarian dynamics and the

[11] The second project was inland valley development for irrigated rice cultivation and will not be presented here.

[12] Industrial plantations are the property of SOGUIPAH and are worked by salaried employees.

[13] A contract between SOGUIPAH and the farmer binds the farmer to reimburse the cost of the plantation and deliver his production to SOGUIPAH.

[14] AGROPARISTECH is a member of the Paris Institute of Technology which is a consortium of 10 of the foremost French Graduate Institutes in Science and Engineering. AGROPARISTECH is a leader Institute in Life Sciences and Engineering.

farmers' strategies, and permits the quantification of ex-post impact but also to devise a model of ex- ante evolution for the following years.

2. Gathering evidence of impact

The data collection was carried out entirely ex post. Several types of surveys and interviews were used to collect evidences of impact.

First, a contextual analysis realized all along the research work with key informants was necessary to describe the project implementation scheme, the contemporaneous events and the existing agrarian dynamics. It was also used to assess qualitatively whether those dynamics were attributable or not to the project. A series of surveys and historical interviews (focused on the pre-project situation) were notably conducted in order to establish the most reliable baseline possible. An area considered "witness" to the agrarian dynamic that would have existed in the project's absence was identified.

Second, a preliminary structured survey (on about 240 households) was implemented, using recall to collect data on the farmers' situation in the pre-intervention period and during the project. It was the basis of a judgment sample to realize in depth interviews (see bellow), which aimed at describing the farming systems and quantifying rigorously the farmers' income.

3. Assessing causal attribution

By conducting an early contextual analysis, the evaluator was able to identify a typology of farming systems which existed before the project. In order to set up a sound counterfactual, a judgment sample was realized amongst the 240 households surveyed, by choosing 100 production units which had belonged to the same initial types of farming system and which had evolved with (in the project area) or without the project (in the witness area).

In-depth understanding of the endogenous and exogenous factors influencing the evolution and possible trajectories of farming systems enabled the evaluator to rigorously identify the individuals whose evolution with or without the project were comparable.

This phase of judgment sample was followed by in-depth interviews with the hundred farmers. The evaluator's direct involvement in data collection was then essential, hence the importance of a small sample. It would not have been possible to gather reliable data on yields, modifications to production structures over time and producers' strategies from a large survey sample in a rural context.

Then, based on the understanding of the way the project proceeded and of the trajectories of these farmers, with or without the project, it was possible to build a quantitative model, based on Gittinger's method of economic analysis of development projects (Gittinger, 1982). As the initial diversity of production units was well identified before sampling, this model was constructed for each type of farming system existing before the project. Understanding the possible evolutions for each farming system with and without the project allowed for the estimation of the differential created by the project on farmers' income, that is its impact.

4. Ensuring rigor and quality

Although the objective differences between each production unit studied appear to leave room for the researcher's subjectivity when constructing the typology and sample, the rationale behind the farming system concept made it possible to transcend this possible arbitrariness. What underlies this methodological jump from a small number of interviews to a model is the demonstration that a finite number of types of farming systems exists in reality.

Moreover, (i) the use of a comparison group, (ii) the triangulation of most data collected by in-depth interviews through direct observation and contextual analysis and (iii) the constant implication of the principal researcher, were key factors to ensure rigor and quality.

5. Key findings

The large survey realized by interviewers on 240 households allowed identifying 11 trajectories related to the implementation of the project. Once each trajectory and each impact was characterized and quantified through in-depth interviews and modeling, this survey permitted as well quantifying a mean impact of the project, on the basis of the weight of each type in the population. The mean impact was only 24 €/year/household in one village poorly served by the project, due to its enclosed situation, whereas it was 200 €/year/household in a central village.

Despite a positive mean impact there were also highly differentiated impacts that existed, depending on the original farming system and the various trajectories with and without the project, which could not be ignored. Whereas former civil servants or traditional landlords beneficiated large contractual plantations, other villagers were deprived of their land for the needs of the project or received surfaces of plantations too limited to improve their economic situation.

Therefore, it seems important that the impact evaluation of a complex development project includes an analysis of the diversity of cases created by the intervention, directly or indirectly.

The primary interest of this new method was to give the opportunity to build a credible impact assessment entirely ex post. Second, it gave an estimate of the impact on different types of farming systems, making explicit the existing inequalities in the distribution of the projects' benefits. Third, it permitted a subtle understanding of the reasons why the desired impacts materialized or not.

6. Influence

The results from this first impact assessment were available after four years of field work and data treatment. They were presented to the Guinean authorities and to the local representatives of the main donors in the rural sector. In the field, the results were delivered to the local communities interviewed and to the farmers' syndicates. The Minister of Agriculture declared that he would try to foster more impact evaluations on agricultural development projects. Unfortunately, there is little hope that the conclusions of this research will change the national policy about these types of projects, in the absence of an institutionalized forum for discussing it between the different stakeholders.

Case 4: GEF Impact Evaluation 2007[15]

Evaluation of Three GEF Protected Area Projects in East Africa

1. Description of Evaluation

The objectives of this evaluation included:

- To *test evaluation methodologies* that can assess the impact of GEF interventions. The key activity of the GEF is "providing new and additional grant and concessional funding to meet the agreed incremental costs of measures to achieve agreed global environmental benefits."[16] The emphasis of this evaluation was therefore on verifying the achievement of agreed global environmental benefits.
- Specifically, to test a *Theory of Change approach* to evaluation in GEF's biodiversity focal area, and assess its potential for broader application within GEF evaluations.
- To assess the *sustainability and replication of the benefits of GEF support*, and extract lessons. It evaluated whether and how project benefits have continued, and will continue, after project closure.

Primary users

The primary users of the evaluation are GEF entities. They include: the GEF Council, which requested the evaluation; GEF Secretariat, which will approve future protected area projects, Implementing Agencies (such as the World Bank, UN agencies and regional Development Banks) and national stakeholders who will implement future protected area projects.

2. Evaluation design

Factors driving selection of evaluation design

The Approach Paper to the Impact Evaluation[17] considered the overall GEF portfolio in order to develop an entry-point which could provide a good opportunity to develop and refine effective and implementable impact evaluation methodologies. Themes and projects that are relatively straightforward to evaluate were emphasized. The EO adopted the DAC definition of impact, which determined that closed projects would be evaluated to assess the sustainability of GEF interventions.

