DEVELOPING A METHODOLOGY FOR IMPACT EVALUATION AT UNIDO: POSSIBLE ROUTES AND OPTIONS

Report prepared by

Lilit Melikyan

On behalf of the UNIDO Evaluation Group

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION

Vienna
May 2008

OUTLINE

ABBREVIATIONS

ADF	Agence de Français
DAC	Development Assistance Committee
DFID	Department for International Development
EVA	Evaluation Unit at UNIDO
GEF	Global Environmental Facility
IE	Impact Evaluation
IEG	Independent Evaluation Group
IFAD	International Foundation for Agricultural Development
IFC	International Financial Corporation
KfW	Kreditanstalt für Wiederaufbau,
M&E	Monitoring and Evaluation
NONIE	Network of Networks on Impact Evaluation
NGO	Non Governmental Organisation
OECD	Organisation for Economic Cooperation and Development
RBM	Results Based Management
UN	United Nations
UNCDF	United Nations Capital Development Fund
UNIDO	United Nations Industrial Development Organisation
WB	World Bank

ACKNOWLEDGEMENTS

I would like to thank the staff of Evaluation Group of UNIDO, as well as staff of other UNIDO branches for their time, information and comments.

The views and opinions expressed in this report are mine and do not necessarily reflect the views of UNIDO.

This document has not been formally edited.

1. INTRODUCTION

There is a growing appreciation within the development community of the merits of conducting impact evaluations of development programs and projects, as an accountability tool and a way of understanding of what works and what does not. There is, also, a growing pressure on foreign aid providers, including international organisations, bilateral agencies and international NGOs to measure the impacts of their interventions, as the focus on poverty reduction and achieving Millennium Development Goals is growing. Impact evaluations help to better understand the extent to which activities reach the poor and the magnitude of their effects on people's welfare.
A number of other factors also contribute to this, including initiatives at the World Bank.[1]

As a response to these developments, *impact evaluation* is an area of increasing interest and attention across the development agencies. Many of them have developed or are in the process of formulating their approaches to impact evaluation.

- Agencies increasingly experiment with different approaches to impact evaluation. While some gravitate towards using quantitative techniques, an understanding is also growing of the fact that many of development projects are not suited for this. Real-life constraints, multifaceted nature of projects and multi-partner implementation arrangements have resulted in increased appreciation of qualitative methods. Valuable experience is being accumulated in both directions. Also, a consensus is growing around the need to asses not just the final impacts of the projects, but also the nature of changes that occur along the results' chain during the project implementation (referred to as a *theory of change* or *theory based approach to impact evaluation*): it provides the context without which the validity of conclusions on final impacts might be questioned. The best examples in impact evaluations combine quantitative and qualitative approaches, putting these in the context of theory of change induced by a project.

- Some agencies, like the WB and the IFC are already carrying out impact evaluations, using sophisticated quantitative techniques available to measure impact. IFC, in particular, currently has 30 experimental and quasi-experimental impact evaluations in its portfolio, with 23 presently underway[2]. Particular attention is given to evaluating pilot projects prior to roll-out and replication, and projects that require testing several approaches to identify the most effective ones.

- The WB piloted the first major impact evaluation (of its Social Investment Funds), where a theory based approach was used, combined with a quasi-experimental design and case studies. GEF has now taken theory based approach as the min pre-requisite to conducting impact evaluations, combining it with different quantitative and qualitative methods.

[1] See the DIME initiative: http://go.worldbank.org/HIYKB2QV00. For IEG work on impact evaluation, see http://www.worldbank.org/ieg/ie

[2] http://www.ifc.org/ifcext/rmas.nsf/Content/home

- A number of other agencies have started experimenting with different impact evaluation methods. DFID, ADF, IFAD and a number of UN agencies (e.g. UNCDF) have by now conducted a number of experimental and quasi-experimental evaluation designs, as well as studies using qualitative methods.[3]

- Some agencies consider assessing impact to be one the key features of conducint routine evaluations: ex-post evaluations carried out by KfW/FZE, for instance, all focus on assessing impact, using a lighter and less scientific approach than the IFC/World Bank model[4].

- Other agencies are investing in more impact evaluations in the future. For example, Oxfam plans to undertake impact evaluation for large donor-driven programmes and is currently carrying out a feasibility study for a randomised control evaluation of a large livelihoods programme in West Africa[5].

- Results Based Management (RBM) is being introduced across organisations, including improvements in M&E systems and management reforms to enhance their efforts to pursue evidence – based strategies.

Unfortunately these efforts are less standardised (at least in terms of reaching a consensus about the definitions of key terms and appropriateness of methods), there is less of pooling of resources for conducting joint evaluations and less of recipient country involvement than is desired.[6] There are, however, developments along these directions as well. In particular:

- A growing number of developing countries also recognise the benefits from impact evaluation, and many are making efforts to institutionalise monitoring and evaluation systems as part of sound governance[7].

- Some aid agencies demonstrate their willingness to join forces for harmonising approaches and concepts and conducting joint evaluations, e.g. through NONIE network (Network of Networks in Impact Evaluation)[8].

[3] Howard White, Michael Bamberger, Ann Flanagan and Shampa Sinha, "Impact Evaluation in Official Development Agencies', 2008, upcoming

[4] "A Comparative Study of Evaluation Policies and Practices in Development Agencies', AFD/ODI, December 2007

[5] ibid

[6] See for example Center for Global Development:" When Will we ever learn? Improving Lives Through Impact Evaluation", Report of the Evaluation GAP Working Group, Washington DC, 2006

[7] See for example, the Report by Subgroup 2 of the NONIE group on "Impact evaluation guidance". January 08

[8] NONIE is a network of networks for impact evaluation comprised of the DAC Evaluation Network, The United Nations Evaluation Group (UNEG), the Evaluation Cooperation Group (ECG), and a fourth network drawn from the regional evaluation associations. Its purpose is to foster a program of impact evaluation activities based on a common understanding of the meaning of impact evaluation and approaches to conducting impact evaluation. http://www.worldbank.org/ieg/nonie

- International Initiative for Impact Evaluation or 3IE ("Triple IE"), a membership organization, was set up recently, designed to dramatically increase the number of rigorous impact evaluations in areas such as health and education.[9]

- The ONE UN approach now being piloted in a number of countries is a step forward towards harmonisation and pooling of resources for impact evaluations within the UN.

Draft UNIDO RBM Implementation Plan[10] (p.17) requires that a practical impact evaluation methodology, which is appropriate for typical UNIDO interventions is *developed and tested* by the end of 2008. It also states that the drive towards greater accountability for development impact confronts UNIDO with specific challenges, given that demonstrating MDG impact, in particular on poverty, is relatively complex for those typical UNIDO interventions that are critical for economic development, but whose MDG impact is subject to extended causal chains and dependent upon complementary interventions, often from other UN organisations. Currently, no impact evaluation methodology that would be practical, is available for typical UNIDO interventions.

UNIDO Evaluation Group (EVA) is seeking an effective approach to developing an impact evaluation methodology matching specific requirements and possibilities of UNIDO and ways of integrating impact evaluation into the toolbox of existing UNIDO evaluation methods.

This report is the first step in the direction of answering that call based on

- interviews at UNIDO HQ to identify the expectations and attitudes of UNIDO senior staff from different branches *vis-a-vis* impact evaluation; and
- desk study, comparing existing approaches and testing their specific matches with UNIDO requirements.

The result is an outline of a proposed tailor made approach for UNIDO for conducting impact evaluations, introducing recommendations for further steps to finalise the development of this approach and for its implementation. In particular, more work will be needed in investigating how suitable the recommended approaches are for UNIDO, based on a detailed review of its portfolio, extending the approach to cover thematic areas not covered in this report, and testing the recommended approaches and evaluation methods.

In Chapter 2, the definitions and concepts behind the notions of *impact* and *impact evaluation* are discussed, along with advantages and problems associated with conducting impact evaluations.

In Chapter 3, the main specifics of UNDO operations, which need to be taken into account when designing its approach to Impact Evaluations are identified.

[9] http://www.cgdev.org/content/opinion/detail/15102/. Current members are: the Mexican Ministries of Health and Education, Ugandan Ministry of Finance, UK Department for International Development, Netherlands Ministry of Foreign Affairs, Canadian International Development Agency, IRC, African Development Bank, Bill & Melinda Gates Foundation, William and Flora Hewlett Foundation, Google, CARE USA, and Save the Children (United States)

[10] RBM Implementation Plan with milestones for the biennium 2008-2009, UNIDO, Final Draft prepared by the RBM Steering Committee Submitted to UNIDO Directors for consideration, 30 November 2007

In Chapter 4, the main existing approaches and methods of conducting impact evaluations are presented.

In Chapter 5 the main principles of approaches to conducting impact evaluation at UNIDO are suggested.

In Chapter 6, the suggested approaches for Impact Evaluation at UNIDO are described in more detail for projects that could be classified under Business Development Services' thematic area (covering the majority of components under UNIDO TC Major Programs C and D).

Chapter 7 summarises with the main Recommendations.

This report has been prepared by Lilit Melykian on behalf of the UNIDO Evaluation Group.

2. IMPACT EVALUATION: MAIN CONCEPTS

Development Assistance Committee (DAC) of the Organisation of Economic Cooperation and Development (OECD) defines five criteria for conducting evaluations of development programs and projects: relevance, effectiveness, efficiency, sustainability and impact. Thus, impact evaluations are only one of the five evaluation criteria, and it is important to keep this perspective in mind: carrying more *impact* evaluations should not come at the expense of other evaluations.

Box 1. DAC evaluation criteria

Relevance
The extent to which the objectives of a development intervention are consistent with beneficiaries' requirements, country needs, global priorities and partners' and donors' policies. Retrospectively, the question of relevance often becomes a question as to whether the objectives of an intervention or its design are still appropriate given changed circumstances.

Effectiveness
The extent to which the development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance.

Efficiency
A measure of how economically resources/inputs (funds, expertise, time, etc.) are converted to results.

Impact
Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.

Sustainability
The continuation of benefits from a development intervention after major development assistance has been completed. The probability of continued long-term benefits. The resilience to risk of the net benefit flows over time.

Source: Glossary of Key Terms in Evaluation and Results Based Management, OECD-DAC 2002

While most of foreign aid providers are conducting evaluations of their interventions along the lines of relevance, effectiveness and efficiency (much less so of sustainability), impact evaluations are still not common.

UNIDO Evaluation Policy[11] defines *evaluation* as an "…assessment, as systematic and impartial as possible, of a project, programme or entire strand of activities under a single thematic or institutional heading…", determining the ***relevance, impact, effectiveness, efficiency and sustainability*** of UNIDO interventions in line with DAC criteria (see Box 1).

Before proceeding further it is important to define the concepts of *impact* and *impact evaluation*.

DAC defines impacts as "Positive and negative, primary and secondary *long-term effects* produced by a development intervention, directly or indirectly, intended or unintended". This involves the main impacts and effects resulting from the activity on the local social, economic, environmental and other development indicators.

Impact evaluation is the systematic identification of the effects – positive or negative, intended or not – on individual households, institutions, and the environment caused by a given development activity such as a program or project.[12]

[11] http://www.unido.org/doc/5122

[12] Glossary of Key Terms in Evaluation and Results Based Management, OECD-DAC 2002

Impact evaluations can range from large scale sample surveys, in which project populations and control groups are compared before and after project interventions, and possibly at several points during program intervention to small-scale rapid assessments and participatory appraisals, where estimates of impact are obtained from combining group interviews, key informants, case studies and available secondary data. Different approaches to Impact evaluations are described in Chapter 4.

Impact evaluations can be used for[13]:

- Measuring outcomes and impacts of an activity and distinguishing these from the influence of other, external factors;
- Helping to clarify whether costs for an activity are justified;
- Informing decisions on whether to expand, modify or eliminate projects, programs or policies;
- Drawing lessons for improving the design and management of future activities;
- Comparing the effectiveness of alternative interventions;
- Strengthening accountability for results.

The advantages of conducting Impact Evaluations include the following:

- They provide estimates of the magnitude of outcomes and impacts for different demographic groups, regions or over time;
- They provide answers to some of the most central development questions: to what extent are we making a difference; what are the results on the ground; how can we do better?
- Systematic analysis and rigor can give managers and policy-makers added confidence in decision-making.

There are difficulties associated with conducting impact evaluations too, such as:

- Some approaches are very expensive and time-consuming, and although faster and more economical approaches are also used, they come at a cost of compromising on the rigour that is applied (See Section 4.2. Box 10);
- Oftentimes, conducting impact evaluations under severe budget and data constraints carries high risks of affecting the reliability of results;
- Their utility could be reduced when decision-makers need information quickly, since rigourous impact evaluations require are time consuming;
- There usually are difficulties in identifying an appropriate counter-factual;
- Strong technical skills in social science research design and analysis are required, with, ideally, a balance of quantitative and qualitative research skills on the part of the evaluation team.

While there is consensus about the general DAC definitions of *impact* and *impact evaluations,* they *have come to mean different things to different people.* The following four meanings of impact evaluation have recently been the most common[14]:

- An evaluation which looks at the impact of an intervention on final development outcomes, rather than only at project outputs, or a process evaluation which

[13] WB/OED, "Monitoring and Evaluation: Some Tools, Methods & Approaches', Washington DC, 2004

[14] Howard White "Impact Evaluation – the Experience of the Independent Evaluation Group of the World Bank", IEG, WB, Washington DC, September 2006

focuses on implementation;

- An evaluation concerned with establishing the counterfactual, i.e. the difference the project made (how indicators behaved with the project compared to how they would have behaved without it);

- An evaluation carried out some time (five to ten years) after the intervention has been completed in order to examine its long-term effects; and

- An evaluation considering all interventions within a given sector or geographical area.

A fifth and increasingly common one uses a variety of ex-ante appraisals, e.g. poverty impact assessments; conflict assessments, environmental impact assessments, and so on. While these are ex-ante appraisals, the term illustrates the multi- dimensional nature of the use of the concept of *impact* in development agencies today. They are also important in terms of enabling better ex-post impact evaluations, as discussed in Section 4.3 of this report.

These definitions are not mutually exclusive. The need and increased emphasis to show *results* supports the idea of seeing impact evaluation as being concerned with *final* development (or welfare) outcomes, which is in line with the first of these definitions. At the same time there is increased awareness that attention needs to be paid to the establishment of a good counterfactual, which is the second one of these definitions.

The WB, through its IEG, along with a number of other agencies, has adopted an approach to Impact Evaluation which combines these two definitions[15].
.
While these meanings of Impact Evaluation are not mutually exclusive, they, at the same time, do not necessarily coincide. For example, final development outcomes need not necessarily be long-term effects.

These different meanings of impact evaluation also emphasize the time factor in relation to impact evaluations:

- conducting ex-ante impact assessments before the commencement of the project;
- assessing the potential for achieving the intended impact during the project implementation (*impact monitoring*);
- projecting the potential impact at the end of the projects; and
- impact evaluations *per se* at some point after the projects are over.

This evolving perspective of the concept of impact evaluation is discussed in Section 4.3. This perspective also highlights the importance of data collection (baseline, and monitoring data), setting up effective and continuing monitoring schemes, and the need for cooperating with recipient governments to ensure that monitoring extends beyond the project.

In the light of the coexistence of this multitude of notions of *impact evaluation,* it is important for UNIDO to clearly state where it stands.

[15] World Bank Web site on impact evaluation: http://www.worldbank.org/poverty/impact/

According to the definition adopted by UNIDO, it is concerned with *long term effects* vs. or, rather, in addition to, immediate project-completion results. Capturing real long term effects however is often difficult for a variety of reasons: most of the time project personnel is no longer in the field, complicating making in-depth inquires about the projects' progress, assumptions, and other factors; committing funds for projects which were completed 3-5 years ago becomes problematic due to changes in management priorities, budget constrains and so on.

Hence, UNIDO might consider including "lighter" version of impact assessments, or "projections of impacts" in its post-project evaluations of individual projects and IP programs. We shall return to this in Section 4.3.

This does not mean discarding the option of long-term impact evaluation for selected projects and programs. However, impact evaluations *per se* should be undertaken only for selected projects/programs. NONIE (2008)[16] suggests that impact evaluation should ideally be conducted when an assessment shows that political, technical, resource and other practical considerations are adequate. It further recommends the following criteria:

- 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 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, and useful evaluation.

UNIDO will need to come up with its own list of criteria for choosing projects and programs for conducting impact evaluations along these lines.

[16] NONIE Subgroup 2, Impact Evaluation Guidance, Draft, 01/08

Across all different types of impact evaluations there are four common tasks, all of which must be undertaken adequately to increase the rigour of for impact evaluation[17]:

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, and environmental - for individuals, families, households, communities and organizations;

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.

Specific approaches and methods can further improve rigour in each task, as discussed throughout the text.

Amidst the growing interest in impact evaluations, there is an explicit call for counterfactual to be established, as the need for more impact evaluation does call for more and more rigorous, impact evaluation, which means capturing the net effects of programs/projects whenever and wherever possible. But limitations of this approach are being recognised too, as well as the need for counterfactual-based evaluations to be well-contextualised, adopting methodological flexibility to fit the conditions under which they are carried out. UNIDO evaluation policy calls for precisely this – requiring the analysis to capture the changes occurring along the entire causal chain of a project: these concepts need to be elaborated more clearly.

A *quality impact evaluation* must:

- Develop a set of indicators that can meaningfully and reliably define and measure project inputs, implementation processes, outputs, intended outcomes and impacts;
- Develop a logically sound counterfactual presenting a plausible argument that observed changes in outcome indicators after the project intervention are in fact due to the project and not to other unrelated factors, such as improvements in the local economy or programs organized by other agencies;
- Determine, in accordance with accepted statistical procedures, whether a project has contributed to the intended impacts and benefited a significant proportion of the target population;

[17] ibid

- Be based on examination of the entire causal chain (as is the case with Theory based evaluations).

We return to this in Chapter 4.

.

While UNIDO evaluation policy, in line with General Assembly resolution 59/250 calls for systematic evaluation of operational activities of the UN system *by assessing their impact on poverty eradication, economic growth and sustainable development,* so far there have been no *specific* impact evaluations of UNIDO projects carried out. Hence, opportunities to learn from lessons of impact evaluations are limited. Therefore, this has to be taken into account in treating the recommendations in this report: there should be a period of piloting and testing, to allow choosing those methods of impact evaluation that most suit UNIDO.

Before making these recommendations and specific challenges that impact evaluations may pose for UNIDO (in Chapter 4), it is important to identify the specifics of UNDO operations, which need to be taken into account when designing its approach to conducting impact evaluations. Chapter 3 addresses these specifics.

3. UNIDO: OVERVIEW AND NEEDS ASSESSMENT FOR IMPACT EVALUATION

UNIDO promotes industrialisation in developing and transition countries, in cooperation with its 172 Member States through strengthening industrial capacities and cleaner and sustainable industrial development. UNIDO's assistance is delivered through two, mutually complementary and supportive core functions: a normative function as a Global Forum, and an operational function, providing Technical Cooperation (TC).

- As a *global forum*, UNIDO generates and disseminates knowledge relating to industrial matters and provides a platform for the various actors in the public and private sectors, civil society organisations and the policy- making community in general, to enhance cooperation, establish dialogue and develop partnerships.

- As a *technical cooperation* agency, UNIDO designs and implements programmes to support the industrial development efforts of its clients. It also offers tailor-made specialised support for programme development.

UNIDO achieves its objectives through:

- Integrated programmes (IPs, at a level of a country or a region) or Country Service Frameworks (CSFs); and

- Stand-alone projects.

UNIDO has sharpened its technical cooperation activities by focusing on three themes (falling under its list of Major Programmes), which directly respond to international development priorities:

- *Poverty Reduction through Productive Activities:* UNIDO addresses Poverty Reduction by promoting employment and income generation among the poor.

- *Trade Capacity Building:* UNIDO combines building up the technical infrastructure required to participate in international trade (e.g. standards, quality, metrology, accreditation and certification) while also strengthening key export sectors that require support services in upgrading productive and export capacities .

- *Energy and Environment*: UNIDO assists countries in the implementation of activities related to the multilateral environmental agreements; the promotion of energy efficiency; and the promotion of sustainable production and consumption practices.

UNIDO's operations are budgeted on a biennial basis, funded by assessed contributions of member states and voluntary contributions.

UNIDO has three policy-making organs: the Programme and Budget Committee; the Industrial Development Board; and the General Conference. The Programme and Budget Committee, consisting of 27 Members is a subsidiary organ of the Board, and assists it in preparing work programmes and budgets. The Programmes and Budgets documents, the current one being the "Programmes and Budgets: 2008-2009" features

non only the strategy and programs envisioned for those years, along the Major Programme Areas (see Box 2), but also specifies the outcome indicators at corporate level for each of these Programme areas.

Box 2 Major programmes and program components

MAJOR PROGRAMME C: POVERTY REDUCTION THROUGH PRODUCTIVE ACTIVITIES

Programme Components

1 Industrial Policy, Business Environment and Institutional Support
2 Rural and Women's Entrepreneurship Development
3 SME Cluster and Network Development
4 Agro-processing and Value Chain Development
5 Rural Energy for Productive Use
6 Sustainable Production in Poor Communities
7 Technology Diffusion
8 Promotion of Domestic Investment, FDI and Alliances

MAJOR PROGRAMME D: TRADE CAPACITY-BUILDING

Programme Components

1 Competitiveness Analysis and Trade-related Policies
2 Enterprise Upgrading for Trade Enhancement
3 Innovation Systems, Technology Management and Foresight
4 Modernization of Export-oriented Agro-industries
5 SME Export Consortia
6 Corporate Social Responsibility for Market Integration
7 Standards, Metrology, Testing and Conformity

MAJOR PROGRAMME E: ENERGY AND ENVIRONMENT

Programme Components

1 Renewable Energy.
2 Industrial Energy Efficiency and Climate Change
3 Cleaner and Sustainable Production
4 Water Management
5 Montreal Protocol
6 Stockholm Convention

MAJOR PROGRAMME F: CROSS-CUTTING PROGRAMMES AND COUNTRY-LEVEL COHERENCE

Programme Components

1 Industrial Research and Statistics
2 Strategic Research
3 Industrial Statistics
4 Special Programmes
5 LDC and South-South Cooperation
6 Human Security / Post-crisis Rehabilitation
7 Partnerships with Organizations of the Private Sector and of Civil Society
8 Country-level and Regional Coherence

Source: "Programmes and Budgets: 2008-2009"

The review of UNIDO's projects' portfolio allows making the following important observations, which need to be considered when identifying the approaches that UNIDO will take to conducting impact evaluations:

1. UNIDO is dependent on voluntary contributions to finance its technical

cooperation activities. Therefore, the need for the availability of funding for conducting extensive impact evaluations needs to be shared by the financing partners. The review of project documents as well as evaluation reports indicates that the budget (as well as time) allocated within individual projects currently, is usually insufficient to allow carrying out impact evaluations on a systematic, and rigorous basis. And this goes not just for impact evaluations, but for routine evaluations. UNIDO (2007)[18] notes that "… Six reports (30%) recommended the inclusion of evaluation costs in the IP budget at the design stage. For larger programmes, inclusion of the same in project budgets was recommended". It could be argued that this would be important not only for large projects, but for all. The same publication also notes that "…four reports recommended more time for evaluations as standard practice. These reports hold that two weeks in the field were inadequate for in-depth independent evaluations'.

2. The portfolio is diverse, ranging from policy advice to very specialised and targeted assistance provided to SMEs. On the other hand, across the components of Major Program Areas, projects are very often composed of similar building blocks- *intervention modes*, as those are referred to in this paper: projects encompass, one or two, or all three of the following:
    - policy advice at government level,
    - capacity building of intermediate organizations, and
    - direct support to end-users (be it SMEs, communities or households).

3. Projects vary significantly in their costs: from small projects below Euro 100K to multi- million Euro projects. The latter, are however rare (like Montreal Protocol), and the majority of projects are medium to small size. The average project size is about Euro 500K[19].

4. Many of the projects, like the environmental projects, are expected to generate impact, that may take many years to be perceived or measured, while most of UNIDO projects span across 2-3 years.

5. Most of the projects do not have clear and agreed baselines, if at all, and the quality of monitoring data is far from sufficient to facilitate conducting quality impact evaluations[20]. Monitoring at UNIDO has focused much more closely on results at the project level, more specifically- outputs, and on implementation issues rather than progress toward achieving results: monitoring is rather a management instrument to keep activities on track. And while mechanisms appear to be in place to guide development of goals and results during project design, implementation, and reporting and individual projects have been assessed against their implementation performance as part of various reporting, there remains a large gap in the effectiveness of such project-level mechanisms in capturing results at the impact level.

[18] "Comparative review of lessons learned from 20 UNIDO Integrated Programmes", UNIDO, Evaluation Group, Vienna, 2007

[19] An estimate based on information on budget allocations and number of current projects, available on UNIDO intranet

[20] Ibid; and 'RBM Implementation Plan with milestones for the biennium 2008-2009', UNIDO, November 2007

6. While the outcome indicators are listed in the Programme and Budget document, there is no clarity about impact indicators across the three major TC Programmes.

7. This list of outcome indicators is too long and contains many conceptual duplications: differences between many indicators are in wordings only. There is significant room for reducing the list and introducing some standardisation in this list, which will allow for aggregation of results at impact level across portfolios of projects and thematic impact evaluations.

8. UNIDO oftentimes implements projects jointly with other donor agencies and assessing impact of only its contribution is next to impossible.

The last three points need further elaboration. Currently UNIDO is undergoing a major exercise whereby a Results Based Management (RBM) system is being implemented - so far on a piloting basis, with the introduction of the Results Reporting System (RRS) being one of the first tasks to be completed. This pilot stage of RRS already addresses these points. If introduced across all the projects UNIDO-wide, RRS will significantly enhance the opportunities for conducting impact evaluations at all levels (projects, portfolio/thematic, country, and UNIDO-wide).

Taken as a whole, the observations above indicate that results measurement within the UNIDO at impact level would pose a challenge at the beginning. UNIDO Evaluation Group will need to develop a strategy for impact evaluations both for short- term and long-term:

- For a short-term, the strategy will need to enable producing the most useful results *within the identified constrains*: It should, *inter alia,* include some of the following elements:
    - maximising the use of existing monitoring and evaluation data;
    - more collaboration with co-funding agencies in conducting joint evaluations;
    - conducting only a few selected impact evaluations as a start, on a piloting basis for different type of projects and with different methods, but with less quantitative emphasis.

- In the mid- to- longer time horizon, presumably the problem of lack of baseline and monitoring data will be gradually resolved, and allocating more financial resources from the project budgets for conducting baseline studies and impact evaluations will be a more widespread practice across UNIDO. This will allow for conducting more of rigorous, and, in particular, quantitative, impact evaluations.

It is important, however, to identify the most suitable approaches to impact evaluation early enough and develop a sound theoretical basis for deriving an understanding of UNIDO impacts and the possibilities to measure those.

Before proceeding to suggesting the evaluations approaches and methodologies that are most suitable for UNIDO, a review of the existing ones is needed. These are summarized in Chapter 4.

4. EXISTING APPROACHES TO EVALUATING THE IMPACT OF DEVELOPMENT PROJECTS

4.1 Existing approaches to Impact Evaluation

There are two main groups of methods to Impact Evaluation[21]

.

- *Quantitative impact evaluation:* analysis based on a representative survey of treatment group and a comparison group, preferably both before and after the intervention. There are a range of possible techniques under this heading;

- *Qualitative impact evaluation:* analysis based on participatory methods amongst beneficiaries.

In fact in some documents, *Quantitative impact evaluation*s are referred to as *rigorous impact evaluations.*[22] Other papers argue that qualitative impact evaluations, if carried out properly, could also be qualified as rigorous impact evaluations[23]. The position of this paper is that quantitative analysis provides for the most possible rigour in addressing attribution issues. At the same time, efforts should be in place to increase the rigour of qualitative impact evaluations in addressing attribution/contribution of the projects to achieving the observed results. At a minimum, qualitative analysis, even if not a full-blown participatory analysis, can help to provide valuable context.

There has been a rise in the use of *quantitative impact evaluation*. There has been however also an increased debate about the feasibility of using quantitative techniques for impact evaluations for many types of development interventions, and also depending on other constraints (data, time, etc). Hence the development community has seen recently an increased appreciation of the qualitative methods in impact evaluations, and the need to combine, when possible the two.

- *Theory-based (program logic) approach* helps to build the story around the intervention and understand why it worked or not, with an analysis tracing the log frame from inputs to outcomes, using a mix of methods to establish causal linkages. It is worth underlining that this is an *approach,* rather than a *method*, in contrast to quantitative and qualitative groups of methods. Moreover, theory based impact evaluations will ideally combine methods of both kinds.

In what follows below the existing impact evaluation methodologies are described briefly, grouped along the lines above. For an extensive and detailed guide with examples and strategies for mitigating common problems of particular designs see Baker (2000).[24]

[21] In this report Impact Evaluation refers to Ex-post Impact Evaluation, unless otherwise noted.

[22] WB/OED, "Monitoring and Evaluation: Some Tools, Methods & Approaches', Washington DC, 2004

[23] NONIE Subgroup 2, Impact Evaluation Guidance, Draft, 01/08

[24] Baker, Judy (2000) Evaluating the Poverty Impact of Projects World Bank. http://imagebank.worldbank.org/servlet/WDSContentServer/IW3P/IB/2000/08/19/000094946_00080705302127/Rendered/PDF/multi_page.pdf

4.1.1 Quantitative impact evaluation: experimental, quasi-experimental and non-experimental designs

a. Methods of Quantitative Impact Evaluation

The cornerstone of quantitative impact evaluation is data collection from a statistically representative sample using a structured questionnaire. It is strongly preferable that data are collected both before the intervention (baseline) and after (endline). A midterm survey is also an advantage. Data should be collected from both the affected population (the treatment group) and a comparison group. Project impact is then calculated as either a single difference (difference in outcome between project and control after the intervention), or double difference (the difference in the change in outcomes for the project and control before and after the intervention).

Selection of an appropriate comparison group is one of the main challenges in impact evaluation. They should be identical to the treatment group except that the latter receive the intervention and the former do not. In practice this is difficult to achieve for two reasons. First, beneficiaries of the intervention may be selected (or self-select) on the basis of certain characteristics. If these characteristics are observed then a comparison group with the same characteristics can be selected. But if they are unobserved then in principle only a randomized approach can eliminate selection bias. Second, the comparison group may be contaminated either by spillover effects from the intervention or a similar intervention being undertaken in the comparison area by another agency[25].

Many of these issues become more problematic when evaluations are conducted under real-world constraints. Some of the common issues include:

- Projects target *all* communities or subjects with particular characteristics, making it difficult to find close matches for a comparison group;
- When individuals or communities self-select into a program, it is difficult to identify factors determining the decision to participate and consequently to find a good comparison;
- Project and comparison groups may differ in terms of factors not covered in the survey *(omitted variables or unobservables)*.

As a result, random assignment is practiced in only a small proportion of development projects. There are a number of ways of addressing these problems within the broad 'experimental paradigm', such as quasi-experimental methods and non- experimental designs.

For most impact evaluations a quasi- experimental design must be used, employing different sampling procedures for the selection of the project and comparison groups. This has important implications for the analysis of project impacts, as post-intervention gain scores may be due to sampling bias (differences in the characteristics of the two groups), rather than to the effects of the project.

Box 3 describes the two main methodologies under randomised experiments.

[25] This section borrows from http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20215333~menuPK:451260~pagePK:148956~piPK:216618~theSitePK:384329,00.html#exp

Box 3. Experimental designs

Randomization

Experimental or randomized designs are considered the most robust of the evaluation methodologies. In theory, the control groups generated through random assignment serve as a perfect counterfactual, free from the troublesome selection bias issues that exist in all evaluations. In practice however this method is only applicable to a narrow range of the interventions supported by development agencies, due to several problems:

1. Randomization may be unethical owing to the denial of benefits or services to otherwise eligible members of the population for the purposes of the study.

2. It can be politically difficult to provide an intervention to one group and not another[1].

3. The scope of the intervention may rule out the possibility of selecting a control group such as with a nationwide program or policy change.

4. Individuals in treatment or control groups may change certain identifying characteristics during the experiment that could invalidate or contaminate the results.

5. It may be difficult to ensure that assignment is truly random.

6. Experimental designs can be expensive and time consuming in certain situations, particularly in the collection of new data.

Encouragement design

The *encouragement design* is a special case of an experimental design that can be used in situations with little control over subjects' compliance. It is particularly useful when one wishes to rigorously estimate the effect of some intervention that cannot itself be randomly administered to some and not others. The key idea is that instead of randomizing the application of the intervention itself, what is randomized is *encouragement* to receive the treatment. Encouragement typically takes the form of additional information or incentives. By randomizing encouragement and carefully tracking outcomes for all those who do and do not receive the encouragement, it is possible to obtain reliable estimates of both the encouragement and the intervention itself. *See "The Encouragement Design for Program Evaluation', A. Diamond and J. Hainmueller (Harvard University & IFC), 2007, http://www.ifc.org/ifcext/rmas.nsf/Content/WorkingPapers*

Quasi-experimental methods relax the probabilistic and population distribution conditions imposed by 'true' experimental research designs by shifting the emphasis from 'cause-effect' in temporal priority to 'association' between variables. See Box 4.

Non –experimental designs can be used when it is not possible to randomly select a control group, identify a suitable comparison group through matching methods or use reflexive comparisons. In such situations, program participants can be compared to non-participants using statistical methods to account for differences between the two groups. See Box 5.

As with quasi-experimental methods, these evaluation designs is relatively cheap and easy to implement since it can draw on existing data sources. However, it poses a number of difficulties. First, the reliability of results is often reduced as the methodology is less robust statistically. Second, the methodology has some statistical complexities. Third, although it is possible to partially correct for selection bias, full correction remains as a challenge.

Box 4. Quasi –experimental designs

Matching

Matching involves identifying non–program participants comparable in essential characteristics to participants on the basis of either a few observed characteristics or those that are known or believed to influence program outcomes. Matched comparison groups can be selected before project implementation (prospective studies) or afterwards (retrospective studies).

The main advantage of evaluations using matching methods is that they can draw on existing data sources and are thus often quicker and cheaper to implement. The principal disadvantages are that the reliability of the results is often reduced, as the methodology may not completely solve the problem of selection bias, and the matching methods can be statistically complex. When operating under real-world constraints it will often be necessary to rely on easily observable or identifiable characteristics. While this may expedite matters, it is important to keep in mind the potential for unobservable differences, to address these as far as possible through qualitative research, and to attach the appropriate caveats to the results.

The most widely used type of matching is **propensity score matching**, in which the comparison group is matched to the treatment group by using the propensity score (predicted probability of participation given observed characteristics). This method allows one to find a comparison group from a sample of non-participants closest in terms of observable characteristics to a sample of program participants. The potential problem is again related to unobservables. If they are constant over time, then their effects can be swept out by taking double difference estimates. But if they are time variant, or correlated with both selection and outcomes, then biased estimates will result. An issue when working under budget constraints is that the use of propensity scores will often require a larger sample to ensure that the best matching variables are identified.

Reflexive comparison

In a reflexive comparison, the counterfactual is constructed on the basis of the situation of program participants before the program. Thus, program participants are compared to themselves before and after the intervention and function as both treatment and comparison group. This type of design is particularly useful in evaluations of full-coverage interventions such as nationwide policies and programs in which the entire population participates and there is no scope for a control group.

There is, however, a major drawback with reflexive comparisons: the situation of program participants before and after the intervention may change owing to myriad reasons independent of the program, and unless they are carefully done, reflexive comparisons may not be able to distinguish between the program and other external effects, thus compromising the reliability of results. This design is widely used both because the elimination of the comparison group can cut data collection costs by up to 50% , and because there are many situations in which data are available for the project group (usually from project surveys and administrative records) but not for a comparison group. In WB/IEG (2006)[1] it is argued that these cannot be considered as quality impact evaluations. In Section 6 we describe particular cases when reflexive comparisons are acceptable as quality impact evaluations- when there is a straightforward causal link between the intervention and the result, provided that measurements are taken right before and after the intervention.

Box 4 (cont-d). Quasi – experimental designs

Regression discontinuity designs

In cases where a program is assigned using a clear threshold for eligibility comprised of one or more criteria, this program assignment rule can be used for evaluation. The basic idea is to compare individuals, communities or units just above the threshold and hence not eligible for the project (the comparison group) with those just below the threshold who are eligible (the treatment group). This procedure requires that the treatment rule is fairly enforced in practice, and the selection criteria are not subject to manipulation by potential beneficiaries.

Multiple comparison group designs.

When projects are implemented in different ways, or participants receive different combinations of services, it may be possible to use different comparison groups for different treatments.

Pipeline

The comparison group is defined as individuals, households or communities selected to participate in the project but who have not yet done so. Clearly the approach can only be used for activities which continue beyond the end of the project being evaluated.

Often large projects such as housing or community infrastructure are introduced in phases over several years and some beneficiaries will not begin to receive services until several years after the start of phase one. When there are no major differences between the characteristics of families or communities scheduled for each phase, the later phases can provide a good comparison group for the earlier phases. These procedures are also economical to use. However, project design and selection criteria must be carefully reviewed because there will often be systematic differences between the phases. For example, phase one may start with the poorest families or alternatively with the more centrally located or better-off areas, and in both of these cases the characteristics of communities in later phases are likely to be different.

Box 5. Non –experimental designs

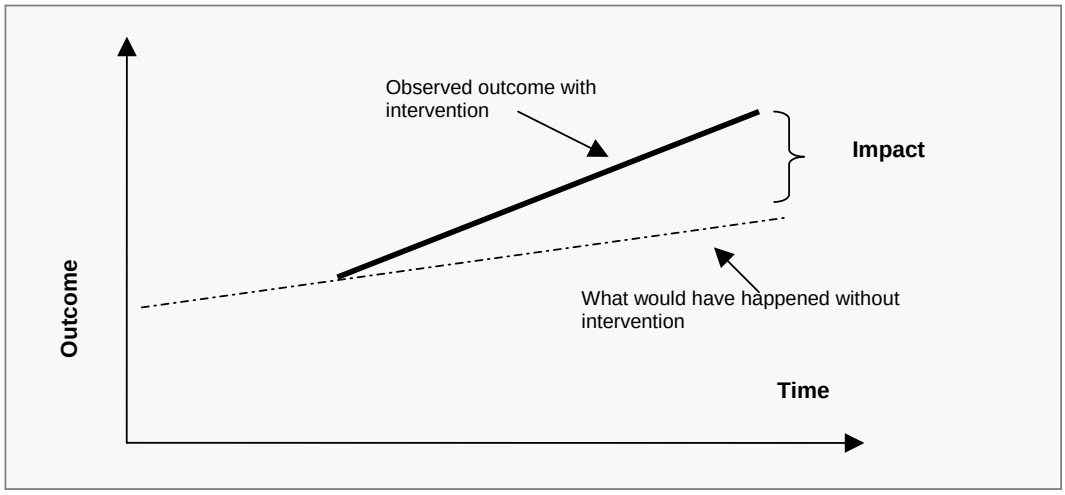
Regression based approach models the determinants of outcomes and possibly also models the determinants themselves. The approach has the advantage of flexibility – it does not lump different activities under the single heading of 'the intervention' – and automatically incorporates differing intensities of participation. It is only when the treatment is a simple, homogenous activity that dummy and mean comparison approaches are appropriate.