Biodiversity and protected areas:

The biodiversity focal area has the largest number of projects within the GEF portfolio of currently active and completed projects. In addition, biodiversity has developed more environmental indicators and global data sets than other focal areas, both within the GEF and in the broader international arena. The Evaluation Office chose protected areas as the central theme for this phase of the Impact Evaluation because: protected areas are one of the primary approaches supported by the GEF biodiversity focal area and its implementing agencies, and the GEF is the largest supporter of protected areas globally; previous evaluations have noted that an evaluation of the GEF support for protected areas has not been carried out, and recommended that such a study be undertaken; protected areas are based on a set of explicit change theories, not just in the

[15] The GEF Evaluation Office section of the GEF website contains the 11 papers produced by the Impact Evaluation 2007, under the heading of "ongoing evaluations."
[16] Instrument for the Establishment of the Restructured Global Environment Facility
[17] GEF EO, "Approach Paper to Impact Evaluation", February 2006.

GEF, but in the broader conservation community; in many protected area projects, substantial field research has been undertaken, and some have usable baseline data on key factors to be changed by the intervention ; a protected areas strategy can be addressed at both a thematic and regional cluster level (as in East Africa, the region chosen for the study); the biodiversity focal area team has made considerable progress in identifying appropriate indicators for protected areas through its "Managing for Results" system.

The choice of projects

Lessons from a set of related interventions (or projects) are more compelling than those from an isolated study of an individual project. To test the potential for aggregation of project results, enable comparisons across projects and ease logistics, it was decided to adopt a sub-regional focus and select a set of projects that are geographically close to each other. East Africa is the sub-region with the largest number of complete and active projects in the GEF portfolio with a protected area component, utilizing large GEF and cofinancing expenditure.

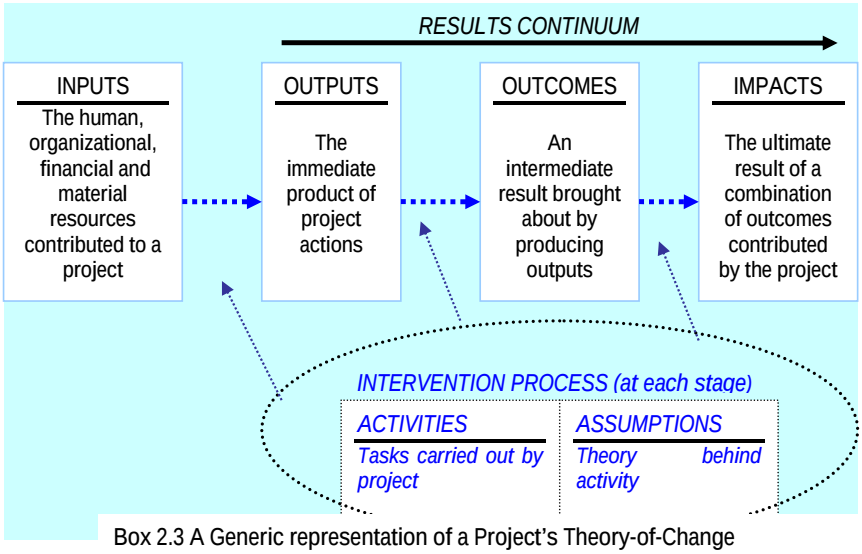
The following three projects were selected for evaluation:

- *Bwindi Impenetrable National Park and Mgahinga Gorilla National Park Conservation Project*, Uganda (World Bank)
- *Lewa Wildlife Conservanc*y, Kenya (World Bank)
- *Reducing Biodiversity Loss at Cross-Border Sites in East Africa*, Regional: Kenya, Tanzania, Uganda (UNDP).

These projects were implemented on behalf of the GEF by the World Bank and UNDP. They have a variety of biodiversity targets, some of which are relatively easy to monitor (gorillas, zebras, rhinos). Also, these projects were evaluated positively by terminal and other evaluations and the continuance of long term results was predicted. The *Bwindi Impenetrable National Park and Mgahinga Gorilla National Park Conservation Project* is a $6.7 million full-size-project and the first GEF-sponsored trust fund in Africa. The *Lewa Wildlife Conservancy* is a medium-sized-project, within a private wildlife conservation company. The *Reducing Biodiversity Loss at Cross-Border Sites in East Africa* Cross project is a $12 million project, implemented at field level by Government agencies, that aims to foster an enabling environment for the sustainable use of biodiversity.

The advantages of a Theory of Change approach

An intervention generally consists of several complementary activities that together produce intermediate outcomes, which are then expected to lead to impact (see **Box 2.3**). The process of these interventions, in a given context, is determined by the contribution of a



Box 2.3 A Generic representation of a Project's Theory-of-Change

variety of actions at multiple *levels*, some of which are

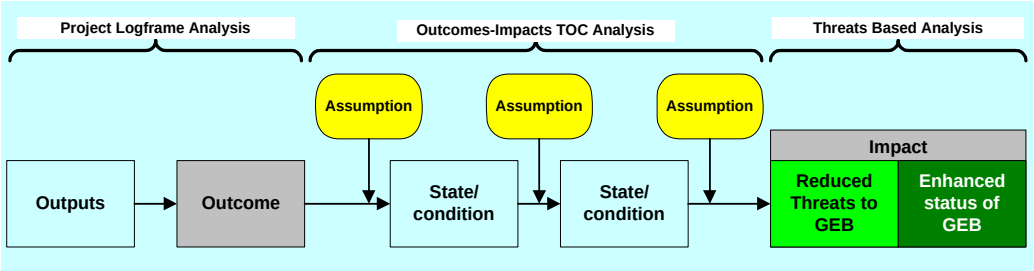
outside the purview of the intervention (e.g. actions of exterior actors at the local, national or global levels or, change in political situations, regional conflicts and natural disasters). Subsequently, an intervention may have different levels of achievement in its component parts, giving mixed results towards its objectives.

Experimental Approaches, which have recently gained prominence in the international debate on development evaluation, were not found appropriate for the purpose of this evaluation. While each GEF intervention using protected area strategies encompasses a variety of activities; it involves context-specific interactions (leading to ecological changes in the case of global environmental benefits), often on a scale that is not amenable to being simulated at another site. Whilst the Counterfactual-Based Approach may sometimes provide rigorous estimates of the level of effect a specific activity has had, it is unable to *deconstruct the changes a set of activities*, or a project brings about; and it provides few insights into the reasons behind the identified changes (the why and how). Finally, given the dynamic inter-play between actors and activities within and outside the project, the Evaluation Office focuses largely on estimating the 'contribution' of GEF activities towards bringing change, rather than attempting to attribute a specific level of global environmental change to GEF interventions.

The Evaluation Office therefore decided to adopt a Theory-based Approach (or Theory-of-Change Approach), combining a range of qualitative and quantitative methods. This approach focused on measuring the success of intermediate activities and relating them to impacts observed at the global level (using global and local environment databases).