Instrumental variables method is one of the econometric techniques that can be used to compare program participants and non-participants correcting for selection bias. It consists of using one or more variables (instruments) that matter to participation but not to outcomes given participation. This identifies the exogenous variation in outcomes attributable to the program, recognizing that its placement may not be random but purposive. The instrumental variables are first used to predict program participation; then the program impact is estimated using the predicted values from the first equation.

b) Addressing attribution issues

Quantitative impact evaluations address attribution issues through explicitly quantifying the results observed *after the project* with a situation *in which the project would not have taken place – the counterfactual* (see Figure 1).

Figure 1 The Impact of the intervention



4.1.2 Qualitative impact evaluation

The key advantage of qualitative approaches is that they usually provide a much stronger context and so a good 'feel' for the intervention. They are less powerful in terms of the strength of causal inference they carry, however. But the latter statement has two caveats to bear in mind. First of all, for certain types of interventions, e.g. for projects on policy advice at national level, qualitative impact evaluations are the only feasible ones. And second, if carried out with the necessary care, it is possible to achieve greater rigour in their application.

a. Methods of qualitative impact evaluation

Participatory approaches

Participatory approaches are intended to empower beneficiaries by enabling them to shape decisions which affect their lives. Participatory research eschews the pre-set agenda implicit in the structured questionnaires used in the quantitative approach. Projects with a participatory orientation are likely to have participation in the design of their M&E system, which means that beneficiaries played a role in defining the indicators to be monitored and perhaps also in the monitoring. Participatory approaches are very likely to be the most suitable for process aspects of impact evaluations, and of course for process evaluations. Participatory evaluations also play an important role in furthering organisational change in institutions and projects: inclusive stakeholder consultations involve ownership of the assessment and facilitate the application of changes needed to be made.

Box 6. Commonly Used Participatory Tools

1. **Stakeholder Analysis** is the starting point of most participatory work. It is used to develop an understanding of the power relationships, influence and interests of the various people involved in an activity and to determine the who should participate and when
2. **Beneficiary Assessment** involves systematic consultation with project beneficiaries and other stakeholders to identify and design development initiatives, signal constraints to participation and provide feedback to improve services and activities
3. **Participatory Monitoring and Evaluation** involves stakeholders at different levels working together to identify problems, collect and analyse information and generate recommendations
4. **Participatory Rural Appraisal** is a planning approach focused on sharing learning between local people, both urban and rural and outsiders. It enables development of managers and local people to assess and plan appropriate interventions collaboratively.

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.

Participatory approaches can be used for[26]:

- Learning about local conditions and local people's perspectives and priorities to design more responsive and sustainable interventions;
- Identifying and trouble-shooting problems during implementation;
- Evaluating a project, program, or policy;
- Providing knowledge and skills to empower poor people.

Participatory methods allow to:

- Examine relevant issues by involving key players in the design process;
- Establish partnerships and local ownership of projects;
- Enhance local learning, management capacity, and skills;
- Provide timely, reliable information for management decision-making.

Disadvantages of these methods include:

- Sometimes they are regarded as less objective;
- They are usually weak in terms of delivering rigorous impact evaluations - that is identifying attribution for changes in development outcomes.
- They are time-consuming if key stakeholders are involved in a meaningful wayl
- There are concerns about the potential for domination and misuse by some stakeholders to further their own interests.

Costs and time required could be low to medium, varying greatly, depending on scope and depth of application and on how local resource contributions are valued.

[26] WB/OED, "Monitoring and Evaluation: Some Tools, Methods & Approaches', Washington DC, 2004

Outcome Mapping

Outcome Mapping is a methodology that focuses on outcomes as behavioural change. 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 (individuals, groups and organisations with whom the intervention interacts directly and with whom the intervention anticipates opportunities for influence) with new tools, techniques and resources to contribute to the development process.

Success Case Method

The Success Case Method 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 impacts. 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.

Most Significant Change

The Most Significant Change (MSC) technique is a form of participatory monitoring and evaluation. 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.

Rapid Participatory Assessment (RPA) Methods

Rapid Participatory Appraisals[27] are quick, low-cost ways to gather the views and feedback of beneficiaries and other stakeholders, in order to respond to decision-makers' needs for information.

Advantages of these methods are: low –to – medium costs (depending on the scale of methods adopted.); speed (four to six weeks, depending on the size and location of the population interviewed and the number of sites observed.); and flexibility to explore new ideas.

Disadvantages include (a) the fact that findings usually relate to specific communities or localities, thus creating difficulties in generalizing from findings; and more importantly, (b) proneness to being less valid, reliable, and credible than formal surveys.

[27] WB/OED, "Monitoring and Evaluation: Some Tools, Methods & Approaches', Washington DC, 2004

Box 7. Rapid Participatory Assessments Methods

- **Key Informant Interviews:** series of open ended questions posed to individuals selected for their knowledge and experience in a topic of interest. Interviews are qualitative, in-depth and semi-structured. They reply on interview guides that list topics or questions.
- **Focus Group Discussions.** A facilitated discussion among carefully selected participants with similar backgrounds. Participants may be beneficiaries or program staff. Facilitator uses a discussion guide. Note-takers record comments and observations.
- **Community Group Interviews** – a series of questions and facilitated discussion in a meeting open to all community members. The interviewer follows a carefully prepared questionnaire.
- **Direct Observations.** Use of a detailed observation form to record what is seen and heard at a program site. The information may be about ongoing activities, processes, discussions, social interactions, and observable results.
- **Mini Surveys.** A structure questionnaire with a limited number of close-ended questions that is administered to 50-75 people. Selection of respondents may be random, or 'purposive" (Interviewing stakeholders at locations such as a clinic for a health care survey).

.

RPA methods can be used for:

- Providing rapid information for management decision-making, especially at the project or program level.
- Providing qualitative understanding of complex socioeconomic changes, highly interactive social situations, or people's values, motivations, and reactions.
- Providing context and interpretation for quantitative data collected by more formal methods.

Case Studies

A case study is a method for learning about a complex instance, based on a comprehensive understanding of that instance obtained through its extensive description and analysis as a whole and in its context. Case studies are appropriate for determining the impacts of program or projects and reasons for success or failure. The method is often used in combination with others, such as sample surveys. This method should not be used when there is a great diversity among sites and projects[28]. Analysis of case study data is generally extensive. Multiple methods could be used in a case study (see Box 8):

WB IEG uses case-study-portfolio reviews, for example, in both sector studies and country assistance evaluations. At the very least these reviews describe trends in the size and composition of the portfolio (for example by region or subsector) and an analysis based on IEG ratings. But they may be more comprehensive, studying the evolution of project objectives and design over time, in which case this method becomes more akin the theory based approach (discussed in Section 4.1.3).[29]

[28] L. Morra and A. Friedlander, "Case Study Evaluations', WB/OED, 1999

[29] In the portfolio review in IEG's study of social funds, design features such as the use of outreach, targeting mechanisms, community contribution, and maintenance arrangements were codified based on a desk review supplemented with task manager interviews. The theory-based approach used in this evaluation suggested the importance of some features for successful operations, such as comprehensive outreach and community-level arrangements for maintenance. Four country case studies were undertaken to provide material with which to test the importance of these factors, and their results

Box 8: Case Study Methods

	Technique	Methodology
1	**Extensive analysis**	Analysis of multiple types of data sources, such as
		• Interviews with all relevant sources
		• Observation over time
		• Participant observation
		• Documents
		• Archives
		• Physical Information
2	**Analysis via triangulation of data**	sis through:
		• Pattern matching
		• Explanation building
		• Thematic review
3	**Comparison of evidence for consistency**	sis through techniques such as:
		• Matrix of categories
		• Graphic data displays
		• Tabulation of even frequencies
		• Chronological time series/ordering

Source: L. Morra and A. Friedlander, "Case Study Evaluations', WB/OED, 1999

Abadie *et al* [30]investigates the application of synthetic control methods to comparative case studies. It discusses the advantages of these methods and apples them to study the effects of Proposition 99, a large-scale tobacco control program that California implemented in 1988. Given that many policy interventions and events of interest in social sciences take place at an aggregate level (countries, regions, cities, etc.) and affect a small number of aggregate units, the potential applicability of synthetic control methods to comparative case studies is large, especially in situations where traditional regression methods are not appropriate.

b. Addressing attribution

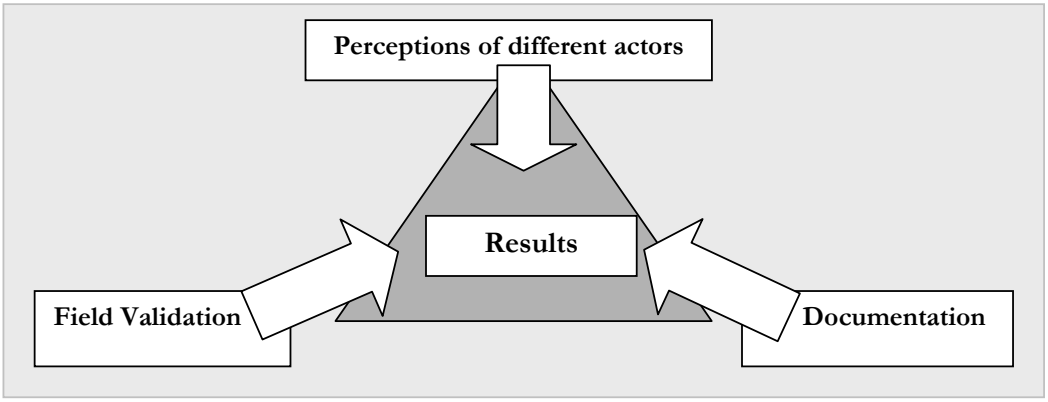
The key analysis technique used to address attribution in qualitative impact evaluations is triangulation, which involves developing the reliability of the findings through multiple data sources of information (see Figure 2), bringing as much evidence as possible into play (from different perspectives) in the assessment of hypotheses and assumptions.

The validity of the findings, especially on trying to determine cause and effect, is derived from agreement among the types of data sources, together with the systematic ruling-out of alternative explanations and the explanation of "outlier results". Examining consistency of evidence across different types of data sources is a means of obtaining verification.

helped identify the most critical design features. The portfolio review showed how social funds had evolved, acquiring better design features over time.

[30] A. Abadie, A. Diamond, and J. Hainmueller Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program , July 2007, Harvard University, Cambridge, MA USA

Figure 2 Method of Triangulation



There are particular strategies for making such comparisons, such as pattern matching, explanation building, and thematic reviews. They involve techniques such as graphic data displays, tabulations of event frequencies, and chronological or time series orderings.

4.1.3 Theory based impact evaluation

a) Main concepts and steps

The best approaches to impact evaluation combine quantitative and qualitative methods. The theory-based evaluation (TBE - hereafter) approach attempts to do just that. Or in other words, it is not method-specific, it is an approach which needs to incorporate robust evaluation methods, quantitative or qualitative, or, ideally, both.

The interest in theory based approaches to impact evaluation is growing, with some agencies, like GEF, taking it as the framework for all their impact evaluations.

A TBE design is one in which the analysis is conducted along the length of the causal chain from inputs to impacts. Many impact evaluations concern themselves only with the final link in the chain: final outcomes/impact. But to do this, is often to lose the opportunity to learn valuable policy lessons about why an intervention has worked (or not), or which bits have worked better than others.

Applying a TBE approach requires mapping out the channels through which the inputs provided by the intervention are expected to affect the outcomes specified in the objectives. In many cases this analysis will already be contained in the project log frame, which may also specify indicators at the various levels. Thus the monitoring system can be a useful source of analysis of process aspects of the intervention – though usually supplemented by qualitative data[31].

TBE has similarities to the LogFrame approach and is based on it.

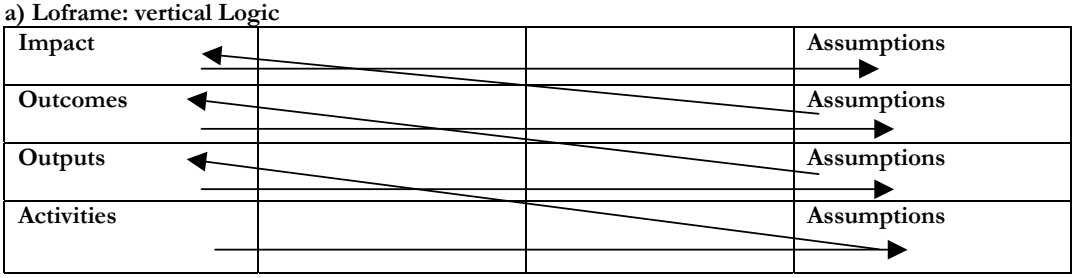
[31] H. White. "Impact evaluation: an overview and some issues for discussion', note prepared by the Independent Evaluation Group of the World Bank, in collaboration with the DAC Secretariat, 4th meeting of the DAC Network on Development Evaluation, 30 – 31 March 2006.

b) From LogFrame to Results Chain

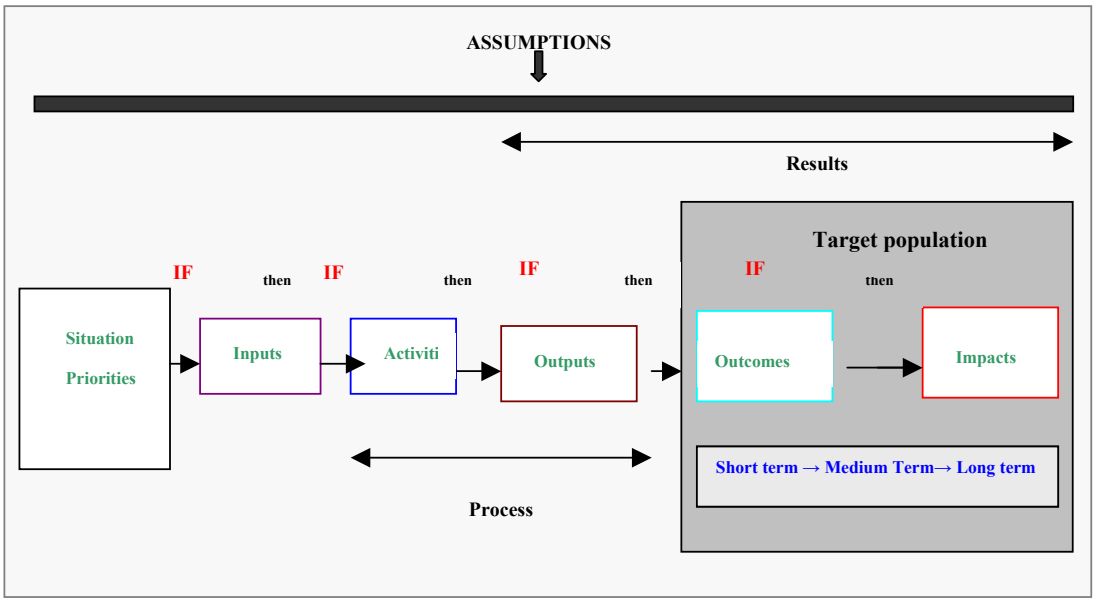
The logical framework or LogFrame[32] is an analytical tool used commonly to plan, monitor, and evaluate projects. The LogFrame helps to clarify objectives of any project, program, or policy. It aids in the identification of the expected causal links - the "program logic" - in the following results chain: inputs, processes, outputs (including coverage or "reach" across beneficiary groups), outcomes, and impact. It leads to the identification of performance indicators at each stage in this chain, as well as risks which might impede the attainment of the objectives. The LogFrame is also a vehicle for engaging partners in clarifying objectives and designing activities. During implementation the LogFrame serves as a useful tool to review progress and take corrective action. Figure 3 presents a generic logic model in a schematic way – as a starting point for developing a Results Chain.

Figure 3: Generic logic model

a) Loframe: vertical Logic



b) Logframe: extended and schematic



[32] The logframe was originally developed by the United States Department of Defense, and adopted by the United States Agency for International Development in the late 1960s. Since then, it has been applied and modified by many bilateral donors, including Germany, the United Kingdom, the European Union, Canada, and Australia. Development of Logframes for UNIDO project is mandatory according to its TC Guidelines.

It is very important to underline the role of assumptions in developing logframes for a project, as they are cornerstones for developing the results chains later. Every next step in the causal chain is expected to materialize only when certain set of assumptions holds true. The latter needs to be explicitly described and monitored to enable to make conclusions about the reasons behind the observed results: whether these were related to the assumptions or project activities.

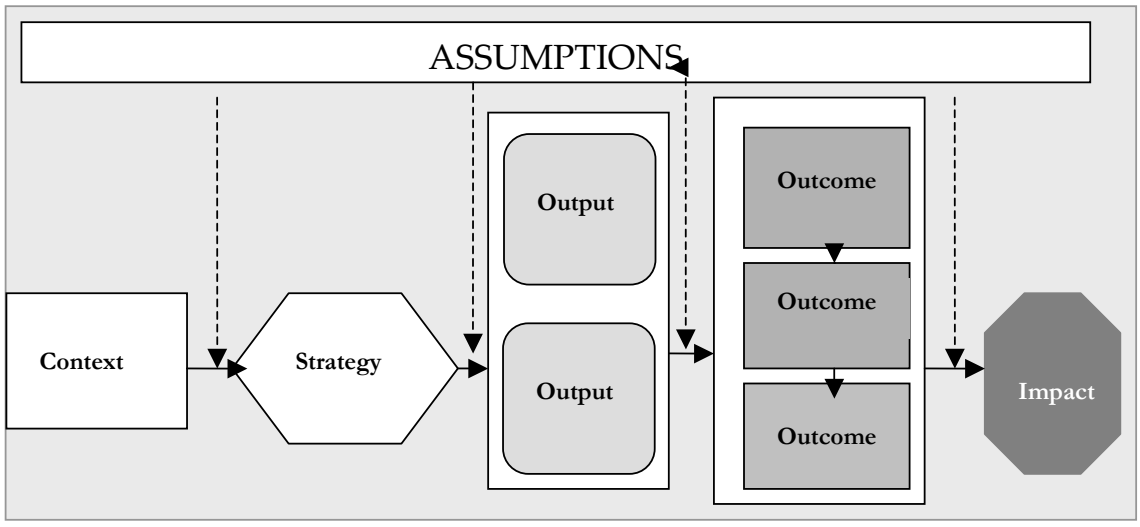
c) Results Chains as a starting point for TBE

Results chain goes a step further from the Logframe: it allows for a more in-depth understanding of the workings of a program or activity—the "program theory" or "program logic". In particular *it does not assume simple linear cause-and effect relationships.* Figure 4 describes a Generic Results Chain. This allows to account for in depth analysis of all cause and effects relationships. In general, results chains are a tool that clarifies assumptions about how project activities contribute to achieving project targets. In particular, a results chain[33]:

- Is a diagram of a series of "if…then" statements ("causal");
- Defines how we think a project strategy or activity is going to contribute to reaching a goal;
- Focuses on the achievement of results – not the execution of activities;
- Is composed of assumptions that can be tested.

There is a series of criteria that must be fulfilled in order to fully develop a results chain. They need to be results oriented, simple, connected in a "causal" manner, demonstrate changes and ideally, show the weights (strength and importance) of the causal links (in line with the principle of 'relativity').

Figure 4 Generic Results Chain



[33] GEF Evaluation Office/Foundations for Success, "GEF Impact Evaluation: Final Report on a Proposed Approach to GEF. Impact Evaluation. Impact Evaluation Information Document No. 2'

The "story' at a project level should try to answer the following questions:

- What was the intended series of cause and effect linkages (mechanisms), which were expected to generate impacts?
- What were the key features of the project context, which interacted with these linkages to determine results achieved (outputs, outcomes, and impacts)?
- How did the project respond to its specific context to generate results?
- How do impacts at the project level relate to the overall objectives of UNIDO?
- Are additional results from the project anticipated in the future, and if so, to what extent?
- Are project results likely to be sustained?
- What does the project tell us about the underlying theory of change on which the intervention was based?

TBEs critically examine the links in the causal/results' chain. There can be missing links if the project design missed some key determinants at the next level it has sought to influence. There may also be weak links in the chain as a result of poor implementation. The links are analyzed through a combination of quantitative and qualitative argument: in some cases statistical analysis may be used to confirm a link, but in others a more qualitative case for plausible association may be made.

The premise of TBEs is that programs and projects are based on explicit or implicit theory about how and why a program will work. The evaluation would then be based on assessing each theory and assumption about a program. The approach puts emphasis on the responses of people to program activities. Theories direct the evaluator's attention to likely types of near-term and longer-term effects.

Among the advantages of TBE are[34]:

- they assess the validity of the theory of change adopted (implicitly or explicitly) and the interaction between the specific local circumstances and the general principles expected to generate the desired change;
- they help to explain how and why effects occurred. If events work out as expected, the evaluation can say with a certain confidence how the effects; were generated. By following the sequence of stages, it is possible to track the microsteps that led from program inputs through to outcomes;
- the theoretical model can clarify the assumptions upon which the intervention was based and also incorporate opposing theories, which may be applied to the same situation;
- they assist in identification of unintended side-effects of the program;
- they help in prioritising which issues to investigate in greater depth, perhaps using more focused data collection or more sophisticated M&E techniques;
- they enable detailed examination of the nature of linkages between a complex set of causes and a set of effects.

If incorporated in project designs (discussed in Section 5.3.1), this approach- very similar then to developing the objective (or, as sometimes referred to- problem tree) - can

[34] WB/OED, "Monitoring and Evaluation: Some Tools, Methods & Approaches', Washington DC, 2004 and Baker, Judy (2000) Evaluating the Poverty Impact of Projects World Bank.

provide early indications of program effectiveness during project[35] implementation. If there are breakdowns during implementation, it is possible to fix them along the way.

The shortcomings of the approach are similar to many of the other approaches. In particular:[36]

- identifying assumptions and theories can be inherently complex;
- evaluators may have problems in measuring each step unless the right instruments and data are available;
- problems may be encountered in testing the effort because theory statements may be too general and loosely constructed to allow for clear-cut testing;
- there may be problems of interpretation that make it difficult to generalise from results (see Weiss 1998[37]);.
- drawing results' chains can easily become overly complex if the scale of activities is large or if an exhaustive list of factors and assumptions is assembled;.
- stakeholders might disagree about which determining factors they judge important, which can be time-consuming to address.

Costs associated with employing a TBE approach to impact evaluation are around medium, depending on the depth of analysis and especially the depth of data collection undertaken to investigate the workings of the program. Time required can vary greatly, depending on the depth of the analysis, the duration of the program or activity, and the depth of the M&E work undertaken.

When drawing results chains it is important to remember a few important issues (see Figure 5):

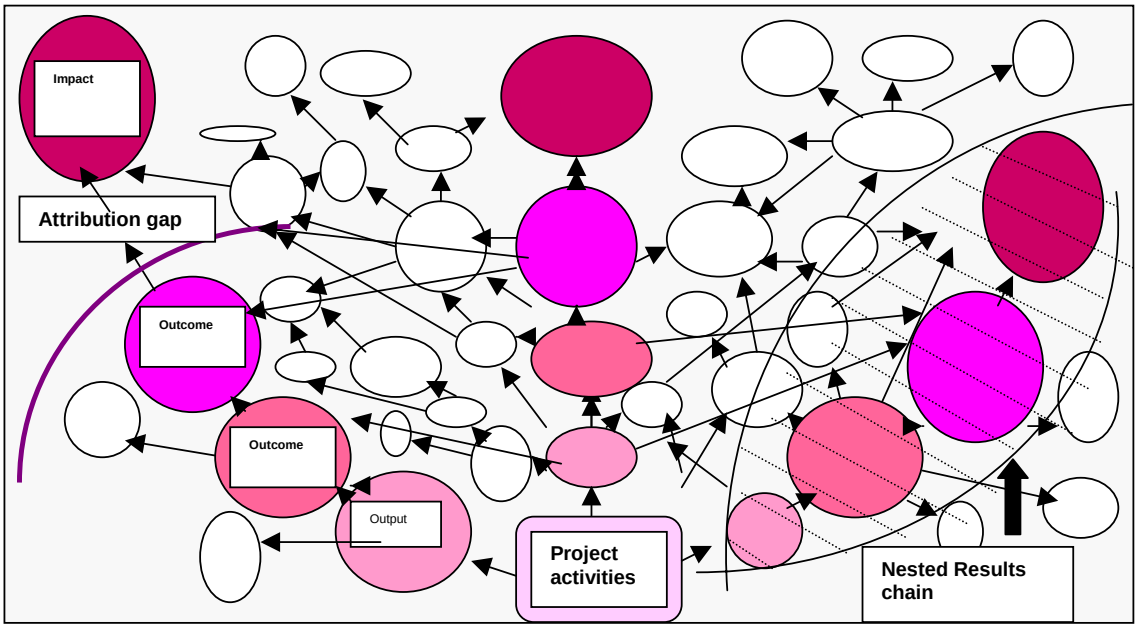
- The results chain for a project will consist of "nested" results chains along the main vectors along which the project achieves its results. For example, a capacity building project for an enterprise will work through: the development of its human resource potential (and thus productivity, performance, etc), its modernisation (thus affecting competitiveness, networks, products' base, etc), and so on.
- There will be an infinite number of causal links that could be drawn as existing for a project. It is important to define the limits, i.e. the main results that we are interested and will be measuring.

[35] GEF Evaluation Office/Foundations for Success, "GEF Impact Evaluation: Final Report on a Proposed Approach to GEF. Impact Evaluation. Impact Evaluation Information Document No. 2'

[36] WB/OED, "Monitoring and Evaluation: Some Tools, Methods & Approaches', Washington DC, 2004 and Baker, Judy (2000) Evaluating the Poverty Impact of Projects World Bank.

[37] Carol H. Weiss (1998) Evaluation: Methods for Studying Programs & Policies 2nd edition. Prentice Hall

Figure 5 Nested Results chains and project boundaries



There are two methodological steps that are at the heart of theory-based evaluation: theory reconstruction and theory testing.

d) Reconstructing theories of change

A common interpretation of an intervention theory is that it starts out from a systematic representation of the expectations and assumptions held by intervention staff and decision makers[38].

There are several methodologies available for reconstructing intervention theories. A methodology called policy-scientific approach (Leeuw, 2003)[39] is based on a five-step procedure to uncover the main assumptions based on documents and interviews. Klein Haarhuis and Leeuw (2004)[40] for example apply this approach in an evaluation on the World Bank anti-corruption program.

Carvalho and White (2004)[41] use a variation of this approach to analyze the impact of the World Bank social fund program. They show how for different assumptions, so-called 'anti-theories' can be defined, for example embodying the assumptions of opponents of particular types of interventions. The subsequent assessment process

[38] J. Vaessen and D. Todd, Methodological Challenges in Impact Evaluation: The Case of the Global Environment Facility (GEF), University of Untwerp, Institute of Development Policy and Management, 2007

[39] Leeuw, F.L. (2003) "Reconstructing Program Theories: Methods Available and Problems to be Solved", American Journal of Evaluation 24(1): 5-20.

[40] Klein Haarhuis, C.M. and F.L. Leeuw (2004) "Fighting Governmental corruption: The New World Bank Programme Evaluated", Journal of International Development 16: 547-561

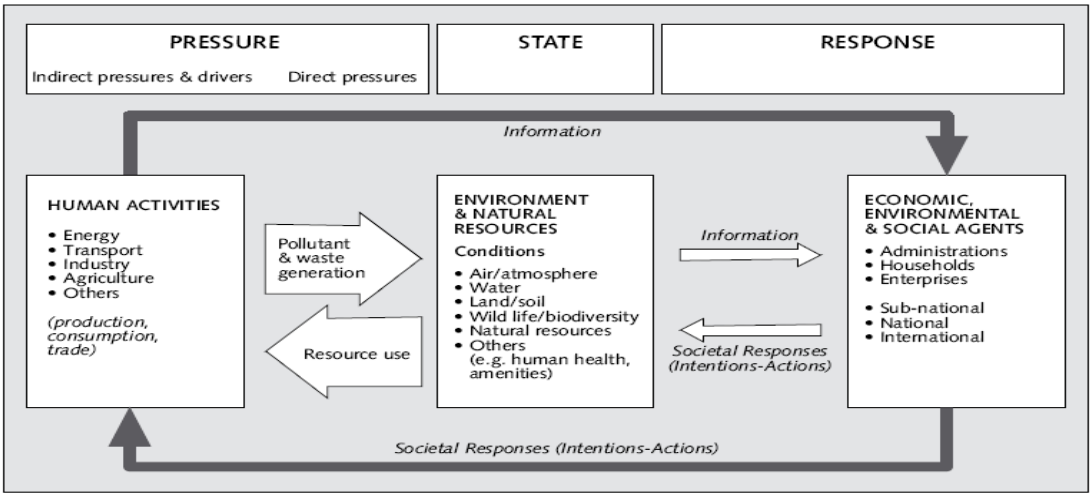
[41] S. Carvalho and H. White "Theory-Based Evaluation: The Case of Social Funds', the World Bank, Washington DC, 2004

serves the purpose of adjudicating between the rival theories and ultimately arriving at better explanations of processes of change leading to impact.

The Pressure-State-Response model (PSR) developed by the OECD in the 1990s (OECD, 2003)[42] is yet another variation. It is described in Figure 4. PSR analysis complemented by case studies and historical narrative approach will assist in examining the linkages between interventions, outcomes and impacts.

This seems to be particularly relevant for UNIDO, from the available information it seems that intentions and assumptions are only in part made explicit in formal documents (such as formal logical frameworks) and thus require further reconstruction Such an approach could also be used to focus on the evolution of a particular UNIDO intervention or a series of UNIDO interventions in a particular region/country, allowing for a detailed understanding of the complex interactions between interventions and economic, business, social, institutional and environmental contexts to develop a detailed picture of the induced changes and impacts.

Figure 6 The Pressure-State-Response Model



Source: OECD (2003).

d) Testing the intervention theory

The specific methods and approaches that evaluators can use to test the intervention theory are almost endless. The links are analysed through a combination of quantitative and qualitative argument: in some cases statistical analysis may be used to confirm a link, but in others a more qualitative case for plausible association may be made. In this sense, theory based evaluation is not method-specific. For example, an intervention theory can constitute the basis for a quasi-experimental evaluation.

An example of theory-based technique was piloted by the Operations and Evaluation Department of the World Bank to evaluate the impact of social investment funds on community-level decision making processes, traditional power structures and relationships, and community capacity, trust, and well-being. This was based on the theory that priority groups can effectively implement a project and operate and maintain the investment created by the project. A set of main assumptions and sub-assumptions has been set out and tested using existing household survey data, as well as specially

[42] OECD (2003) OECD Environmental Indicators: Development Measurement and Use, OECD, Paris.

designed survey instrument for a smaller sample, and focus groups and other PRA techniques. The information from each of these data sources was triangulated in the analysis.

While attribution is usually addressed at the level of final levels of the results chain, what happened across the results' chain is very important in answering that question. In theory-based evaluation exercises the evaluator relies primarily on logical argumentation, by carefully tracing all the assumptions underlying the theory (from inputs to outputs to impacts)[43]. Depending on the type of assumption, different sources of evidence come into play. Whenever quantitative, hard evidence is not possible to come up with, qualitative techniques are used.

For the purpose of impact evaluation it is useful to distinguish process theory and impact theory (see Figure 2.b, also see Rossi et al., 2004)[44]. The former refers to the assumptions and expectations underlying the processes of inputs leading to outputs, while the latter concerns the assumptions regarding particular outputs inducing processes of change resulting in final impacts. This differentiation is important in order to determine whether an observed lack of change is (mainly) due to problems of implementation, often referred to as implementation failure, or whether the concept of intervention (the idea that particular outputs lead to desired impacts) is fallacious, which is called theory failure .

As a result, in impact evaluation evaluators should have at their disposal substantial data about intervention outputs and the implementation process producing these outputs as a basis for further analysis. Only then can the more complex question of attribution in impact theory, i.e. the interaction between intervention outputs and external variables, the (potentially) complex and diffuse causal chain linking outputs to impacts, be addressed. TBE is best suited to address this challenge.

Indeed, using experimental and quasi-experimental studies most often enables receiving the best possible robust estimates of attribution. 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 counterfactual – 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, as was demonstrated in Section 4.1.

Some types of interventions lead directly to specific impacts which are unlikely to have occurred in the absence of the intervention, for example the introduction of a specific agricultural technology. In these cases we can talk about 'causal attribution', i.e. when:

- 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');

[43] J. Vaessen and D. Todd, Methodological Challenges in Impact Evaluation: The Case of the Global Environment Facility (GEF), University of Antwerp, Institute of Development Policy and Management, 2007

[44] Rossi, P.H., M.W. Lipsey and H.E. Freeman (2004), Evaluation – A Systematic Approach, Seventh Edition, Sage Publications, Thousand Oaks

- 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 difficult 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. 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:

- 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' characterized by large numbers of unknown variables and unknown causal connections between variables, by interactions, feedbacks and nonlinear relationships, and high sensitivity to small perturbations.

Mayne (2001) suggests a 5 step approach to contribution analysis, as in Box 9>

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.

Box 9 Contribution Analysis

Step 1. Develop the results chain	Develop the program theory model/program logic/results chain describing how the program is supposed to work. Identify as well the main external factors at play that might account for the outcomes observed. This program theory should lead to plausible association between the activities of the program and the results sought. Some links in the results chain will be fairly well understood or accepted. Others will be less well understood or subject to explanations other than that the program was the "cause". In this way you acknowledge that attribution is indeed a problem.
Step 2. Assess the existing evidence on results.	The results chain should provide a good idea of which intended results (outputs, intermediate and end outcomes) could be measured. What evidence (information from performance measures and evaluations) is currently available on the occurrence of these various results? The links in the results chain also need to be assessed. Which are strong (good evidence available, strong link or wide acceptance) and which are weak (little evidence available, weak logic or little agreement among stakeholders)?
Step 3. Assess the alternative explanations	Outcomes by definition are influenced by the action of the program but also by external factors- other programs, as well as social and economic factors. In addition to assessing the existing evidence on results, there is a need to explicitly consider the extent of influence these external factors might have. Evidence or logical argument might suggest that some have only a small influence and that others may have a more significant influence on the intended results
Step 4. Assemble the performance story	With this information you will be able to set out the your performance story of why it is reasonable to assume that the actions of the program have contributes (in some fashion, which you may want to try and characterize) to the observed outcomes. How credible is the story? Do reasonable people agree with the story? Does the pattern of results observed validate the results' chain? Where are the main weaknesses in the story? There always will be weaknesses. These point to where additional data or information would be useful In getting additional evidence it is not possible (at least for now), that this is the most you can say about the extent to which the program has made a difference.
Step 5 Seek out the additional evidence	To improve your performance story you will need additional evidence. This could involve information on both the extent of occurrence of both specific results in the results chain and the strength of certain links in the chain.
Step 6 Revise and strengthen the performance story	With the new evidence you should be able to build a more credible story, one that a reasonable person will be more likely to agree with. It will probably not be foolproof, but will be stringer and more credible.

Source: John Mayne, "Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly', The Canadian Journal of Program Evaluation Vol. 16 No. 1 Canadian Evaluation Society, 2001

4.2 Impact Evaluation under data, budget and time constraints

4.2.1 Constrains and tradeoffs in conducting Impact Evaluations.

It was already discussed in this document that currently UNIDO faces a number of challenges in conducting impact evaluations- mainly budget and data constraints

Randomized evaluation designs, involving the collection of information on project and control groups at two or more points in time, provide the most rigorous statistical analysis of project impacts and the contribution of other factors. But in practice it is rarely possible to use these designs, and the reasons are often related to costs (along with methodological problems and/or ethical constraints).

- A number of World Bank impact evaluations have ranged from $200,000 - $900,000 depending on program size, complexity and data collection.
- They can take up to two years or more.

Blomquist (2003)[45] provides examples of costs of impact evaluations (Annex 2) of social development projects, from which it follows that the average cost of these was about 350K USD, with 85% of this amount spent on average on data collection, and 15% on data analysis. Thus most impact evaluations use less expensive and less rigorous evaluation designs.

WB/OED (2004)[46] describes the average costs of some of the main evaluation techniques (see Box 10). It is important to note however that the ranges given are quite large depending on the size and complexity of the project, sample – in case of surveys, extent of using international consultant time, etc. For example, impact evaluations can be conducted for significantly less that $100,000 and in some cases for as little as $10,000 - $20,000. This will come, however, at a cost of threats to validity. In the list of most common impact evaluation techniques, listed in Bos 10, each successive model sacrifices methodological rigor, in return from which there are significant reductions in cost and time requirements.

Budget, time or data constraints can act as a disincentive to conduct rigorous evaluations. Conducting impact evaluations under these constraints increases the difficulty of dealing with four sets of threats to the quality of the design and the validity of the conclusions.

- Threats to overall quality of the evaluation design and implementation
- Threats to statistical analysis
- Theoretical coherence and adequacy of the counterfactual
- Generalisability of findings

While these four sets of *threats to the validity of evaluation conclusions* can affect all evaluations, they are more difficult to manage when working under real-world constraints.

[45] John Blomquist, "Impact Evaluation of Social Programs: A Policy Perspective', *Social Protection Human Development Network,* 2003

[46] WB/OED, "Monitoring and Evaluation: Some Tools, Methods & Approaches', Washington DC, 2004

Box 10. Features of the main models of evaluation

Model	Design	Example	Indicative cost and time
1. Randomized pre-test post-test evaluation.	Subjects (families, schools, communities etc) are randomly assigned to project and control groups. Questionnaires or other data collection instruments (anthropometric measures, school performance tests, etc) are applied to both groups before and after the project intervention. Additional observations may also be made during project implementation.	Water supply and sanitation or the provision of other services such as housing, community infrastructure etc where the demand exceeds supply and beneficiaries are selected by lottery. Example: Bolivia Social Fund.	1-5 years depending on time which must elapse before impacts can be observed. Cost can range from $50,000 - $1million depending on the size and complexity of the program being studied.
2. Quasi-experimental design with before and after comparisons of project and control populations.	Where randomization is not possible, a control group is selected which matches the characteristics of the project group as closely as possible. Sometimes the types of communities from which project participants were drawn will be selected. Where projects are implemented in several phases, participants selected for subsequent phases can be used as the control for the first phase project group.	These models have been applied in World Bank low-cost housing programs in El Salvador, Zambia, Senegal and the Philippines.	Cost and timing similar to Model 1.
3. Ex-post comparison of project and non-equivalent control group.	Data are collected on project beneficiaries and a non-equivalent control group is selected as for Model 2. Data are only collected after the project has been implemented. Multivariate analysis is often used to statistically control for differences in the attributes of the two groups.	Assessing the impacts of micro-credit programs in Bangladesh. Villages where micro-credit programs were operating were compared with similar villages without these credit programs.	$50,000 upwards. The cost will usually be one third to one half of a comparable study using Models 1 or 2.
4. Rapid assessment ex-post impact evaluations.	Some evaluations only study groups affected by the project while others include matched control groups. Participatory methods can be used to allow groups to identify changes resulting from the project, who has benefited and who has not, and what were the project's strengths and weaknesses. Triangulation is used to compare the group information with the opinions of key informants and information available from secondary sources. Case studies on individuals or groups may be produced to provide more in-depth understanding of the processes of change.	Assessing community managed water supply projects in Indonesia.	$25,000 upwards (the Indonesia study cost $150,000). Some studies are completed in 1-2 months; others take a year or longer.