The use of a hybrid evaluation model:

During the process of field-testing, it was decided that, given the intensive data requirements of a theory-of-change approach and the intention to examine project impacts, **the evaluation would mainly focus on the later elements of each project's theory-of-change, when outcomes are expected to lead to impact**. Based on this approach, the evaluation developed a methodology composed of three components (see **Box 2.4**):



Box 2.4 Components of Impact Evaluation Framework

- **Assessing implementation success and failure**: To understand the contributions of the project at earlier stages of the results continuum, leading to project outputs and outcomes, a *Logframe analysis* is used. Though the normally complex and iterative process of project implementation is not captured by this method, the Logframe provides a means of tracing the realization of declared objectives. GEF interventions aim to "assist in the protection of the global environment and promote thereby environmentally sound and sustainable economic development."[18]

[18] See the Preamble, "Instrument for the Establishment of the Restructured Global Environment Facility"

- **Assessing the level of contribution (i.e. impact)**: To provide a direct measure of project impacts, a *Targets-Threats analysis (Threats Based Analysis)* is used to determine whether global environmental benefits have actually been produced and safeguarded.[19] The robustness of global environment benefits identified for each project (or 'targets') is evaluated by collecting information on attributes relating to the targets' biological composition, environmental requirements and ecological interactions. This analysis of targets is complemented by an assessment of the level of 'threat' (e.g., predation, stakeholder attitude and behavior) faced by the global environment benefits. For targets and significant threats, trends over time (at project start, at project close, and currently) and across project and non-project areas are sought, so that a comparison is available to assess levels of change.
- **Explanations for observed impact**: To unpack the processes by which the project addresses and contributes to impact, an *Outcomes-Impacts Theory-of-Change analysis* is used. This theory-of-change approach constructs and validates the project logic connecting outcomes and ultimate project impact. It involves a comprehensive assessment of the activities undertaken after project closure, along with their explicit and implicit assumptions. This component enables an assessment of the sustainability and/or catalytic nature of project interventions, and provides a composite qualitative ranking for the achievements of the projects. Elements of the varied aspects of sustainability include behavior change and the effectiveness of capacity building activities, financial mechanisms, legislative change and institutional development.

The model incorporates three different elements that it is suggested are involved in the transformation of project outcomes into impacts. These are as follows, and were each scored for the level of achievement of the project in converting outcomes into impacts.

- Intermediary States. These are conditions that are expected to be produced on the way to delivering the intended impacts.
- Impact Drivers. These are significant factors or conditions that are expected to contribute to the ultimate realization of project impacts. Existence of the Impact Driver in relation to the project being assessed suggests that there is a good likelihood that the intended project impact will have been achieved. Absence of these suggests that the intended impact may not have occurred, or may be diminished.
- External Assumptions. These are potential events or changes in the project environment that would negatively or positively affect the ability of a project outcome to lead to the intended impact, but that are largely beyond the power

3. Data Collection and Constraints:

- *Logical Framework and Theory of Change Model:*

The approach built on existing project logical frameworks, implying that a significant part of the Framework could be relatively easily tested through an examination of existing project documentation, terminal evaluation reports and, where available, monitoring data. Where necessary, targeted consultations and additional studies were carried out.

- *Assessing conservation status and threats to Global Environment Benefits:*

A data collection framework for assessing the status of the targets and associated threats was developed, identifying indicators for each, along with the potential sources of information. For the Bwindi and Lewa projects, the task of collecting and assessing this information was undertaken by scientists from the Institute

[19] This is based on Nature Conservancy's 'Conservation Action Planning' methodology

of Tropical Forest Conservation, headquartered in Bwindi Impenetrable National Park, and the Lewa Research Department respectively. For the Cross Borders project, this exercise was done by Conservation Development Center, based on the existing project documentation, a field visit to the project site and consultations with key informants. The objective of this exercise was to provide quantitative measures for each indicator from **before the project** (baseline), at the **project close**, and **present day**. Where quantitative data were not available, detailed qualitative data were collected.

Improving rigor

- *Internal validity*

The evaluation used a participatory approach with substantial involvement of former project staff in drawing out theories-of-change and subsequently providing data for verification. These data were verified by local independent consultants, via a process of triangulating information from project documentation and external sources. Given that all three projects are now closed, the participation from former project staff enabled a candid and detailed exchange of information (during Workshops in Uganda and Kenya). The participants in return found the process to be empowering, as it clarified and supported the rationale for their actions (by drawing out the logical connections between activities, goals and assumptions) and also enabled them to plan for future interventions.

- *External validity*

Given the small number of projects, their variety and age (approved in varied past GEF replenishment phases), the evaluation did not expect to produce findings, which could be directly aggregated. Nevertheless, given the very detailed analysis of the interventions a few years after project-closure, it did provide a wealth of insights into the functioning of protected area projects, particularly elements of their sustainability after project closure. This allowed limited generalization on key factors associated with achievement of impact, on the basis of different levels of results related to a set of common linkages in the theoretical models. On this basis, the Evaluation Office recommended that the GEF Secretariat ensure specific monitoring of progress towards institutional continuity of protected areas throughout the life of a project.

4. Weaknesses

Impact evaluations are generally acknowledged to be highly challenging. The objective of this particular study, of examining GEF's impact at a 'global' level in biodiversity, make the study particularly complex. A few concerns include:

- The nature of changes in biodiversity is still under debate. Such changes are often non-linear, with uncertain time-scales even in the short-run, interactions within and across species, and exogenous factors (like climate change). Evidence regarding the achievement of global environment benefits and their sustainability must therefore be presented with numerous caveats.
- Numerous explanations and assumptions may be identified for each activity that is carried out. One frequent complication, during the construction of the Theory of Change, is the justification for one chosen theory to the exclusion of others.
- The approach may not always uncover unexpected outcomes or synergies, unless they are anticipated in the theories or assumptions of the evaluation team.[20] However, fieldwork should be able to discern such outcomes, as was the case in the Bwindi case study, which produced evidence of a number of unexpected negative impacts on local indigenous people.

[20] Sanderson, I. (2000) '"Evaluation in Complex Policy Systems", *Evaluation* 6(4): 433–54.

- The association between activities and outcomes in the Theory of Change approach depends on measuring the level of activities carried out, and then consciously (logically) linking them with impact through a chain of intermediate linkages and outcomes. Information on these intermediate outcomes may be difficult to obtain, unless former project implementers participate fully in the evaluation.

5. Concluding thoughts on the evaluation approach

For biodiversity, GEF's first strategic priority is "*Catalyzing Sustainability of Protected Area Systems"*, which aims for an expected impact whereby "biodiversity [is] conserved and sustainably used in protected area systems."