Source: WB/OED, "Monitoring and Evaluation: Some Tools, Methods & Approaches', Washington DC, 2004

When resources are constrained, trade-offs will almost always have to be made on how resources are to be used. Some of these involve saving money by, for example, hiring less expensive interviewers or reducing the number and depth of case studies; while others involve decisions on whether to invest scarce resources to increase sample size, improve the coverage and quality of the sampling frame or reduce non-response rates by requiring more revisits to households.

Oldsman and Hallberg (2002)[47] suggest that significance of investments made in a particular program/portfolio should be one of the key guiding principles in choosing the

[47] E. Oldsman and K.Hallberg,, "Framework for Evaluating the Impact of Small Enterprise Initiatives", Nexus Associates, Inc., 2002, Washington DC

appropriate method (along with the availability of time and data), the robustness of which diminish from experimental studies to quasi-experimental, further to '*before and after studies'* and further down to participants' judgment. .

4.2.2 Quality Impact Evaluation under data, budget and time constraints

Despite the fact that major challenges arise when evaluators must work under serious budget, time and data constraints, evaluators are constantly asked to address important operational and policy questions regardless. In addition many evaluations are required to:

- Assess the distribution of benefits among different sectors of the target population;
- Identify factors influencing the magnitude and distribution of the impacts;
- Assess the sustainability of impacts over time.

While these trade-offs may be inevitable, framing them in the context of what dimension they are likely to compromise the results is a useful way to inform the decisions. There may also be ways to lessen the constraints, and there also may be things which should never be compromised.

IEG/WB (2006) [48] provides advice to those planning an impact evaluation, so that they can select the most rigorous methods available within the constraints they face. It also clarifies the nature of trade-offs between evaluation rigor and the budget, time and data which are available for an impact evaluation.

The challenge for the evaluator and the client is to decide whether it is possible to conduct a quality impact evaluation under the real-world constraints, and to select the strongest possible design within the particular set of budget, time and data constraints. For example: at what point does the sample become too small, or too limited in its coverage to permit sound statistical analysis? What are the criteria for assessing the adequacy of secondary data for estimating baseline conditions? And when is it possible to construct a valid counterfactual in the absence of a baseline study?

As an illustration, the following designs although weaker than an ideal randomized design, could still ensure a reasonable degree of analytical rigor:

- Pre-and post-intervention project group and post-intervention comparison group (allowing for 25% data collection cost-saving):

- Post-intervention project and comparison groups with no baseline data. (cost savings of up to 50%) .

As a comparison, a design, with Post-intervention project group without baseline data or a comparison group, is a very weak design and although it is widely used to estimate project effects, cannot be considered as producing rigorous quantitative estimates of

[48] This Section draws heavily from "Conducting quality impact evaluations under budget, time and data constraints', IEG/WB, 2006

project impact[49].

IEG/WB (2006) discusses a number of ways to address the effects of budget, time and data constraints on the validity of the evaluation conclusions. [50](see Box 11)

A number of other organizations have also come up with recommendations on how to address the need to conduct impact evaluations under constrains. The Shoestring Evaluation approach is one of the more known ones, mostly in line with the recommendations above. .

There are limits however to the extent to which costs could be saved in conducting impact evaluations. And it needs to be realised that allocation of sufficient resources for impact evaluations by the funding agencies (and in this particular case - UNIDO and its Board) is a necessary pre-condition for conducting quality impact evaluations.

[49] Frequently this design is used when the evaluator is operating on an extremely tight budget (sometimes as little as $10,000) and may only be able to spend a few weeks in the field. Estimates of change (impacts) are based on a combination of qualitative data such as recall, key informants, focus groups and participatory group techniques such as participatory rapid assessment (PRA), project records and secondary data from public service agencies (e.g., ministry of health or education), censuses and other government data. Secondary data are only used to obtain global comparisons of the project communities with similar areas, and not for household-level analysis. Depending on the nature of the design, costs can be reduced by 75% or more.

[50] "Conducting quality impact evaluations under budget, time and data constraints', IEG/WB, 2006

Box 11 Ways to address the effects of budget, time and data constraints on the validity of the evaluation conclusion

a) Strengthening the overall quality of the evaluation design

- Even when operating under budget and time constraints, it is always important to allow sufficient time to meet with clients and key stakeholders to understand *their* information needs, deadlines and constraints. Any design decisions on ways to save money or time must be made in consultation with clients so that they fully understand and accept the trade-offs involved.

- When time is the main constraint, it is possible to get off to a quick start by organizing video conferences with local agencies and researchers and by commissioning preparatory studies to be completed *before* the arrival of foreign consultants.

- Consider the cost implications of how selection bias and instrumental variables are addressed. Working with project staff to address selection bias during project design will often be more economical than conducting costly surveys during the post-implementation evaluation to match participants and comparison groups on a set of difficult-to-identify selection criteria. Project managers may agree to introduce more explicit participant selection criteria to avoid some biases, or administrative data collection may be strengthened during participant selection so that actual selection criteria will be better documented. Applying and documenting clear selection and rejection criteria strengthens the possible types of analysis.

- Developing a program theory (i.e., logic) model and articulating the *effects chain* through which impacts are expected to be achieved can help identify the critical assumptions and issues on which the scarce evaluation resources should be focused.

- Peer review, ideally a standard component of any evaluation, is very helpful when assessing the threats to validity of measures to address real-world constraints, by bringing another perspective to bear on the tradeoffs being made.

- When the budget does not permit the use of a household sample survey, Participatory Rapid Appraisals (PRA) and other qualitative techniques can often provide community level estimates of, for example, water consumption or use of new sanitary facilities. However, if these techniques are to be used for impact evaluation a sufficiently large sample of communities will be required to permit statistical analysis and it will be necessary to assess whether there are still cost savings.

- When operating with smaller than ideal sample sizes, it is sometimes possible to develop cost-effective mixed-method approaches that strengthen validity by providing two or more independent estimates of key output, outcome or impact indicators. For example, if key informant estimates of changes in quantity and reliability of water supply are consistent with household surveys, this can increase the credibility of the estimates. However, the use of mixed-methods increases data collection costs so it is important to determine whether this strategy does contribute to overall cost reduction. It is also important to remember that if the sample is too small, it will still not be possible to apply statistical tests for hypotheses testing — even if the mixed-methods increase the plausibility of a causal relationship.

b) Strengthening the sample design and statistical analysis

- Sample size issues: Reducing sample size is a tempting way to save money, but smaller samples increase the risk of false negatives (wrongly assuming the project did not have an impact). Statistical power analysis is a useful way to ensure the proposed sample will be large enough for the purposes of the analysis.

Box 11 (cont-d): Ways to address the effects of budget, time and data constraints on the validity of the evaluation conclusion

c) Strengthening the theoretical framework and the validity of the counterfactual

- Rapid assessment studies are a cost-effective way to develop the program theory models discussed earlier.
- When working with small samples, mixed method approaches can provide a cost-effective way to understand key concepts and to improve their measurement.
- The counterfactual can be strengthened by reconstructing baseline conditions (even when using propensity score matching designs that rely on post-test comparisons without baseline data) and where necessary strengthening comparison groups.
- Ensure time and resources are allocated to develop program theory. Simpler models can be developed in a relatively short period of time. Often the evaluator will find that although program staff all have their own ideas of the program objectives and how they will be achieved there is no official formulation of the program model and this has to be elicited during interviews, workshops and review of program documents. It is, however, possible to do this in a cost-effective and rapid manner.

d) Strengthening the generalizability of conclusions

- Use rapid assessment methods (key informants, focus groups, observation, etc) to identify the similarities and differences between the project and comparison groups and to understand how this might affect generalizability of findings. Also use mixed methods to interpret the results — good qualitative work can shed light on the process that got you the measured outcome, even if you did no process analysis initially.
- Use contextual analysis (qualitative or quantitative description and analysis of the local economic, political, institutional and socio-cultural context in each project location) to understand how local factors might affect outcomes and to assess how far the operation of these factors is unique to this particular context and how far they could be generalized.15
- Multivariate analysis should be used where possible to strengthen the match of project and comparison groups and hence strengthen the validity of projections of the conditions under which the project could be replicated.

e) Assessing when it is feasible to conduct an impact evaluation

Before implementing the evaluation, an *evaluability assessment* should be conducted to determine whether the requirements for conducting a quality impact evaluation (see Section 1) can be met with the available resources, timeline and data availability. This is conducted after all possible avenues have been explored for strengthening the evaluation design, using the techniques discussed earlier, and the best available design has been proposed. If an acceptable *impact* evaluation cannot be conducted within these constraints, the resources and time frame must be renegotiated, the scope and objectives of the evaluation revised, or the evaluation cancelled. Remember it is often possible to conduct an operationally useful assessment of potential project effects even when conditions do not permit a rigorous impact evaluation.

4.3 Enabling Ex-Post Impact Evaluations

4.3.1 Ex-ante Impact Assessments

There are two notions of ex-ante assessments used in policy and evaluation literature. In a more strict sense, it is the prediction of the impact of a policy/program, using econometric forecasting, model building, simulation techniques and alike: in this context ex-ante assessments are more suitable for important government policies, since carrying out such evaluations is an expensive exercise.

A more conventional notion of ex-ante impact assessment stands for a tool for effective management- a process that supports the preparation of proposals for new or renewed project level interventions. Its purpose is to gather information and carry out analyses that help to define objectives, to ensure that these objectives can be met, and that the instruments used are cost-effective and that reliable later evaluation will be possible[51]. For this notion, other terms, such as appraisal, policy analysis, impact assessment and feasibility study are also widely used.

The term as used in this document is a shorthand for the information needed to support good quality programme preparation and the analyses that produce this information.

How to actually carry out an ex- ante impact assessment should be decided pragmatically, taking into account the real information needs of each situation. The time and effort put into an ex- ante impact assessment should be proportional to the scale of the intervention that it supports. Existing information and evidence from earlier evaluations, studies and other sources should be fed into the ex-ante process whenever possible.

The content of ex-ante assessment would differ in each case, but the list below could serve as a guide. It is not the purpose of this paper to provide a guide to program design, therefore in what follows below the attention is focused only on those aspects of project design, which would facilitate impact evaluations later on:

- **Problem analysis and needs assessment-** assessment of future needs and future external environment;
  - Defining of the key aspects of the situation to be addressed by the programme;
  - Identification of the key factors that are likely to influence the key problem;
  - Identification of the main groups of actors that influence or that are being influenced by the situation;
  - Analysis of the cause-effect relations between the factors identified and the interests and motivations of the actors;
  - Constructing of a visual presentation of these relationships, for example in the form of a "problem tree"

- **Assessment of Alternative delivery mechanisms-** assessment of options for intervention (intervention strategies, instruments, channels and levels of

[51] 'Ex ante evaluation a practical guide for preparing proposals for expenditure programmes", EU Commission, 2001

intervention) including appropriateness and quality of delivery mechanisms, a Cost-Benefit Analysis; and Risk Assessment. The latter is important in order to identify suitable courses of action to prevent or mitigate their impact, and, at the impact evaluation stage, to be able to track down the assumptions made.

- **Developing logframes and results chains as part of the project designs.** This will enable managing the projects better, providing early indications of program effectiveness during project implementation. If there are breakdowns during implementation, it is possible to fix them along the way. And it will ease impact evaluation later on, since the assumptions, expectations, risks, and factors in the external environment will be described.

- **Setting of objectives and related indicators-** setting of general and specific objectives, associated results and measures of achievement; expected level of future results and associated costs in terms of operative credits and human resources. From impact evaluation perspective, setting concrete, measurable, or at least verifiable, objectives is fundamental to the success of the programme because it lays a basis for later evaluations of what has been achieved. If objectives are vague and too general, it is difficult to assess whether the intervention has been successful or not.

- **Targets.** In order to trace changes resulting from program activities through to poverty impact, programmes must track changes at the intermediate levels of the causal model. To track these changes, there must be at least one indicator associated with each change expected in the causal model. While using quasi-experimental studies for measuring change may reduce the need to measure all intermediate indicators, it is rare that a quasi-experimental study can cover more than one or two levels of a causal model. Therefore, both programs doing quasi-experimental studies and those using other strategies for measuring changes should track changes in intermediate indicators. To be useful, all indicators should be specific to the change expected, measurable, relevant to the programme and attainable as a result of changes in the previous level of the causal model.

- **Lessons from the past**. The following important knowledge can usually be gathered from evaluation reports and related studies:
  - Relevance of the existing strategy and possible needs for its amendment;
  - Effectiveness of existing delivery instruments;
  - Critical factors affecting the implementation of a programme;
  - Types of problems encountered while monitoring and evaluations were carried out.

- **Planning future M&E**. Establishing clear objectives and indicators is the first stage for putting in place a good quality monitoring system. The ex- ante process should address the following aspects of monitoring:
  - Planning of the necessary arrangements for collecting data on the foreseen indicators (both physical and financial) and other factors relevant for later analysis of market development, etc;.
  - Analysis of the soundness and reliability of the proposed methods and concrete instruments for collecting follow-up data, storing and processing this data and ensuring its validity;.

- Ensuring that the monitoring system is fully operational from the outset of the programme;

The ex ante assessment should then go on to analyse:

- what types (mid-term, final or ex post) of evaluations are needed and when
- what should be the main focus of these exercises
- who is responsible for carrying them out.

The ex- ante exercise should pay particular attention to the overall approach and timing of each evaluation. When planning the different evaluations, it is important to set up a clear link between the evaluation process, its results and decision-making.

4.3.2 Projections of Impact and Impact Monitoring

The impact of some programs may not be significant until several years after the program has ended. In addition, programmes that intervene more directly may achieve greater impact in the short term than those that aim to influence systems more indirectly, but those that intervene more indirectly, giving precedence to sustainability, may achieve significantly greater scale of impact over the long term as change permeates through a system or systems. In order for donors, programme managers and other stakeholders to effectively judge the success of a program, impact must be realistically projected out beyond the life of the programme.

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.

To facilitate comparisons among programmes, all programmes should make realistic projections for impact at least two years beyond the end of the programme, using transparent methods and calculations.

These projections should be made as targets at the start of the programme. An alternative is to identify the magnitude of potential impacts - according to some scale, for example in line with a methodology called Poverty Impact Assessment (PIA), described in OECD (2006)- whereby the potential impacts are assessed by the direction of their effect (+ or -) and strength (e.g. ++).[52]As the program matures, quantitative projections could be made and updated regularly, e.g. annually, based on evidence and trends in data gathered to date. This process is referred to as "impact monitoring", and is a key strategy employed by GTZ[53]. Projections should take into account the multipliers developed for the programme. These multipliers should be adjusted based on annual impact assessment results.

4.4 Conducting ex-post Impact Evaluations: Summary of Main

[52] 'Promoting Pro-Poor Growth: Harmonising ex ante poverty impact assessment", OECD 2006

[53] GTZ 2004: Result-based Monitoring: Guidelines for Technical Cooperation Projects and Programmes

messages

This section attempts to provide a summary of the main challenges that impact evaluations pose, points out those that are/will be more pronounced for UNIDO due to the nature of its projects, and some guidance on the ways to deal with these.

Attribution

The term stands for attributing the changes observed over time in the treatment group in a project vs. other factors, or in other words comparing the performance of the treatment group with a hypothetical situation of what would have happened without the project (the counterfactual). Thus, an impact evaluation is:

- explicitly about establishing causality;
- always challenged with an essential problem *of missing data,* because establishing the counterfactual means coming up with performance measures in a situation that never happened (since the project was carried out);
- explicitly concerned with assessing the final socio-economic results of the project. But since what happens during the project implementation affects these results, as the project progresses along the results' chain, keeping the 'process' in the perspective is vitally important so that not to allow misjudgments about the causes of certain observed effects.

Trade-offs

Conducting impact evaluation is always charged with tradeoffs:

- The main trade-off is one between the constraints (time, resources and data) and the reliability of results;
- The trade-offs necessitate accepting some compromises when deciding on the choice of methodology and the design features of conducting impact evaluations, and these compromises manifest themselves in affecting the reliability and robustness of findings. Having said that, it is also important to underline that commitment to *quality impact evaluation* is essential: threats to validity should be acceptable;
- The trade-offs could be significant enough to render conducting an impact evaluation unfeasible- or, in other words, threats to validity would be so significant that there is no point in contemplating to carry out a particular impact evaluation. Thus, before making commitments on conducting a particular mpact evaluation and/or thinking about the details of the design, it is necessary to assess the *impact evaluability* of a given project given the real-life constraints.

Timing

Articulating time horizons for planned impacts should span along the project life-cycle and beyond.

- *Design of Impact Evaluations* should start even before the project implementation starts- with:

    - developing the results chain, deciding on the boundaries of the project and identifying all the main nested effects chains;

    - thinking of what design will be used to evaluate its impact several years after its end, and thus deciding on the way the baseline data is collected (e.g. covering both treatment and control groups or not);

    - selecting indicators, based on their: Relevance, Validity, Reliability, and Practicality;

    a

    - conducting ex-ante impact assessments, i.e. laying out the potential vectors for the expected impacts and carrying out all the preparatory work to allow a quality impact evaluation later on; and, where justified, conducting an assessment of the magnitude of the expected impacts.

- Large and long in duration projects should ideally have a mid-term evaluation in their design: while conducting these, the initial assessments of the vectors and magnitudes of anticipated impacts could be re-assessed and re-adjusted;

- Better use of project completion evaluations is needed with the aim of capturing the early impacts of the projects (this will, indeed, depend on the nature of the project) and coming up with *projections* of expected impacts.

- An ideal timing for conducting impact evaluation *per se* should allow sufficient time for impacts to materialise *after* the project's completion, but not too late to render assessing attribution effects impossible.

Choice of the Methodology and Design of Impact Evaluation

The choice of the methodology and design of impact evaluation is affected by:

- Availability of time, data and financial resources
- Nature of the projects
- Selection method for participation in the program
- Other factors

Experimental and quasi-experimental designs are expensive and require many months and even years to complete. Exclusively qualitative studies are less expensive and shorter , but are at the same time less reliable in terms of robustness of their findings- especially those specifically aimed to complete in a few months and very limited resources.

Nature of the projects affects the selection of the methodology for impact evaluation. For instance, policy advice projects, targeting central governments do not most often allow for employing experimental and quasi- experimental methods.

The way the projects come up with the target group, e.g. enterprises, also affects the choice of method for impact evaluation. When the choice of enterprises is not random, but there are pre-announced or de-facto evident selection criteria, experimental designs

are not feasible and quasi-experimental designs, e.g. propensity score matching, are more suitable. .

It is important however, also to emphasise that it is rarely the case when only one method is possible. Also, a typical project would consist of several components, including, possibly, a policy advice component. Therefore, every project needs to be considered as a separate case, with the potential methods/their combinations identified, and assessed against the constraints.

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 about the methodology to be used 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?

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
- Scenario 2: Tracking emergent, complex interventions: Developmentally-driven impact evaluation
- Scenario 3: Ensuring accountability for results: Accountability-driven impact evaluation
- Scenario 4: Learning from comprehensive collaborative interventions: Contribution-driven impact evaluation

Rigorous impact evaluation

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. 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)

Data collection

Regardless of the approach, all impact assessment methods require accurate and reliable data. At a minimum, it is required to collect data at the outcome level for the entities that will be affected by the initiative. In all cases, baseline data is strongly preferred: while there are techniques allowing assessment of impact without having a baseline or allowing for the reconstruction of baseline, having actual baseline data allows for more rigorous analysis.

There are only two options to collect the required data:

- Drawing on existing data maintained by Governments and other organizations, or piggy-backing on the surveys carried out by them;

- Commissioning special surveys. The type of survey selected, the wording of the questions, sampling strategies, follow-up and data entry procedures are important for the accuracy and utility of the survey results. Probability sampling: stratification should be considered for greater efficiency and to ensure that the sample accurately represents the target population.

In addition, administrative data will always be required to identify and characterise service recipients. Moreover, certain techniques require additional information. For example to employ regression discontinuity, the program must maintain data on variables used to determine eligibility and/or qualification for participation. .

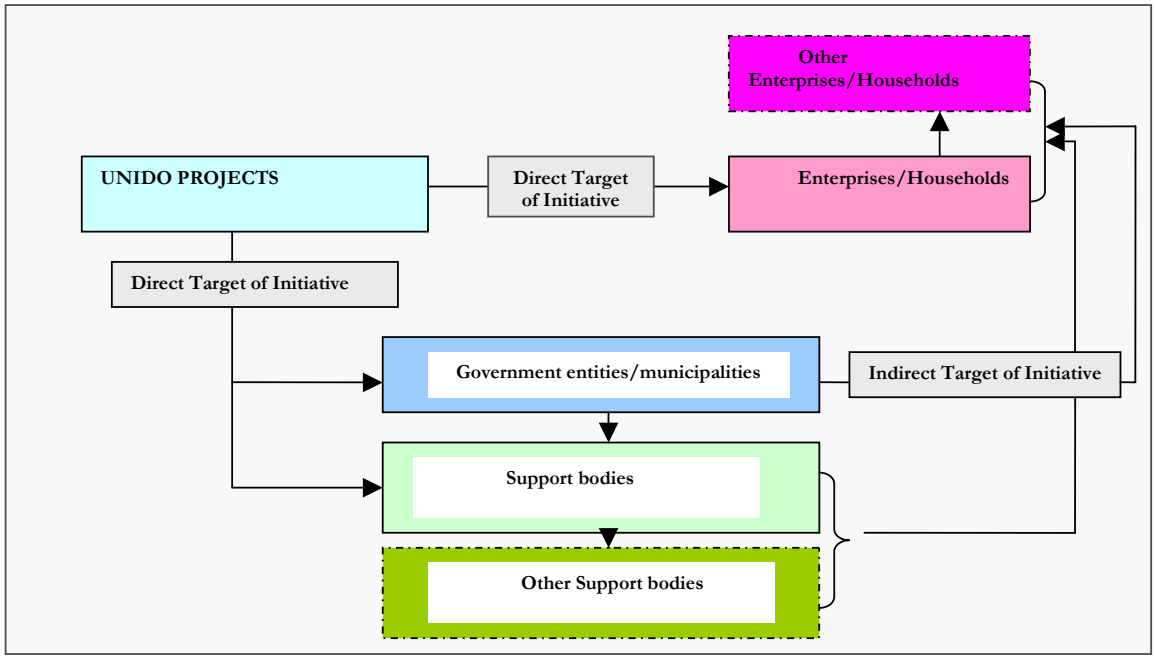
5. SUGGESTED APPROACH FOR IMPACT EVALUATION AT UNIDO

5.1 Main principles

5.1.1 UNIDO main intervention modes and instruments

As a starting point, a detailed review of the projects' portfolio of UNIDO is needed to identify the intervention typology. The initial (based on information from Programmes and Budget document) review of the portfolio of UNIDO in these Major Program Areas enables us to develop a *tentative* typology of projects. Based on the analysis and interviews at UNIDO it could be concluded that the interventions across all TC programs are conducted using three main intervention modes, each of which uses similar intervention instruments in broad terms. This provides a fertile ground for developing theories of change across portfolios and for using theory based impact evaluations approach. In Table 1 these modes and instruments are mapped across the Components of the main three TC Major Programme Areas - C, D and E. Figure 7 describes the main three nested results chains in line with the identified three modes.

Figure 7. The three nested results chains in typical UNIDO Interventions



For every major programme area/important thematic area the above should be adapted with specifics and potential evaluation designs laid out.

Table 1 Typology of UNIDO interventions

PROGRAM COMPENENTS	**INTERVETNTION MODES**											
	Interventions aiming at improving legal and policy frameworks				**Interventions aiming at strengthening support bodies and structures**				**Direct support to enterprises**			
	Intervention Instruments				**Intervention Instruments**				**Intervention Instruments**			
	TA to government agencies	Support for public – provide dialogue	Direct advocacy for specific polices	*etc*	Hardware And Software	Consulting and training	Development of new services	*etc*	Hardware and Software	Consulting and training	Brokerage	*etc*
Poverty reduction through productive activities												
Program components												
Trade Capacity Building												
Program components												
Energy and Environment												
Program components												

5.1.2 Theory Based Impact Evaluation as the core approach

Ideally TBE needs to be adopted for all the impact evaluations at UNIDO. This will well suit evaluating impacts of UNIDO interventions due to several reasons:

- UNIDO interventions target different aspects of a given problem, combining reforms in laws and regulations, supporting capacity building in infrastructure, and direct support to business and communities, oftentimes building up the projects in a phased way. These are usually not one-off and strictly targeted homogeneous projects, but rather, complex interventions, in which one can trace certain "belief' systems/stories, including for example, a strong emphasis on working with pilot enterprises with the belief that their positive performance will trigger others' buying in and multiplier effects; capacity building for support institutions as agents of change, and so on.

- TBE approach provides a good platform for employing qualitative methods of evaluations, which might be the only feasible methods given the shortage of data for most of the impact evaluations to be carried out in a short-time horizon.

Besides, UNIDO evaluation policy states that "… evaluations focus on the analysis of expected and achieved accomplishments, examining the results chain, processes, contextual factors and causality, in order to ascertain the degree of achievement or the lack thereof."- very much in line with TBE approach.

Moreover, it needs to be adopted at the design stage. By mapping out the determining or causal factors judged important for success, and how they might interact, it can then be decided which steps should be monitored as the program develops, to see how well they are in fact borne out. Where the data show these factors have not been achieved, a reasonable conclusion is that the program is less likely to be successful in achieving its objectives. Needless to say, that a necessary starting point then is that all the projects should have logframes and results chains developed at the deign stage.
An approach then needs to be developed on how generalize (or build theories of change[54]) from project level intervention to impact theories at the level of thematic areas of intervention, see section 5.2.2.

5.1.3 Methodologies to employ- generally applicable UNIDO-wide remarks

a) Main points

- Due to the fact that that UNIDO projects are rather varied in nature and size, there is no strong rationale supporting the use of only one specific methodology to use for impact evaluation. For large, multi-million Euro projects, especially for those which are being implemented/replicated in many countries, it might be justified in financial terms to employ experimental and quasi-experimental designs. The problem of the current lack of baseline and monitoring data (for current and completed projects) is, hopefully, one that would be resolved in a long-term, and, for the short-term,

[54] http://www.evaluationtools.org/plan_theory.asp

could, oftentimes, be resolved by different techniques for reconstructing baseline data (see below, in (b)). What should be taken into account is, whether using experimental and quasi-experimental designs are at all possible, due to political, ethical or contamination issues or deriving from the nature of the projects (e.g. when technical assistance is provided to a single government agency for policy formulation).

- For several program components, provided that the resources are available, the experimental and quasi experimental designs will be very suited indeed, e.g. in the case of Agro-processing projects, some of TCB projects, due to the fact that often the interventions here are rather straightforward, oftentimes involving well defined assistance to enterprises, either directly and/or through support institutions. Quasi-experimental designs with matched comparisons, pipeline approach and reflexive comparisons, seem to be suitable tools to employ. .

- These techniques (above) are not suitable for a wide range of activities that UNIDO, like many other development agencies supports, including policy advice, economic and sector work, and institutional development activities more generally, where these are carried out either at the central level or in a very small number of districts. For such cases, employing qualitative methods will be needed.

- For small to medium size projects, where the funding is a major concern, less expensive methodologies (than experimental and quasi-experimental designs) seem to be more feasible. And again, the fact that for the completed and ongoing projects there is no baseline and monitoring data, is only a secondary and hopefully, a temporary problem. Here PRAs (combined with quantitative analysis wherever possible) could be utilised. This could be suitable for environmental projects, projects focusing on policy advice or network development projects. Cluster/thematic evaluations should be used more, perhaps joining the financial resources with other concerned agencies.

b) Addressing current data limitations

All the projects need to collect baseline and monitoring data: this is more a management decision and we should not elaborate on it at length in this paper. In the short term however, there are some partial remedies.

IEG/WB (2006)[55] suggests a list of measures that could be contemplated to reduce the costs of data collection:

- Reducing the length or complexity of the survey instrument
- Reducing the number of interviews conducted, through
- Replacing individual interviews with community-level data collection.
- Reducing interview costs
- Electronic data collection

[55] 'Conducting quality impact evaluations under budget, time and data constraints', IEG/WB, 2006

Impact evaluators usually have to define the post-implementation comparison group as the *counterfactual*, and these designs cannot control for *unobserved* differences between the two groups at the time of project launch. This is particularly problematic when participants are *self-selected*, as enterprises that elect to participate in a project are often those most likely to succeed. From the interviews conducted at UNIDO HQ, it became apparent that this might be particularly true for UNIDO projects. The lack of baseline data makes it then particularly difficult to separate project effects from pre-existing differences. Strategies for reconstructing baseline data include the use of:

- Secondary data from national surveys, censuses and similar studies, conducted around the time of project launch may identify the project population and include relevant data on the project population and/or potential comparison groups;
- Administrative records collected by the project;
- Records from partner agencies;
- Recall

When no secondary data are readily available, qualitative techniques such as *Participatory Rapid Assessment, Key informants Interviews* might be used.

c) Instituting a clear and transparent system for documenting the impacts of the projects

It is important that UNIDO institutes a clear and transparent system for documenting impact evaluation, if it has to be carried out, including.

- Indicator measurement should be accompanied by description on how attribution/contribution is measured/estimated/assessed;
- The most rigorous method that is appropriate and practical should be used. Less rigorous methods should be justified;
- If attribution for changes at the enterprise level is established during a pilot phase and then used to extrapolate changes later, this should be clearly described. Where attributable change is extrapolated, the system should include a method for regularly verifying that the attributable changes established are still reasonably accurate;
- All assumptions are clearly outlined:
- All estimates are clearly described and justified with reasonable evidence;
- All public contributors to change are acknowledged. All public programmes (donor and government) which have contributed to the attributable changes claimed are clearly identified with the financial value of their contribution, as accurately as possible;;
- Attributable change claimed by the program is adjusted to take into account other public contributors to that change. Change attributable to the programme is adjusted based on the relative financial value of the contributions to that change from all involved public programmes.

5.1.4 Multiplier effects

Often the impacts of programs will reach to enterprises beyond those directly covered by the program. This effect of programmes is sometimes called "crowding in," "copy cats," or "spontaneous replication". It is important to include this effect in impact assessment, because it provides the incentive to programme designers and managers to design programmes which emphasise scale and sustainability. Including this effect also provides

a more accurate picture of the actual impact of a programme.

This effect must be taken into account in impact assessment by investigating crowding in at the enterprise level of a causal model as well as at any other appropriate levels. Research must include strategies to measure the extent of "crowding in." The research should be appropriate to the size of the programme. For example, a very large programme should provide evidence from enterprise surveys with statistically significant samples.

A small programme may provide evidence from a handful of focus group discussions, a series of in-depth interviews or a reasonable number of case studies. Evidence must include not only evidence of a change in enterprises but also evidence of attribution to programme activities. As above, the most rigorous method to establish attribution, given the programme size and circumstances, should be chosen. However, it is recognized that establishing attribution for "crowding in" can be challenging, so participant opinion, expert opinion or other qualitative methods are acceptable as long as they are well thought out and executed.

Changes at one point in a market or other system are very likely to produce changes at other points, for example in forward and backward linkages, related value chains or related regulatory systems. Programme managers may include impact produced by changes at other points in systems if reasonable evidence is provided. As above, reasonable evidence depends on the size of the programme. Evidence must include evidence of attribution to programme activities.

Paying specific attention to multiplier impacts in the case of UNIDO projects seems particularly important as it has developed a large portfolio and institutional history of specifically addressing pilot enterprises with the assumption that they will serve as successful example for others to copy them. Results have been mixed, however, as became evident during the interviews conducted at UNIDO HQ and as is articulated in UNIDO (2007)[56]: "…Reports identified a number of success stories in pilot approaches. However, UNIDO did not have a strategy to promote this approach and, so far, has not developed a common understanding of the pilot approach concept. There is an urgent need to address this weakness in order to capitalise on successes and experiences and to meet a growing demand for advice and sharing of information within a South-South context". The reasons for this mixed outcome need to be investigated, and assumptions during the project design stages laid out: this will help to understand where are the broken links. For example, the same UNIDO report [57] suggests that the drawback is perhaps in the fact that the pilot activities do not often include " …proper demonstration and dissemination activities', which is a very likely cause, but there might be other relevant factors, like regulatory barriers, industrial structure in particular industries and localities, and so on. This might as well be other potential reasons. For example the latest WB Global Economic Prospects report suggests that in developing countries the spreading and take up of advanced technologies is hindered by such factors as infrastructure, and low levels of skilled labor.

[56] "Comparative review of lessons learned from 20 UNIDO Integrated Programmes", UNIDO, Evaluation Group, Vienna, 2007

[57] ibid

The report rightly stresses the importance of ensuring that "…the responsibility and modality for scaling up pilot activities and developing policy requirements at the country level should be addressed upfront at the project design stage".

Indeed, projects working directly with poor communities and pilot companies have their rightful place within UNIDO's poverty reduction strategy. However, the effectiveness of using these pilot interventions' as models for replication and upscaling is justified only when this assumption is supported both by program design and country characteristics.

In cases when pilot approach is justified, and well designed, this could be very useful for the impact evaluation perspective: impact evaluations of pilots could be much less expensive than full-scale interventions. Using control groups to establish attributable change at the enterprise and/or poverty levels during piloting is often more manageable than later in the program because it is usually clearer at this stage who the programme is and is not affecting. Attributable change established during pilot implementation can be used to extrapolate attributable change as the programme scale and impact expands provided that it is periodically verified through other means (such as small surveys or case studies).

And replication strategies then could be designed and implemented after conducting such impact evaluations.

5.1.5 Unintended impacts

It is impossible for programme designers and managers to anticipate all the consequences of programme activities. Unexpected changes, either positive or negative, often occur outside of the causal model. It is important that significant unintended changes at any level of the causal model, but particularly at the impact level, are regularly explored. Finding these changes may lead to revisions in the causal model, the programme strategies or both. Even if no changes are made to the programme, it is important to explicitly explore and document unintended results of programme activities. Because good practice in this area for BDS development thematic area has not yet been firmly established, programmes may choose any reasonable methodology for identifying and exploring unintended results. Qualitative methods, such as focus group discussions or in depth interviews, are probably most practical for many programs.

5.1.6 Sustainability

OECD Development Assistance Committee glossary states that 'Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn. Projects need to be environmentally as well as financially sustainable.'[58] Aid agencies have each come up with their definitions of sustainability which look similar, but beyond definitions there is little agreement. White (2005) [59] proposes that analysis of sustainability would be best served by more serious attention to risks and their likelihood, and to do this suggests to use theory based approaches to be adapted to the analysis of sustainability. The theory lays out a set of

[58] Development Assistance Committee Working Party on Aid Evaluation (2002).

[59] H. White, "Challenges in evaluating development effectiveness', IDS Working Paper 242, Institute of Development Studies, Brighton, UK, March 2005

conditions under which X will lead to Y. If Y cannot be observed – as it cannot in the case of future benefits – then the evaluation can test whether or not the conditions for X to lead to Y are in place. Whereas TBE per se seeks to test these conditions, the analysis of sustainability uses reference to other literature to validate the theory linking X to Y.

This approach was adopted in the WB OED's analysis of the sustainability of sub-projects supported by social funds. Based on a review of the literature on project sustainability, conditions were derived that should be met if sub-projects are to be sustainable, and the importance of these conditions was validated through the four country case studies. The portfolio review recorded which projects had the required design features for sustainability and which did not. The study found that social funds were learning the lessons of experience and beginning to adopt the required design features for sub-project sustainability more widely.

TBE enables identification of factors critical for sustainability, but it does not tell anything about the probability of different events. It would seem a minimum requirement that identified risks should be labeled as "very likely" to "not at all likely", with some common understanding of the range of probabilities referred to by each label. Where the risks are of a vague sort ("lack of government commitment") the evaluator should be required to spell out in more detail the implications for the project. Figure 8 provides an example of some reasonably generic factors that might be considered in a sustainability analysis

Figure 8 Possible factors in sustainability analysis

Aspect	Factor	Data
Financial		
Government	Strength of tax base	Historical analysis of revenue variation and factors behind it (e.g. commodity prices).
	Ability to finance	Shortfalls between actual and budgeted expenditure.
	Willingness to finance	Likely government priorities. If government may change, what are priorities of opposition parties?
Communities	Ability to pay	Forecasts of charge as a percent of average income (requires analysis of livelihood of community).
	Willingness to pay	Possible changes in pattern of demand, perhaps because of competing suppliers.
		Will the project continue to produce benefits of sufficient value? (e.g. prices fall so output not worth as much, so producers won't pay for technical services).
Donors	Continued donor finance	Is it planned?
Institutional		
Government	Capacity	Are the skills required for project implementation or supervision present and will they stay there?
	Existence	If the project is under a special unit, will it continue to exist or are arrangements in place for the takeover of functions by exiting government units?
NGO	Organisational viability	Will the organisation continue to exist without external project finance?
	Capacity	Are the skills required for project implementation present and will they stay there?
Community	Organisation	Are there command-based organisations to be responsible for community level responsibilities? Can these responsibilities be enforced?
	Capacity	Do community members have the required skills, or are mechanisms in place so that they acquire them?
Technical		
Design	Soundness	Are physical structures sound or do they have structural problems?
Operations and maintenance	Ability	Do those responsible have the required skill for operation and maintenance?
Environmental		Is project depleting stocks of a non-renewable source? Are there adverse environmental consequences of project technology that will undermine its effects?

Source: H. White, "Challenges in evaluating development effectiveness', IDS Working Paper 242, Institute of Development Studies, Brighton, UK, March 2005

Including sustainability analysis in impact evaluations at UNIDO seems to be well justified, since UNIDO has a major focus< as one of its intervention modes, capacity

building of support institutions. These are seen as dynamic agents of change, and it is believed that it is through strengthening of these meso-level institutions and building up a strong private sector response capacity that UNIDO can help to put economies on a robust and pro-poor growth path.

These support institutions are both in the public and private sectors and could also be non-profit organisations with proven growth potential (the "missing middle").

It is pointed out in the UNIDO Poverty Reduction Strategy Forum[60], that : "… we have ample experience showing that conducive policies (while necessary!) cannot guarantee a sufficient response capacity on the part of enterprises. While the "costs of doing business" may be low, it takes management experience, resources, technology, skill upgrading and more to become competitive". This prompts to the need of more attention to be paid to needs assessment at the project design stage and sustainability analysis both at the design stage and during impact evaluations.

According to UNIDO (2007)[61], "…Good practices in achieving sustainability were found in programmes with a strong national leadership in design and implementation, involving private sector institutions, and those that succeeded in applying a comprehensive multidisciplinary approach and programmes in areas where UNIDO expertise was recognized and had a proven track record'. Main factors that affected sustainability negatively were: supply orientation, poor design, weak monitoring and lack of funding.

Sustainability is closely interrelated with counterpart and beneficiary ownership. It needs to be built up during the IP formulation and implementation, and not merely at the end of the implementation phase. The absorptive capacity of the counterpart and the availability of adequate human, financial and physical resources are decisive and call for realistic assessment. UNIDO and the counterparts must agree on a post-assistance scenario before entering the implementation phase.

5.1.7 Defining indicators

As mentioned in Chapter 3, impact indicators are not spelled out in UNIDO strategic documents[62]. Hopefully this drawback will be resolved as a result of introducing RBM and RRS. Currently however, there is no clarity about anticipated impacts or impact indictors across the Major Programmes. For example, while, most likely poverty reduction is an expected impact for Major Programme C, is it also for Major Programmes D and E? What are other impacts expected across all three Major Programmes? It seems that at the moment there is no consensus about this within UNIDO across departments[63]. Introducing clarity in identifying expected impacts and impact indicators for the major programme areas is a matter of urgency from the point of

[60] UNIDO intranet: Poverty Reduction Strategy Forum

[61] "Comparative review of lessons learned from 20 UNIDO Integrated Programmes", UNIDO, Evaluation Group, Vienna, 2007

[62] Interviews at UNDO HQ also demonstrated that the staff feels that there is no clarity about UNIDO's fundamental objectives and the way forward towards achieving them. The new UNIDO corporate strategy – upcoming- among other issues, will also clarify this.