The advantage of the hybrid evaluation model used was that by focusing towards the end of the results-chain, it examined the combination of mechanisms in place that have led to a project's impacts and ensure sustainability of results. It is during this later stage, after project closure, that the ecological, financial, political, socio-economic and institutional sustainability of the project are tested, along with its catalytic effects. During project conceptualization, given the discounting of time, links from outcome to impact are often weak. Once a project closes, the role of actors, activities and resources is often unclear; this evaluation highlighted these links and assumptions.

Adopting a Theory of Change approach also had the potential to provide a mechanism that helped GEF understand what has worked and what has not worked and allows for predictions regarding the probability of success for similar projects. The Evaluation Office team concluded that the most effective combination for its next round of impact evaluation (Phase-out of Ozone Depleting Substances in Eastern Europe) should seek to combine Theory of Change approaches with opportunistic use of existing data sets, which might provide some level of quantifiable counterfactual information.

Case 5: Impact of Lewa Wildlife Conservancy (Kenya)[21]

Context

The Lewa GEF Medium-Sized Project provided support for the further development of Lewa Wildlife Conservancy ("Lewa"), a not-for-profit private wildlife conservation company that operates on 62,000 acres of land in Meru District, Kenya. The GEF awarded Lewa a grant of $0.75 million for the period 2000 to the end of 2003, with co-financing amounting to $3.193 million.

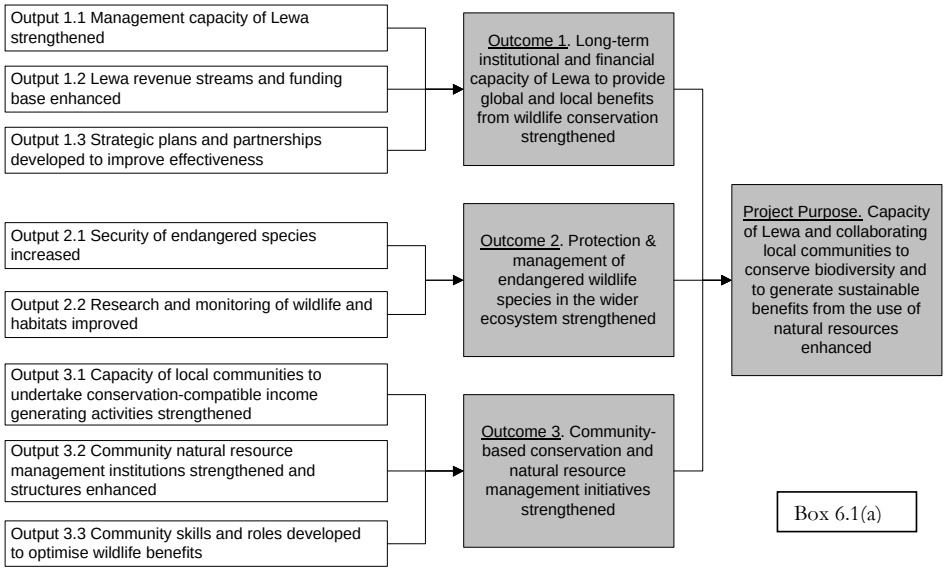
Since the GEF grant, Lewa has been instrumental in initiating the formation of the Northern Rangelands Trust (NRT) in 2004. NRT is an umbrella local organization with a goal of collectively developing strong community-led institutions as a foundation for investment in community development and wildlife conservation in the Northern Rangelands of Kenya. The NRT membership comprises community conservation conservancies and trusts, local county councils, the Kenya Wildlife Service, the private sector, and NGOs established and working within the broader ecosystem. The establishment and functioning of the NRT has, therefore, been a very important aspect in understanding and assessing the ultimate achievement of impacts from the original GEF investment

The Lewa Case study implemented the three elements of the Impact Evaluation Framework which are summarized below.

Assessed implementation success and failure

Given that no project logical framework or outcomes were defined as such in the original GEF project brief, the GEF Evaluation Office Study of Local Benefits in Lewa (2004), with the participation of senior Lewa staff, identified project outcomes and associated outputs that reflected the various intervention strategies employed by the project and identified missed opportunities in achieving the project goals. The assessment was as follows, and provided an understanding of the project logic used (Box 6.1) and a review of the fidelity with which the project activities were implemented (Box 6.2):

[21] Full Case Study at: http://www.thegef.org/uploadedFiles/Evaluation_Office/Ongoing_Evaluations/Ongoing_Evals-Impact-8Case_Study_Lewa.pdf



Box 6.1 (b) Project Outcome	Assessment
Outcome 1: Long-term institutional and financial capacity of Lewa to provide global and local benefits from wildlife conservation strengthened	**Fully achieved (5)**
Outcome 2: Protection and management of endangered wildlife species in the wider ecosystem strengthened	**Well achieved (4)**
Outcome 3: Community-based conservation and natural resource management initiatives strengthened	**Well achieved (4)**

Assessed the level of contribution (i.e. impact)

A *Targets-Threats analysis* of those ecological features identified as global environment benefits (Black Rhinos and Grevy's Zebra) was undertaken with input from scientists from Lewa and the Northern Rangelands Trust research departments. Box 6.2 (a) and (b) provide an overview of the variables considered to increase robustness of the understanding of ecological changes that have taken place since before the project started.

Box 6.2 (a) Change in Key Ecological Attributes over time

Key Ecological Attribute	Indicator	Unit	Conservation Status			Trend
			Baseline	Project end	Now	
Black Rhino						
Population size	Total population size of Black rhino on Lewa	Number	29	40	54	⇧

Box 6.2 (a) Change in Key Ecological Attributes over time

Key Ecological Attribute	Indicator	Unit	Conservation Status			Trend
			Baseline	Project end	Now	
Productivity	Annual growth rates at Lewa	%	12	13	15	⇧
Suitable secure habitat	Size of Lewa rhino sanctuary	Acres	55,000	55,000	62,000	⇧
Genetic diversity	Degree of genetic variation	-	No data available			
Grevy's zebra						
Population size	Total population size of Grevy's zebra on Lewa	Number	497	435	430	⇔
Productivity	Annual foaling rates on Lewa	%	11	11	12	⇔
Population distribution	Number of known sub-populations and connectivity		No data available			
Suitable habitat (grassland & secure water)	Community conservancies set aside for conservation under NRT	Number	3	4	15	⇧
Genetic diversity	Degree of genetic variation		No data available			

Box 6.2 (b) Current Threats to the Global environment benefits(GEBs)

	Severity[22] Score (1-4)	Scope[23] Score (1-4)	Overall ranking
Black rhino			
Poaching and snaring	3	3	**3**
Insufficient secure areas	2	3	**2**
Habitat loss (due to elephant density)	1	1	**1**
Grevy's zebra			
Poaching	2	2	**2**
Disease	4	2	**3**
Predation	3	1	**2**
Habitat loss/ degradation	3	3	**3**