[63] Ibid

view of moving ahead with Impact Evaluations.

Many at UNIDO are concerned that the impact of UNIDO's interventions on poverty has low visibility and are keen to see changes with this regards. While one of the reason behind the low visibility might be insufficient external communication, the core of the problem is in vagueness in articulating expected impacts, lack of agreed upon set of impact indictors and measuring the results (for impacts) - both at project and aggregate portfolio levels.

It could be argued that most of UNIDO's TC programmes and projects are directly or indirectly poverty-reducing. For some types of interventions, the causal chains are likely to be complex (e.g. linking technology foresight exercises or strengthened metrology laboratories to the poverty goal) or importantly, may require the consideration of an extended time line (e.g. when looking at the impact of industrial policy advice). In other cases, the impact is more obvious (e.g. in projects supporting SME clusters or leather product development). And finally, certain types of projects are demonstrably and directly poverty-reducing (e.g. training of women entrepreneurs or support to reduce post-harvest losses).

A position needs to be taken also towards setting impact indicators in relation to Millennium Development Groups. Again if we take *Poverty reduction* as one of the expected impacts, then the next task is to choose related impact indicators. This is an area of intense debates among the development community.

- Many argue that indicators such as changes in output and productivity, employment levels and owner income of participating companies are proxies good enough to stand as measures of poverty reduction. They argue that for small-to-medium size projects it is next to impossible to estimate the effects of a program on poverty, measured in changes of poverty indicators.

- A more strict line of thought argues that these would be of questionable validity as proxies for poverty impacts, and evaluations need to be based on valid indicators of changes in poverty indicators.

It is important to keep the flexibility here. Ideally one should try to trace up the impacts of projects on poverty indicators. This might however not be feasible for most of UNIDO projects due to small size or nature. Therefore, the recommendation is to agree on broader set of indicators, which would, in addition to poverty indicators per se include also indicators like Net number of jobs created, changes in owner income of participating enterprises, and so on (see Section 6.2). But the list has to be defined for each Major Program Area and Program Component. In terms of poverty indicators per se, it is also important to have a broad vision of poverty concepts and, hence, indicators, capturing impact on income and access aspects of poverty.

- Below, we first examine the income aspect of poverty.

    - It would be necessary to establish a threshold level of income or consumption below which people are considered poor if poverty is to be measured. Oldsman (2002)[64] suggests to take the income level as the most common definition for poverty: "In the absence of a good alternative,

[64] ibid

donor agencies may want to use $1 per day (PPP) as the relevant poverty line …since this would be consistent with the Millennium Development Goals". However, the choice of the poverty line(s) to use in an evaluation of a particular program should be based on the objectives of the program as well as social norms as to what constitutes minimum living conditions.

- In general, there are four different econometric poverty indicators, which could be used:
  - *"poverty rate (headcount)*: the proportion of the population living in households with income per person below a specified poverty line;
  - *poverty gap index:* a measure of the depth of poverty within a population in terms of the gap between income or consumption levels and the specified poverty line;
  - *squared poverty gap index:* a measure takes the severity of poverty into account by placing a higher weight on individuals living in households that are furthest from the poverty line;
  - *Sen-Shorrocks-Thon Index: a* measure that attempts to combine the incidence, depth and severity of poverty into one indicator[65].

Developing results chain is extremely important in analyzing the causal linkages that lead to the expected results of poverty reduction. In ILO (2005)[66]it is, for example, demonstrated that technological developments will not necessarily result in poverty reduction at least in the short-to medium term, as one of the potential results of developments in technology is job reduction, especially in less developed countries (for the continuation of the discussion see Section 6.1)

The next stage is selecting indicators for Outputs, Outcomes and Impacts. RRS exercise carried out at UNIDO currently is addressing the issue of improving the sets of indicators, and we shall not pose here for long. Also, it is not the task of this assignment.

The necessity to reduce the number of outcome indictors and introducing standardisation in those has already been mentioned. What is of direct interest for the purposes of this report are impact indicators.

The Table in Annex 3 is an attempt to map the Outcomes identified in the Programme and Budget document for Program Components according to the three intervention modes identified earlier in the paper. This was developed by the EVA and it is our recommendation that improvements in the list of indictors for UNIDO within the context of RRS/RBM takes into account the perspective of addressing all three identified intervention modes.

5.1.8 Better use of existing UNIDO evaluations for the purpose of evaluating early impacts of UNIDO projects

Some elements of impact evaluations could be incorporated in planned UNIDO project

[65] Oldsman 2003, p. 9, cf. also Appendix A

[66] "World employment report 2004-05: Employment, productivity and poverty reduction", International Labour Organization, Geneva 2001, First published 2005

completion evaluations. Indeed each case will be different, but it is possible that some of the impacts could be identified already soon after the projects are over.

Project evaluations, conducted by EVA at their completion already attempt at assessing project impacts. However, these attempts contain inherent problems, since:[67]

- Figures collected at the conclusion stage are often based on self-declarations of the partners and hence represent mere indications of intentions;
- Unaccounted information could pose a significant problem, e.g. companies choosing not to disclose the actual figures
- Attribution issues are rarely addressed.

To be able to improve the incorporation of some elements of impact evaluation at the project completion stage, these limitations must be addressed in line with the recommendations above as well as those in Section 4.3.2.- incorporating and applying the concepts of impact monitoring and impact projections.

IP evaluations, which are in essence country level evaluations currently have small sections on assessing impacts of the projects, but these also have problems. An objective one is the fact that they cover projects at different stages of implementation and are called to provide a comparable and summative evaluation of all the projects in a country/region. In addition, IP self-evaluations do not seem to be in a state to facilitate using IP evaluations for more rigorous analysis: UNIDO (2007)[68] notes that "… 30% of the sample criticise UNIDO's IP self-evaluation mechanisms as inadequate'.

On the positive side however, it is a mechanism to look at the impacts of completed (with a few years time leg) in a given country, and provided that necessary resources were available, impact evaluations of those could have been carried out. The same is true for thematic evaluations carried out by EVA. In both cases, TBE approach should be best used to arrive to aggregation of the identified early impacts.

5.1.9 Relating Impacts to Programme Costs

In order to judge and improve the efficiency of programmes, impacts must be related to the costs of achieving them. Therefore, programmes must also keep track of the costs for inputs required to achieve the impacts stated

Programmes are encouraged to also separate costs by the three major intervention modes identified in order to provide useful management information.

Programmes are encouraged to represent impacts related to programme costs using ratios appropriate to the programme. These might include, for example:

- Cost per enterprise benefited;
- Additional income generated divided by programme costs;
- Cost per full time equivalent job created;
- Cost per person out of poverty;

[67] Independent Evaluation ITALY UNIDO Investment and Technology Promotions Office in Italy

[68] "Comparative review of lessons learned from 20 UNIDO Integrated Programmes", UNIDO, Evaluation Group, Vienna, 2007

Ideally Impact evaluations should include also Cost-Benefit Analysis to assess the costs at which the claimed impacts were achieved with – to enable to judge about the efficiency of the program.

5.2 From projects to programs: Enabling aggregation at portfolio level

Vaessen and D. Todd (2007)[69] focus on the connection between the thematic area of intervention and policy instruments. They argue that this link is crucial as theories of the effectiveness of (combinations of) policy instruments constitute essential building blocks of intervention theories at thematic area level. Some work on establishing this type of connection has already been done.

- Intervention theories can be developed in relation to different levels of impact. Depending on the type of intervention, and state of knowledge and available data, impacts are captured at four levels: catalytic effects[70], institutional changes, behavioral changes, and changes in socio-economic performance of the target population. The evaluators concentrate on developing causal theories that link particular thematic areas of intervention to the levels of impact which (in first instance) are deemed most relevant for these areas. Subsequently, the causal analysis is extended to end-impacts.

- The above-mentioned evaluation strategy of linking thematic areas to levels of impact should be triangulated through case study analyses (and other sources) in order to analyse the interplay between different intervention mechanisms (policy instruments, institutional structures), specific contextual variables and effect patterns at all levels of change.

- In projects, in which data are not available, but where it is relatively easy to identify the nature of the causal linkages (i.e. the intervention theory) between intervention outputs and processes of change, the PSR can constitute the basis for defining questionnaires and indicators in order to collect ordinal data concerning intervention effects on impact variables. This type of data can also be relatively easily aggregated across interventions and compared to national or international trends.

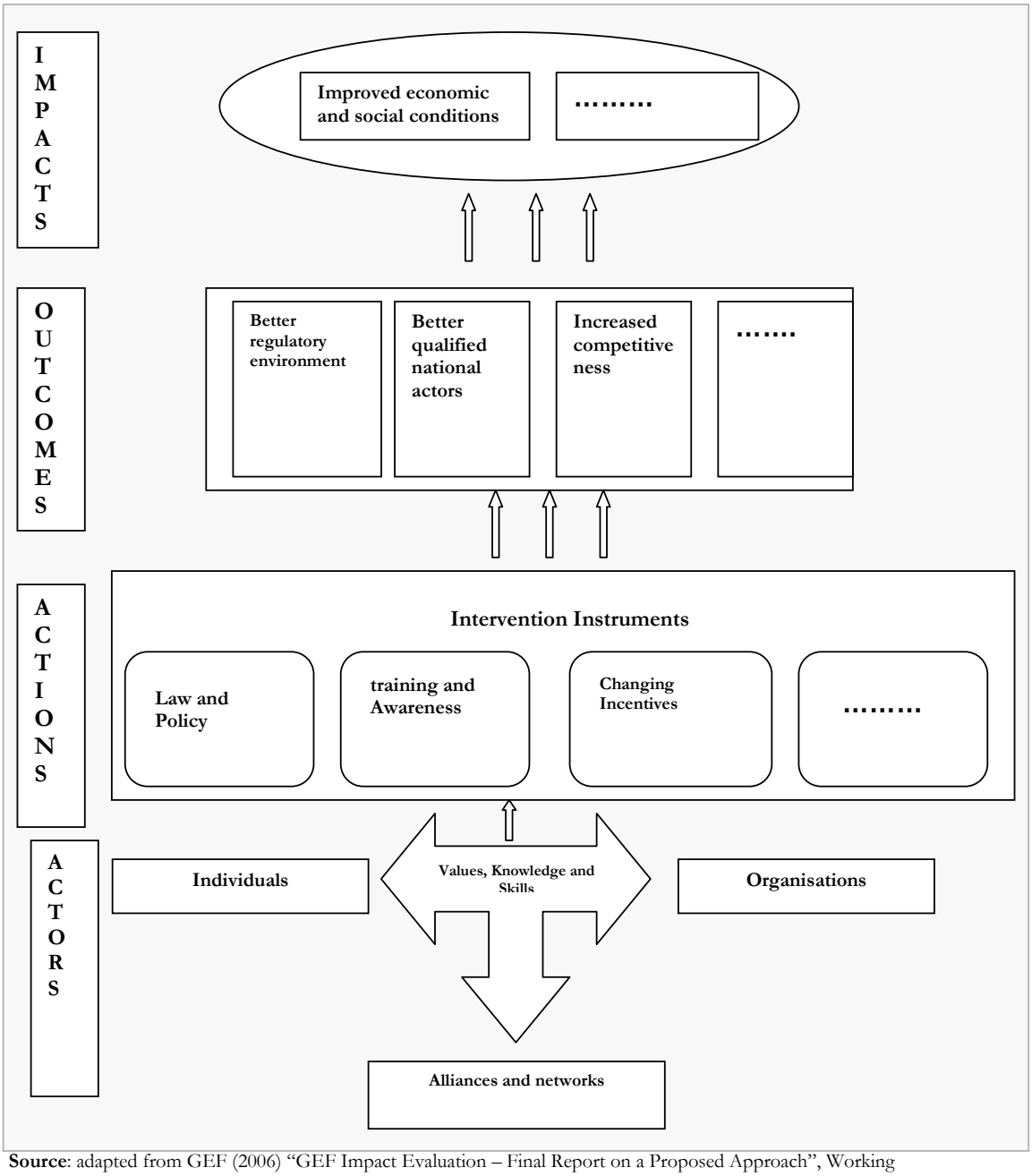
- Finally, in projects that comprise intervention activities about which little is known regarding possible causal patterns towards impacts, more attention should be paid to intervention theory reconstruction (preferably complemented by field assessments, expert interviews, and consultation of academic literature). This will provide the basis for collecting data on relevant proxy indicators.

As a next step for major Intervention theories could be developed along the lines described in Figure 8.

[69] J. Vaessen and D. Todd, Methodological Challenges in Impact Evaluation: The Case of the Global Environment Facility (GEF), University of Untwerp, Institute of Development Policy and Management, 2007

[70] A separate line of research will need to be conducted to develop an approach to identify the catalytic role of UNIDO activities. Recently GEF has commissioned such a research for its programs. There is limited systematic reporting at UNIDO on catalytic effects.

Figure 9 Basic intervention theory mapping



Source: adapted from GEF (2006) "GEF Impact Evaluation – Final Report on a Proposed Approach", Working Document Prepared for the GEF Evaluation Office, Foundations of Success, Washington D.C.

These main Intervention theories might not coincide with Major Program Area Components: for convenience we shall call them in this document "portfolios".

The increased emphasis on results in the context of the Millennium Development Goals has increased the importance of reports that aggregate impacts across evaluations. Development of such typical intervention theories along project portfolios will allow aggregation of results (at all levels, including impact).

A distinction could be made between the aggregation of quantitative data from project level to higher levels of intervention (thematic area/portfolio, global trends) and the process of generalisation or theory building from project level intervention to impact theories at the level of thematic areas of intervention.

- To enable aggregation of quantitative data results indicators must be, at least to some extent standardizes. We understand that under the current exercise of piloting RBM with the introduction of the RRS, this is already under discussion at UNIDO.

- There are techniques that allow aggregation of qualitative data - in a way that reveals, rather than distorts, the patterns in the data. The failure to use a systematic approach may lead to "cherry picking". Well-established methods are available for the systematic analysis of qualitative data; they include content analysis and computer programmes (e.g. QSRNudist) for conducting such analysis. For example White (2005)[71] believes that a tabular summary can suffice.

Table 2: Schema for summarizing qualitative data[72]

	Country a	**Country b**	**Country c**	**Country d**
Question 1				
Question 2				
Question 3				

6. A TOOLKIT FOR IMPACT EVALUATION OF PROJECTS IN THE THEMATIC AREA OF *BUSINESS DEVELOPMENT SERVICES*

In this Chapter a detailed approach is developed on how a TBE approach could be applied to the projects targeting SME/BDS development (which could be modified for others), and how the impacts identified for projects could be aggregated at Component or Programme level and theories of change developed for them. The same exercise could be replicated for other thematic areas.

The Toolkit could be applied for all the components of

- Major Programme C except perhaps, for Programme Component on Sustainable Production in Poor Communities, and Rural Energy for Productive Use
- Major Programme D, except perhaps, for Component D6, D7, .Corporate Social Responsibility for Market Integration.

For the Program Components related Environment, under Major Programme E, Energy and Environment, methodology developed by GEF could be used as a starting point.

Similar analysis, as in this Chapter could be conducted for the Energy portion of the Major Programme E.

[71] H. White, "Challenges in evaluating development effectiveness', IDS Working Paper 242, Institute of Development Studies, Brighton, UK, March 2005

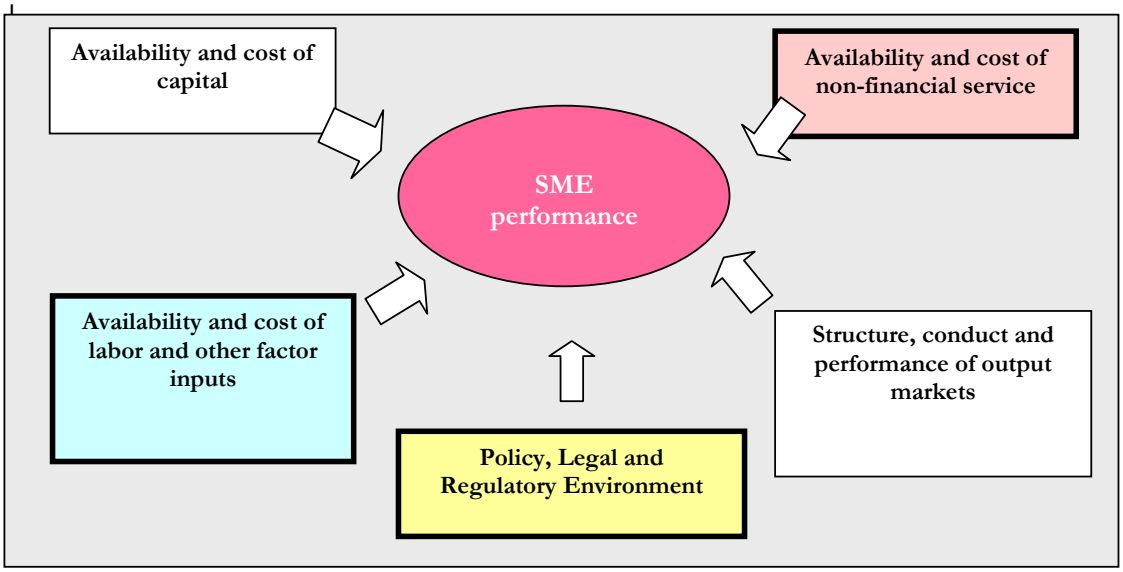
[72] Ibid

6.1 Developing Logic Models and Results Chains

We shall start by understanding what are the factors affecting SME performance. From the five factors identified in Figure 10, UNIDO addresses three:

- Policy, Legal and Regulatory Environment
- Availability and cost of non-financial service
- Availability and cost of labor and other factor inputs

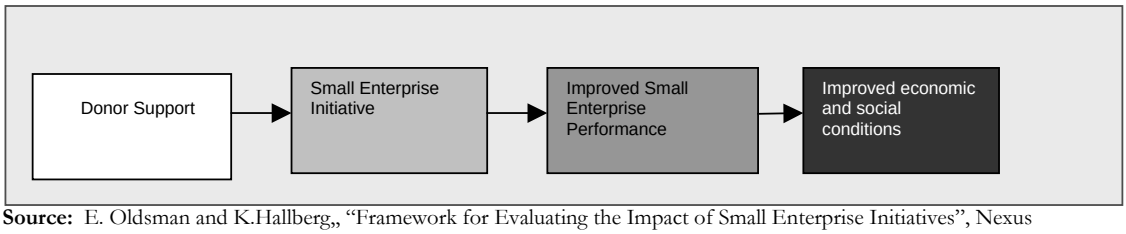
Figure 10 Factors affecting SME performance



Source: Adapted from E. Oldsman and K.Hallberg,, "Framework for Evaluating the Impact of Small Enterprise Initiatives", Nexus Associates, Inc., 2002, Washington DC

Regardless of particular focus and instruments used, all initiatives aim at boosting the performance of SMEs and in turn, improving economic and social conditions as shown in Figure 11.