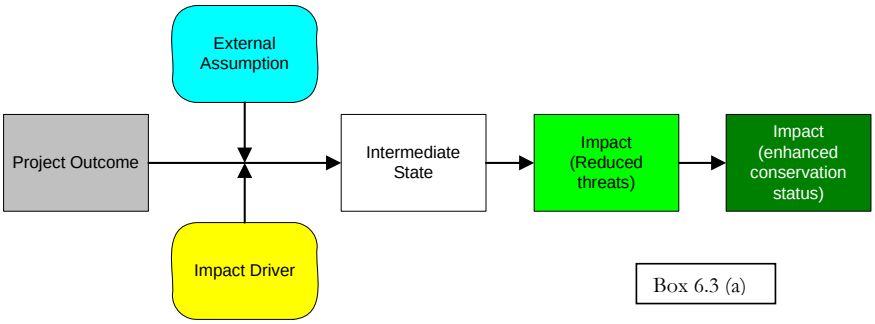
[22] **Severity** (level of damage): Destroy or eliminate GEBs/Seriously degrade the GEBs/Moderately degrade the GEBs/Slightly impair the GEBs

[23] **Scope** (geographic extent): Very widespread or pervasive/Widespread/Localized/ Very localized

Insufficient secure areas	2	2	**2**
Hybridization with Burchell's zebra	1	1	**1**

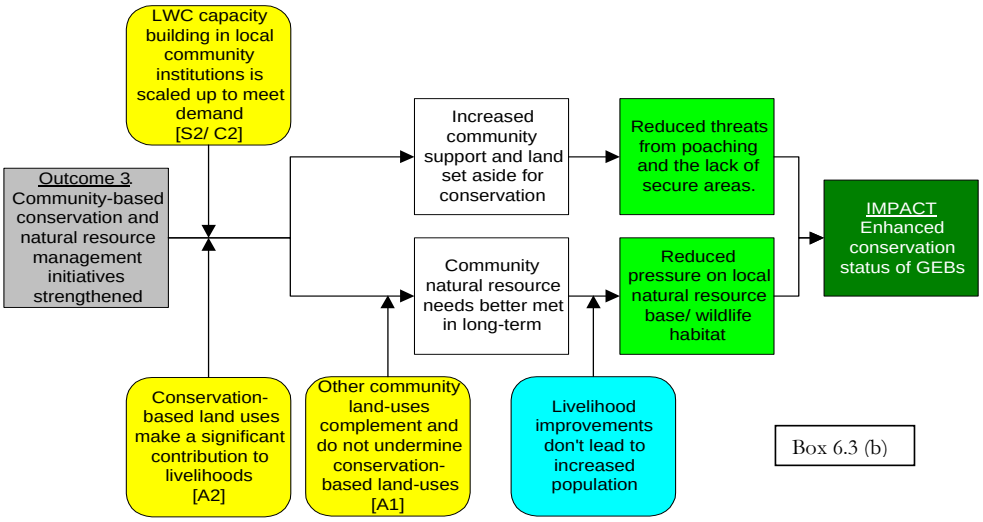
Explanations for observed impact

Theory of Change models were developed for each project Outcome to establish contribution; the framework reflected in Box 6.3(a) was used. This analysis enabled an examination of the links between observed project interventions (discussed in 6.1) and observed impact (discussed in 6.2). As per GEF principles, factors that were examined as potentially influencing results included the *appropriateness* of intervention, the *sustainability* of the intervention and its *catalytic effect* – these are referred to as 'impact drivers.' The next step involved the identification of '*intermediary states*': examining whether the successful achievement of a specific project outcome would directly lead to the intended impacts and, if not, identifying additional conditions that would need to be met to deliver the impact. Taking cognizance of factors that are 'beyond project control', the final step identified those factors that are necessary for the realization and sustainability of the intermediary state(s) and ultimate impacts, but outside the project's influence.



Box 6.3 (a)

An illustrative example is provided by a consideration of Outcome 3 that via *Community-based conservation and natural resource management initiatives strengthened*, expected to achieve enhanced conservation of Black Rhinos and Grevy's Zebras. The *theory of change* model linking Outcome 3 to the intended impacts is illustrated below, in Box 6.3(b). The overall logframe assessment of the project's implementation for community-based conservation and natural resource management was *well achieved (*see Section 6.1*)*. All intermediate factors/impact drivers/external assumptions that were identified received a score of *partially to well achieved*, indicating that together with all its activities, this component was well-conceived and implemented.



In sum for Lewa

The analysis provided indication that the Black rhino and Grevy's zebra populations on the Lewa Conservancy are very well managed and protected. Perhaps the most notable achievement has been the visionary, catalytic and support role that Lewa has provided for the conservation of these endangered species in the broader ecosystem, beyond Lewa. Lewa has played a significant role in the protection and management of about 40% of Kenya's Black rhino population and is providing leadership in finding innovative ways to increase the coverage of secure sanctuaries for Black rhino. Regarding the conservation of Grevy's zebra, Lewa's role in the establishment of community conservancies, which have added almost one million acres of land set aside for conservation, has been unprecedented in East Africa and is enabling the recovering of Grevy's zebra populations within their natural range. However, the costs and resources required to manage and protect this increasing conservation estate are substantial and unless the continued and increasing financing streams are maintained, it is possible that the substantial gains in the conservation of this ecosystem and its global environmental benefits could eventually be reversed.

In conclusion

The assessment of project conceptualization and implementation of project activities in Lewa has been favorable, but, this is coupled with indications that threats from poaching, disease and habitat loss in and around Lewa continue to be severe. Moreover, evaluation of the other case studies *Bwindi Impenetrable National Park and Mgahinga Gorilla National Park Conservation Project*, Uganda and *Reducing Biodiversity Loss at Cross-Border Sites in East Africa*, Regional: Kenya, Tanzania, Uganda, confirmed that to achieve long-term results in the generation of global environment benefits the absence of a specific plan for institutionalized continuation would, in particular, reduce results over time – this was the major conclusion of the GEF's pilot impact evaluation.

Annex 2: Biographies of Reference Group

REFERENCE GROUP PROFILES
Marie-Hélene Adrien is the President and Senior Consultant of a Canadian Consulting firm, Universalia (www.universalia.com) that has specialised over the past 25 years in Monitoring and Evaluation (M&E). Over her career, Marie-Hélène has conducted more than 100 assignments in M&E. She has served for three years on the Board of the Quebec Program Evaluation Society and three years on the Board of the Canadian Evaluation Society. Marie-Hélène has published several books on M&E. A Canadian citizen, Marie-Hélène is originally from Haiti and is fluent in French, English and Spanish.
Paul Balogun Not available
Michael Bamberger has worked on program evaluations and gender impacts of development programs in more than 30 developing countries in Africa, Asia, Latin America, and the Middle East. He worked for 13 years with nongovernmental organizations in Brazil, Colombia, Costa Rica, El Salvador, Peru, and Venezuela before joining the World Bank in 1978. During his 23 years with the World Bank, he worked as an advisor on monitoring and evaluation with the Urban Development Department, as Asia training coordinator for the Economic Development Institute (during which time he organized training programs on monitoring and evaluation for governments and civil society in some 15 Asian countries and in various parts of Africa), and as senior sociologist in the Gender and Development Department. Since retiring from the World Bank in 2001, he has carried out consulting and evaluation training assignments for the World Bank, DFID, USAID, UNICEF, Asian Development Bank, International Food Policy Research Institute (IFPRI), UNDP, U.N. Department of Economic and Social Affairs, IFAD, World Food Program, the U.N. Evaluation Office and the National Planning Department of the Government of Colombia. He has also carried out evaluation consulting assignments for several private consulting agencies. He is on the faculty of the International Program for Development Evaluation Training (IPDET) and the Foundation for Advanced Studies in International Development in Tokyo. Recent evaluation-related publications include: *Integrating Quantitative and Qualitative Research in Development Projects* (2002); senior author of the *Gender* Chapter of the World Bank *Poverty Reduction Sourcebook* (2002); (editor) *Influential Evaluations: Evaluations that improve performance and impacts on development programs* (2004); Lead author (with Jim Rugh and Linda Mabry) *"RealWorld Evaluation: Working Under Budget, Time, Data and Political Constraints."* (2006).
Fred Carden joined IDRC's Evaluation Unit in 1993 and became the Director in March 2004. He has written widely in the areas of evaluation, international cooperation, and environmental management. His current work includes the development of use-oriented evaluation tools and methods. Recent co-publications include "Outcome Mapping," "Organizational Assessment," and "Evaluating Capacity Development." Forthcoming publications (2008) include Knowledge to policy: making the most of development research. Ottawa & Delhi: IDRC & Sage.and *"Using Comparative Data: A systems approach to a multiple case study"*, in, Handbook of Case-Centred Methods, edited by David Byrne and Charles Ragin. Sage. Forthcoming 2008. He has taught and carried out research at York University, the Cooperative College of Tanzania, the Bandung Institute of Technology (Indonesia) and the University of Indonesia. He holds a PhD from the Université de Montréal and a Master's degree in environmental studies from York University. In 2007-2008, he is Research Fellow in Sustainability Science at Harvard University's Center for International Development. He serves on the Performance and Evaluation Committee of the Social Sciences and Humanities Research Council (Canada).
Stewart Donaldson is Professor and Chair of Psychology, Director of the Institute of Organizational and Program Evaluation Research, and Dean of the School of Behavioral and Organizational Sciences, Claremont Graduate University. He has conducted numerous evaluations, developed one of the largest university-based evaluation training programs, and taught and published widely on applied research and evaluation topics. His recent work includes a book with Michael Scriven about the future of evaluation practice - *Evaluating Social Programs and Problems: Visions for the New Millennium* (2003), a book Applied Psychology: *New Frontiers and Rewarding Careers* (2006), and a forthcoming book *Program Theory-Driven Evaluation Science: Strategies and Applications*. He is co-founder of the Southern California Evaluation Association, has served as Co-Chair of AEA's Theory-Driven Evaluation and Program Theory Topical Interest Group, and is on the Editorial Boards of the American Journal of Evaluation and New Directions for Evaluation.

Osvaldo Feinstein Author and editor of books and articles on evaluation, development and economics. Currently adviser to the Spanish Evaluation Agency, professor at the Master on Evaluation of the Universidad Complutense de Madrid, member of the Panel on Monitoring and Evaluation of the CGIAR Science Council, consultant of the evaluation offices of the GEF, IFAD and UNDP, and adviser to several international evaluation networks. In the past, he was a former manager and adviser at the World Bank's Operations Evaluation Department, member of the US National Academy of Sciences Panel on Evidence Based Policy in the Social Sciences, consultant with the Inter-American Development Bank, the African Development Bank, UN Technical Cooperation Department, CEPAL, ILPES and ILO, among other development agencies. Designer of PREVAL. International lecturer on evaluation and development.

Ted Freeman, a partner in the firm Goss Gilroy Inc. in Ottawa Canada, has been a professional evaluator for over 25 years. He began his career with the Government of Canada in the evaluation of regional development programming in the North and soon moved to international development evaluations for a wide range of bilateral and multilateral development institutions and developing countries. In the past two decades he has specialised in leading larger scale joint (bilateral agencies, multilateral agencies and host governments) evaluations of sectors and programs. He was senior evaluation researcher on the Global Evaluation of External Support to Basic Education and most recently, team leader of the External Joint Evaluation of the Health Sector in Tanzania. In the area of impact evaluation he recently served as senior evaluation advisor to a large scale Danida impact evaluation in Tanzania.

Sulley Gariba is an evaluation specialist with over 20 years experience in the design and implementation of systems for institution building, organization development, social policy analysis, training, monitoring and evaluation of development effectiveness. Most of this experience has been related to the design, training and field-based application of Participatory Monitoring, Evaluation and Results-based Management (RBM) systems. Sulley Gariba has been a leader in the international evaluation movement, having been founding President of the International Development Evaluation Association (IDEAS) from 2002 to 2005. He has served on several expert panels, the most recent being the 7-member panel of evaluation experts established by the OECD-DAC and the United Nations Evaluation Group (UNEG) to assess the evaluation function in UNICEF. He has moderated high-level professional development of major international organizations using an RBM framework; designed and delivered training for over a dozen African Parliaments; and supported the design of water supply improvement projects for small towns in Ethiopia; capacity building for decentralized development in Nigeria; human resources development strategies for water sector managers and community development funds in Eritrea; institutional development using RBM for the Economic Community of West African States (ECOWAS) in Abuja and member states. He has recently led a major evaluation of UNDP's HIV and AIDS programmes in 10 Southern Africa countries and Ethiopia, adding to his expertise in the capacity to integrate HIV and AIDS considerations in development planning, management, monitoring and evaluation. Sulley Gariba heads of the Institute for Policy Alternatives, Ghana, and leads major evaluations and training of Parliaments and civil society on evaluation.

Jennifer C Greene has been an evaluation scholar-practitioner for over 25 years. She received her doctorate in educational psychology from Stanford University in 1976 and has held academic appointments at the University of Rhode Island, Cornell University, and presently, the University of Illinois, Urbana-Champaign. Her evaluation scholarship has broadly focused on probing the intersections of social science method with policy discourse and program decision making, with the intent of making evaluation useful and socially responsible. Greene has concentrated specifically on advancing qualitative, mixed methods, and democratic approaches to evaluation. Her evaluation practice has spanned multiple domains of practice, with an emphasis on the domains of education, community-based family services, and youth development. She currently has a grant from the National Science Foundation to pursue an "educative, values-engaged" approach to evaluating science and mathematics education programs. She has published widely in journals and books on program evaluation; she has held leadership positions in AERA and AEA and was recently co-editor-in-chief of *New Directions for Evaluation.* In 2003, she received AEA's Paul F Lazarsfeld award for contributions to evaluation theory. In 2007 her book on mixed methods social inquiry will be published.

Ernie House Ernest R. House is a Emeritus Professor in the School of Education at the University of Colorado at Boulder. Previously, he was at the Center for Instructional Research and Curriculum Evaluation (CIRCE) at the University of Illinois, Urbana-Champaign.He has been a visiting scholar at UCLA, Harvard, and New Mexico, as well as in England, Australia, Spain, Sweden, Austria, and Chile. His primary interests are evaluation and policy analysis. Books authored include *Evaluating with Validity* (1980), *Jesse Jackson and the Politics of Charisma* (1988), *Professional Evaluation: Social Impact and Political Consequences* (1993). He is the 1989 recipient of the Harold E. Lasswell Prize presented by *Policy Sciences* and the 1990 recipient of the Paul F. Lazarsfeld Award for Evaluation Theory, presented by the American Evaluation Association. He was editor of *New Directions in Program Evaluation* (1982 to 1985) and columnist for *Evaluation Practice* (1984-89). Studies include evaluation of the Illinois Gifted Program for the Illinois legislature (1968-1972), assessment of the Michigan Accountability Program for the National Education Association (1974), critique of the National Follow Through Evaluation for the Ford Foundation (1977), audit of the Promotional Gates Program evaluation for the Mayor's Office in New York City (1981), assessment of environmental education policies in Europe for OECD (1992), and evaluation of science, engineering, and technology education programs across federal departments for the Federal Coordinating Council for Science, Engineering, and Technology in Washington (1993).

Mel Mark is Professor of Psychology at Penn State University. He is Past-President of the American Evaluation Association, and has served as Editor of the *American Journal of Evaluation* where he is now Editor Emeritus. His interests include the theory, methodology, practice, and profession of program and policy evaluation. He has been involved in evaluations in a number of areas, including prevention programs, federal personnel policies, and various educational interventions including STEM program evaluation. Among his books are *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs* (Jossey-Bass, 2000; with Gary Henry and George Julnes) and the recent *SAGE Handbook of Evaluation* (Sage, 2006; edited with Ian Shaw and Jennifer Greene), as well as forthcoming books *Exemplars of Evaluation* (Sage; with Jody Fitzpatrick and Tina Christie) and *Social Psychology and Evaluation* (Guilford; with Stewart Donaldson and Bernadette Campbell).

John Mayne is an independent advisor on public sector performance. He has been working with a number of organizations and jurisdictions, including the Scottish Government, the United Nations, the International Development Research Centre, the OECD, the Asian Development Bank, the European Union, and several Canadian federal organizations on results management, evaluation and accountability issues. Until 2004, he was at the Office of the Auditor General where he led efforts at developing practices for effective managing for results and performance reporting in the government of Canada, as well as leading the Office's audit efforts in accountability and governance. Prior to 1995, Dr. Mayne was with the Treasury Board Secretariat and Office of the Comptroller General. He has authored numerous articles and reports, and edited five books in the areas of program evaluation, public administration and performance monitoring. In 1989 and in 1995, he was awarded the Canadian Evaluation Society *Award for Contribution to Evaluation in Canada*. In 2006, he became a Canadian Evaluation Society Fellow.

Masafumi Nagao is Research Professor at the Center for the Study of International Cooperation in Education at Hiroshima University (Japan) where his primary work is to conduct research relating to the evaluation of aid programs and projects in the field of education. From 1999 to 2006 he served as the leader of a JICA team to support a secondary mathematics and science teacher retraining project in South Africa. He is a member of the Advisory Committee of JICA on evaluation and also serves on the School Evaluation Committee of Japan's Ministry of Education. Prior to joining this center, from 1987 to 1998, Mr. Nagao worked for the Sasakawa Peace Foundation, one of the largest grant-making foundations in Japan, as a Chief Program Officer and Program Director. From 1975 to 1987, he served as Economic Affairs Officer in the Technology Transfer Division of the United Nations Conference on Trade and Development (UNCTAD) in Geneva. He received a B.A. in economics from Carleton College in Minnesota and an M.A. in economic policy from Hitotsubashi University in Tokyo. He has published many papers on educational cooperation, school evaluation and evaluation of aid projects. He has just co-edited a book on "Mathematics and Science Education in Developing Countries: Issues, Experiences and Cooperation Prospects" (Quezon City: University of the Philippines Press, 2007).

Michael Quinn Patton lives in Minnesota where, according to the state's poet laureate, Garrison Keillor, "all the women are strong, all the men are good looking, and all the children are above average." It was this lack of interesting statistical variation in Minnesota that led him to qualitative inquiry despite the strong quantitative orientation of his doctoral studies in sociology at the University of Wisconsin. He is former president of the American Evaluation Association and one of only two recipients of both the Alva and Gunnar Myrdal Award for Outstanding Contributions to Useful and Practical Evaluation from the Evaluation Research Society and the Paul F Lazarsfeld Award for Lifelong Contributions to Evaluation Theory from the American Evaluation Association. The Society for Applied Sociology awarded him the 2001 Lester F Ward Award for Outstanding Contributions to Applied Sociology. He was on the faculty of the University of Minnesota for 18 years, including five years as director of the Minnesota Center for Social Research, where he was awarded the Morse-Amoco Award for innovative teaching. Dr Patton has authored five other Sage books: *Utilization-Focused Evaluation*, *Creative Evaluation*, *Practical Evaluation*, *How to Use Qualitative Methods in Evaluation*, and *Family Sexual Abuse: Frontline Research and Evaluation*. His creative nonfiction book, *Grand Canyon Celebration: A Father-Son Journey of Discovery*, was a finalist for 1999 Minnesota Book of the Year.

Ray Pawson is Professor of Social Research Methodology in the School of Sociology and Social Policy, University of Leeds, UK. His main interest, perforce, is in research methodology and he has written widely on the principles and practice of research, covering methods - qualitative and quantitative, pure and applied, contemporaneous and historical. Publications include *A Measure for Measures* (1989), *Realistic Evaluation* (1997) and *Evidence-Based Policy: A Realist Perspective* (2006). He was elected president of the Committee on Logic & Methodology of the International Sociological Association (94-98). He has served much time in prison (for research purposes) being a former UK director of the International Forum for Education in Penal Systems (95-97). He has held the post of visiting professor at the University of Rome, the University of Victoria, BC Canada and RMIT, Australia as well as visiting fellow at the UK ESRC Centre for Evidence Based Policy and Practice, University of London. He is best known for his writing on evaluation methodology, research synthesis and evidence based policy, work which has been supported over the years by three ESRC senior fellowships. He has acted as researcher and consultant on programme evaluation for various UK and European agencies.

Robert Picciotto, Visiting Professor, Kings College, London, is a trustee of the Oxford Policy Institute and a member of the United Kingdom Evaluation Society Council. He graduated from the Ecole Nationale Superieure de l'Aeronautique (France) in 1960 and from the Woodrow Wilson School of Public and International Affairs (Princeton) in 1962. His career in development spans over 40 years. In his last post at the World Bank, he reported to the Board of Directors of the World Bank as Director-General, Evaluation (1992-2002). He previously served as Vice President for Corporate Planning and Budgeting. His other operational assignments include development banking specialist, agricultural economist in the New Delhi office, Division Chief, Agricultural Industries, Assistant Director, Agriculture and Rural Development, and Projects Director in three of the World Bank's Regions. Over the past four years, he has advised the Council of Europe Development Bank, the Asian Development Bank, the African Development Bank, the United Nations Development Program, the International Fund for Agriculture Development, the Organization for Economic Cooperation and Development, Sweden's Ministry for Foreign Affairs and the Department for International Development of the United Kingdom. He has published widely on evaluation, security and development topics.

Patricia Rogers is Professor in Public Sector Evaluation, CIRCLE (Collaboration for Interdisciplinary Research, Consulting and Learning in Evaluation) at RMIT University (Royal Melbourne Institute of Technology), Australia. Dr Rogers has worked in public sector evaluation and research for more than 20 years across a wide range of types of programs and with a large number of clients in the USA, Japan, South Africa, Malaysia, Australia and New Zealand. Her most recent publications include a paper on using program theory for complex and complicated programs, entries on accountability, program logic and logic models for the Encyclopaedia of Evaluation and a chapter on using evaluation for improvement and organisational learning for the Sage Handbook of Evaluation.. She has twice won the AES' Caulley-Tulloch Prize for Pioneering Literature in Evaluation, and in 2003 was awarded the AES' 'Evaluation Training and Services Award', its highest award, for outstanding contributions to the profession of evaluation. In 2005 she became the only non-American to be awarded the American Evaluation Association's Gunnar and Alvar Myrdal Award for Evaluation Practice - presented to an outstanding evaluation practitioner who has made substantial cumulative contributions to the professional practice of evaluation. She is on the Editorial Boards of Evaluation, The American Journal of Evaluation, the Journal of Multidisciplinary Evaluation and New Directions in Evaluation (Associate Editor), on the advisory board of the Western Michigan University's Evaluation Checklists project, and a founding member and former board member of the Australasian Evaluation Society (which covers Australia and New Zealand). She was the inaugural Chair of the American Evaluation Association Topical Interest Group on Systems in Evaluation.

Thomas A Schwandt is University Distinguished Teacher/Scholar and Professor of Education at the University of Illinois at Urbana-Champaign (UIUC) where he holds appointments in the Department of Educational Psychology, the Department of Educational Policy Studies, and the Unit for Criticism and Interpretive Theory. He has served in various other academic positions, including a faculty appointment at the University of Illinois Chicago medical school. He was also employed in the private sector as an organizational consultant and program evaluator and has lectured and taught throughout Scandinavia, including as visiting professor at Roskilde University in Denmark and research professor at SKUR, the National Center for Comprehensive Rehabilitation Research and Development in Norway. He has authored more than sixty papers and chapters on issues in theory of evaluation and interpretive methodologies. He is the author of *Evaluation Practice Reconsidered* (Peter Lang, 2004); *Evaluating Holistic Rehabilitation Practice* (Oslo, Kommuneforlaget, 2004); *Dictionary of Qualitative Inquiry* (Sage, 1997, 2001, 2007), and with Edward Halpern, *Linking Auditing and Meta-evaluation* (Sage, 1988); he has co-edited *Evaluating Educational Reforms: Scandinavian Perspectives* (Information Age Press, 2003) with Peder Haug, *Exploring Evaluator Role and Identity* (Information Age Press, 2002) with Katherine Ryan, and *Knowledge Production: The Work of Educational Research in Interesting Times* (Routledge, forthcoming) with Bridget Somekh. In 2002 he received the Paul F. Lazarsfeld Award from the American Evaluation Association for his contributions to evaluation theory. He is currently an AEA Board member as well as member of the U.S. National Academy of Sciences (Division of Behavioral and Social Sciences and Education) Standing Committee on Social Science Evidence for Use.

Nicoletta Stame teaches Social Policy at the University of Rome "La Sapienza". She was a co-founder and the first President of the Italian Evaluation Association (AIV, 1998-2001). She is the past President of the European Evaluation Society (2004-05). Nicoletta is interested in the theory and methods of evaluation. She is the author of "L'esperienza della valutazione" (Rome, 1998), editor of "Per una valutazione delle politiche di emersione" (Milano, 2005), co-author (with R. Rist) of "From Studies to Streams. Managing Evaluative Systems" (New Brunswick, NJ, 2006), and of many essays in volumes and journals. Her professional background includes researches on socio-economic development, industrial policies and family business, welfare policies. She has evaluated programs of enterprise creation, aid to SMEs, labor regularization and social integration. Her work aims at enhancing the evaluation capacities of public administrators, program implementers and beneficiaries.

Bob Williams is known in evaluation circles, through his evaluations, his work within the Australasian Evaluation Society (AES) and the American Evaluation Association (AEA), his contributions to a range of internet discussion groups including "Evaltalk", plus a wide variety of lectures, publications, books and workshops. His speciality includes the application of systems concepts, action research and large group processes in evaluation settings.

Bob co-edited the AEA Monograph Systems Concepts in Evaluation - an expert anthology. This publication explored the potential of systems based approaches to evaluation. This is the first publication of its kind and can be downloaded from his website (http://www.bobwilliams.co.nz). He is co-editor of the forthcoming Evaluation Journal of South Asia, which addresses directly the kinds of issues close to NONIE's heart. He is also a member of the Editorial Boards of the American Journal of Evaluation and New Directions in Evaluation and current Chair of the American Evaluation Association Topical Interest Group on Systems in Evaluation..