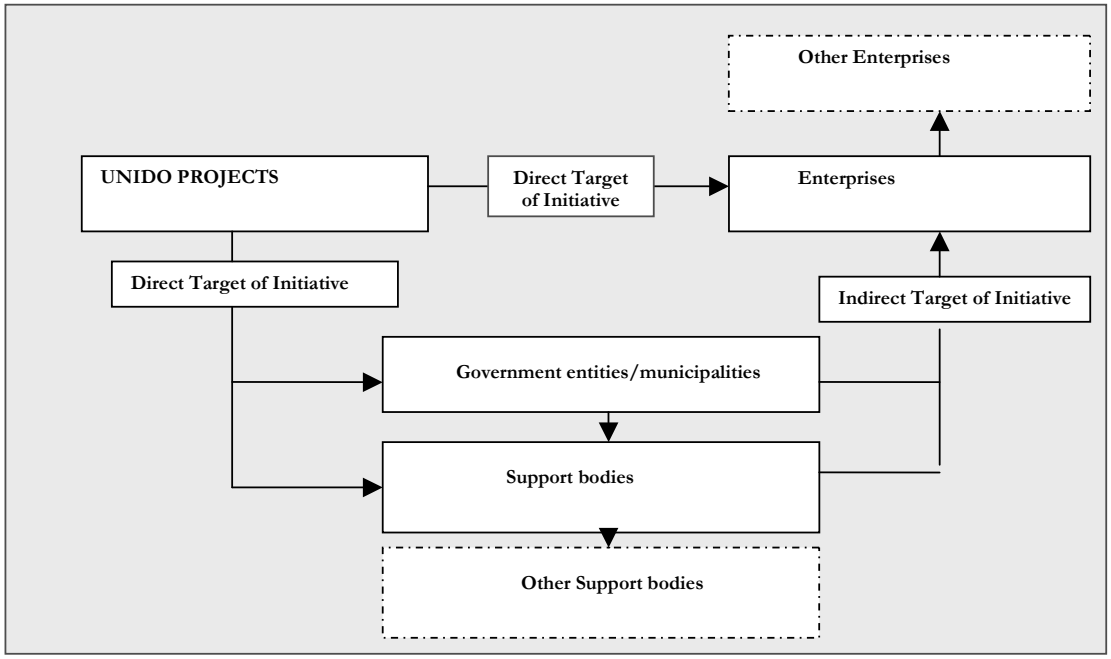
Figure 11 Basic Logic Model



Source: E. Oldsman and K.Hallberg,, "Framework for Evaluating the Impact of Small Enterprise Initiatives", Nexus Associates, Inc., 2002, Washington DC

Often the initiatives target SMEs of specific size and/or those located in a specific region and/or belonging to specific industry. However, SMEs are either direct or indirect targets of initiatives, as in Figure 12.

Figure 12 Direct and Indirect Targets of Small Enterprise Initiatives



Regardless of the specific nature of the project the intervention modes used could be grouped into three broad categories:

- Interventions aiming at improving policy frameworks
- Interventions aiming at strengthening support bodies and structures – in this case BDS providers, BMOs
- Direct support to pilot enterprises

We shall demonstrate it by mapping typical activities along the Programme Components /Intervention Modes matrix, as developed in Table 1. Ideally this should be done through detailed analysis of the entire projects' portfolio For this exercise, given the time constraints the information utilised was taken from Programme and Budgets 2008-2009 and UNIDO intranet.

These three intervention modes are operationalised through instruments/blocks of activities, which could also be categorized, to include for example:

- **Interventions aiming at improving policy frameworks**
  - TA and capacity building to government agencies and self-government bodies;
  - Support for public –private dialogue;
  - Direct advocacy for specific polices; etc
- **Interventions aiming at strengthening support bodies and structures**
  - Hardware/Software support of BDS providers and BMOs;
  - Training and consulting for BDS providers and BMOs;
  - Development of Service packages for BDS Providers and BMOs; etc
- **Direct support to pilot enterprises**
  - Hardware/software support to companies;
  - Training and consulting for companies;
  - Brokerage/value chain development; etc

Table 3: Typical activities of BDS Projects under the Major Programme C

Note: the lists of activities are taken from the Programmes and Budgets Document 08-09

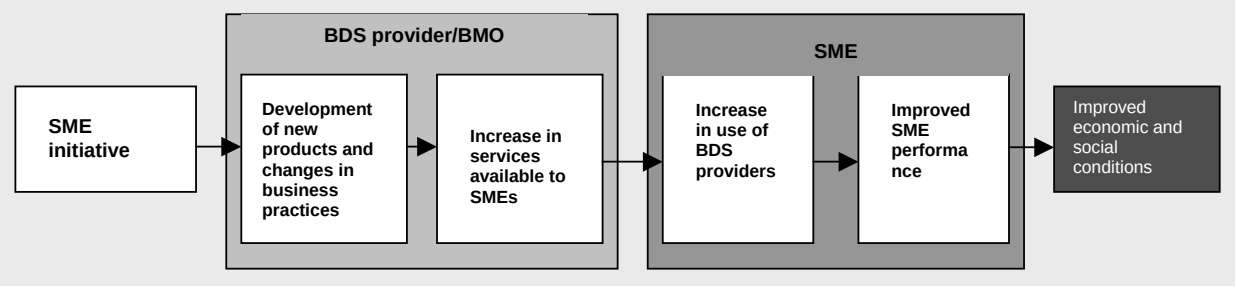
BDS PROJECTS UNDER THE MAJOR PROGRAMME C: POVERTY REDUCTION THROUGH PRODUCTIVE ACTIVITIES				
Interventions aiming at improving policy frameworks				
Instruments	**TA and capacity building to government agencies and self-government bodies**	**Support for public –provide dialogue**	**Direct advocacy for specific polices**	**etc**
Industrial policy, business environment, institutional strengthening	- Industrial policy formulation, business environment reform and effective institutional infrastructure for SME development - Policies and strategies that further industrial competitiveness; - Proposals for adjusting existing legal/regulatory frameworks governing business entry and Operations.	Strengthening public-private dialogues		
Rural and Women's Entrepreneurship Development	Improving capabilities of administrative institutions that facilitate entrepreneurial initiatives, particularly of women and the youth	Encouraging entrepreneurial initiatives facilitated by government and civil society institutions, resulting in pro-poor economic growth	EGM and manuals for promoting rural, women and youth entrepreneurship produced and disseminated.	
SME Cluster and Network Development				
Agro-processing and Value Chain Development				
Rural Energy for Productive Use	Equipping Rural communities with reliable energy services for e.g. water pumping for irrigation, grain grinding and handicrafts activities			
Sustainable Production in Poor Communities	Poor communities trained and equipped with technologies to reduce health and environmental exposure		Information materials, books, reports, brochures, web pages published. Country/sector specific publications, benchmarks, country rankings, indices and other country specific indicators relevant for investment	

			developed.	
Technology Diffusion				
Promotion of Domestic Investment, FDI and Alliances	Governments assisted with the development of Effective policies, strategies and incentives			
Interventions aiming at strengthening support bodies and structures				
Instruments	**Hardware/Software support for BDS providers and BMOs**	**Training for BDS providers and consulting**	**Development of Service packages for BDS Providers and BMOs**	**etc**
Industrial policy, business environment, institutional strengthening		Training of staff of public and private sector institutions capable of designing and implementing SME policies and support programmes;	Industrial performance assessment tools, institutional diagnosis tools, policy/legal and regulatory analysis tools	
Rural and Women's Entrepreneurship Development		Equipping Institutions and training centers with capabilities to operate entrepreneurial trainings	New curricula for human resource development institutions	
SME Cluster and Network Development			Methodologies and/or tools for cluster, networking and value chain development	
Agro-processing and Value Chain Development		Training for staff in support bodies (technology centres, extension services,		
Rural Energy for Productive Use	- Community ICT equipment and household lighting systems installed - Support structure (bodies) for modern energy services introduced and operational		Innovative financial schemes to support rural energy mini-grids and linked income generation activities;	
Sustainable Production in Poor Communities	Capacity-building of environmental and safety control bodies.			
Technology Diffusion	Enhanced contacts and network linkages related to access to technology information	Training for staff in support bodies	Technical manuals and information tools dedicated to foster technology upgrade are	

			prepared	
Promotion of Domestic Investment, FDI and Alliances	- Institutional capacities established to monitor investment related decisions and to measure impact on local economy;	Training to monitor investment related decisions and to	Methodologies for assessing and calibrating investment climate parameters	Training packages developed
Direct Support to Enterprises				
Instruments	**Hardware/Software support**	**Training and Consulting**	**Brokerage**	**etc**
Industrial policy, business environment, institutional strengthening				
Rural and Women's Entrepreneurship Development				
SME Cluster and Network Development				
Agro-processing and Value Chain Development	Upgrading of client SMEs.		Agro-value chains developed	
Rural Energy for Productive Use	Enterprises assisted in the use of renewable energy sources for energy needs.			
Sustainable Production in Poor Communities	Households equipped with well filters and trained to use technologies to reduce health and environmental exposure			
Technology Diffusion				
Promotion of Domestic Investment, FDI and Alliances				

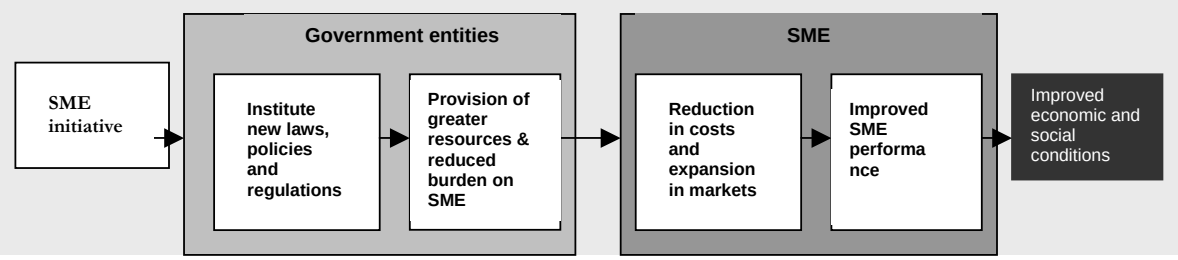
As a next step, logic models need to be developed for the selected projects/programs. The two figures below present logic models for initiatives targeting BDS providers/BMOs (Figures 13) and government entities (Figure 14).

Figure 13 Logic Model for Initiative targeting BDS providers



Source: E. Oldsman and K.Hallberg,, "Framework for Evaluating the Impact of Small Enterprise Initiatives", Nexus Associates, Inc., 2002, Washington DC

Figure 14 Logic Model for Initiative targeting Government Entities



Source: as above

These logframes could be used as general and broad guides for developing concrete logrames for particular projects and theories of change of particular Programme components/projects, some of which may include only direct assistance to SMEs, some - both direct assistance and support to strengthen BDS providers and BMOs, and some- in addition or as stand-alone projects- technical assistance to governments.

In demonstrating that a particular intervention resulted in a specific outcome, certain conditions need to be met[73]:

- First, changes engendered by the intervention need to be shown to produce the effect, or in other words, the outcome must be responsive to the intervention;.
- Second, other plausible explanations for the observed outcome need to be ruled out, i.e. rival hypothesis must be disproved;
- Third, the mechanism by which the outcome was achieved need to be explained- a theory linking the intervention to the outcome need to be articulated.
- Finally, it must be possible to replicate the results in the similar settings.

With proper research apparent correlations could be translated into plausible causal explanations. In developing logframes and results' chains particular attention should be paid to laying out the assumptions behind the intervention- these will be taken to a more detailed level at the results chain. Careful consideration should be given in developing the results chain in the part concerning the ways the interventions were expected to impact poverty. Oldsman (2003)[74] suggests that there are basically three different chains of impact for poverty reduction in BDS-programs (p. 5f). Poverty impact occurs:

- **When the owners of participating small enterprise are poor:** Some BDS-programs target specific groups for intervention such as poor farmers in a particular area. Because these programs work with business owners that are poor, improved business performance may yield sufficient income to lift these business owners (and unpaid family workers) out of poverty. These programs aim to reduce poverty within these specified groups.

[73] Source: E. Oldsman and K.Hallberg,, "Framework for Evaluating the Impact of Small Enterprise Initiatives", Nexus Associates, Inc., 2002, Washington DC

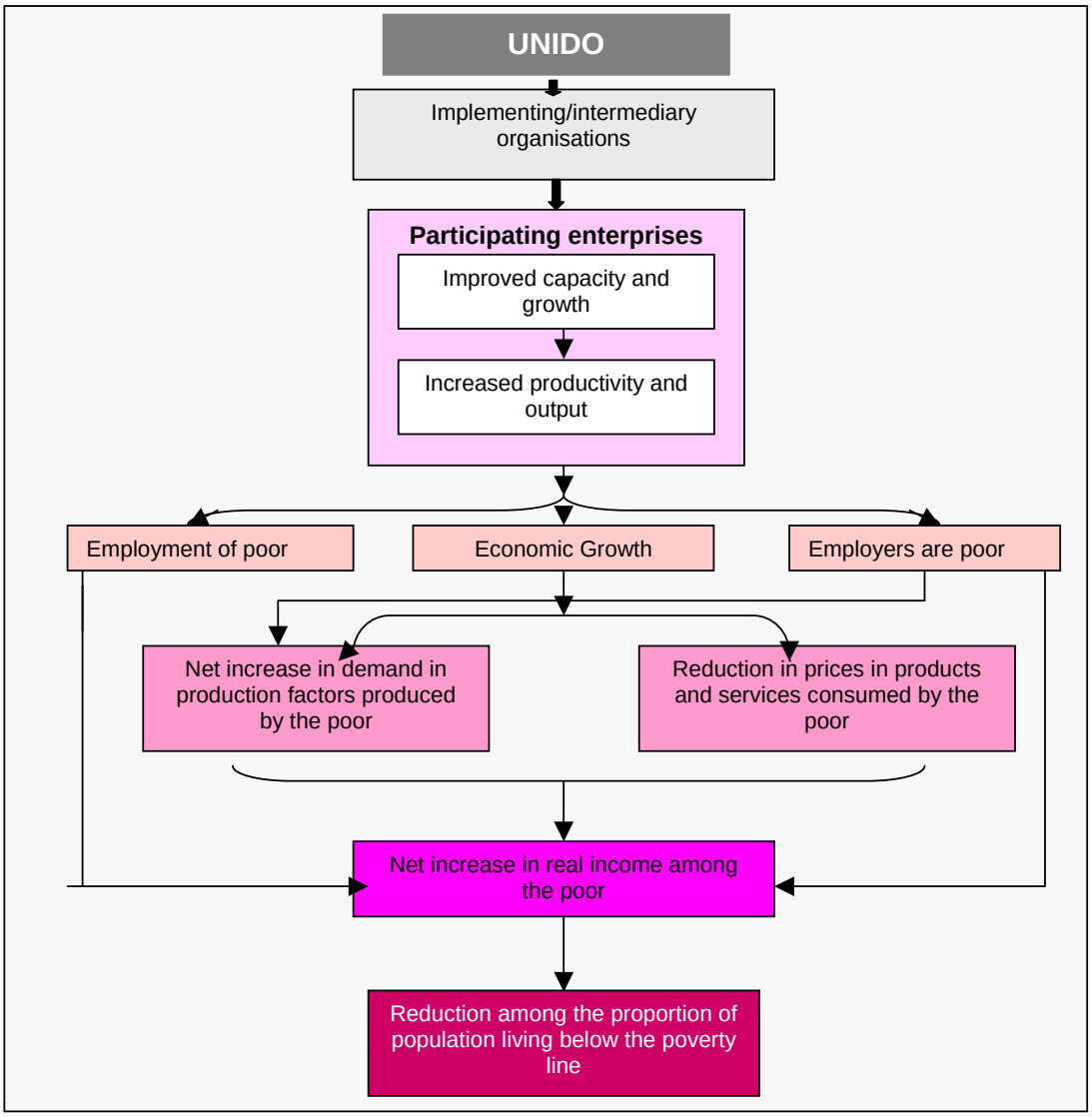
[74] Oldsman, Eric: Assessing the Poverty Impact of Small Enterprise Initiatives, prepared for the Working Group for Impact Measurement and Performance, Committee of Donor Agencies for Small Enterprise Development, Nexus Associates, Inc. 25 December 2003, 24p.

- **When participating small enterprises employ the poor**: Indeed, most BDS-programs are not designed to address business owners who are themselves poor. Those that focus on the development of BDS-markets are not aimed at particular groups of poor people; rather, these programs are intended to reduce poverty within the general population. In these instances, the links between programs and poverty reduction is more complex, hinging to a great extent on induced labour demand within participating firms and the broader economy.

- **Through economic growth generated by participating small enterprises:** The improved performance of participating small enterprises may spur economic growth leading to higher real household income and a concomitant reduction in the poverty rate (see Figure 15). Initiatives benefit the poor *to the extent* that they result in pro-poor economic growth, i.e., growth that provides greater opportunities for the poor to generate higher real income.

In Figure 15 a Results' chain for these routes is described, *without assumptions*.

Figure 15 Poverty Impact of BDS initiatives



Source: Adapted from Brigitte Späth for SDC, "Current State of the Art in Impact Assessment: With A Special View on Small Enterprise Development", 2004

The vision in Figure 15 is still short of an ideally comprehensive analysis of possible results chain. First of al it does not depict the role of assumptions. Secondly, some important interplays are overlooked. For example, it assumes a very simplistic linkage along the employment–productivity dynamics as the result of structural change, discussed below.

Trade-offs between employment and productivity growth exist, most often, in the short and medium term as a result of short-run deviations and structural and frictional changes in the economy. Structural change in the form of shifts in employment across sectors is part of the "creative destruction" process and is necessary to achieve long-run sustainable growth. Does productivity help or harm employment growth? In the longer term, in most economies, productivity and employment growth go hand in hand. This is particularly the case in the industrialized economies, but not in developing ones, as is well documented in ILO (2005)[75]. Job destruction is one of the aspects if impact analysis in BDS projects (and, private sector development projects in general) that needs to be taken into account, incorporating methods for exploring if and to what extent displacement is occurring. The results of this research should be reported alongside other programme results.

6.2 Defining Indicators

The next stage is selecting indicators for Outputs, Outcomes and Impacts. As already mentioned, the RRS exercise carried out at UNIDO currently is addressing the issue of improving the sets of indicators. Therefore, here, we would only direct the reader to a few important papers [76] for a recommended list of indictors for projects targeting BDS projects. .

In Section 5.2.1 the discussion around poverty indicators was started. In what follows below this discussion is now taken a step further with reference to SME development projects.

While programs may measure other indicators for poverty reduction (including indicators for capturing *access* aspects of poverty) and/or for other goals, it is recommended that the following four indicators must be included at the poverty reduction level:

- Number of poor people benefited by the program on an annual basis – this indicator should include a clear definition of "benefited" which represents a significant improvement to a poor person's situation. Only those people under the poverty line as officially defined in the country in question may be included.

[75] "World employment report 2004-05: Employment, productivity and poverty reduction", International Labour Organization, Geneva 2001, First published 2005

[76] "A Guide to IFC TAAC Indicators", developed by E. Oldsman, Nexus Associates, 2005; and E. Oldsman and K.Hallberg,, "Framework for Evaluating the Impact of Small Enterprise Initiatives", Nexus Associates, Inc., 2002, Washington DC

- Amount change in income as a result of the programme on an annual basis – this may be calculated at the enterprise (profits and wages) or household levels but must include an estimate of the percentage of any additional income (or income losses) which is received (or lost) by poor people as defined by the national poverty line in the relevant country. (An alternative would be average change in income for people affected by the program – calculated as an actual amount).

- Change in the number of full time equivalent jobs as a result of the programme on an annual basis (i.e. the jobs created in a particular year minus those lost in the same year) – programmes may also choose to report jobs saved or sustained but this should be reported separately.

- Change in the number of people in poverty as a result of the programme on an annual basis – poverty should be defined by the national poverty line in the relevant country

It is our understanding that Donor Committee for Enterprise Development[77] is working on recommendations along the same lines to be introduced by a number of agencies.

In order to measure these indicators effectively, programmes must identify what they will measure in order to calculate the indicators above. For example, to calculate additional income, a programme may choose to measure changes in profits of effected enterprises and changes in wages for those enterprises. Programmes are free to choose how to calculate each indicator as long as there is at least one indicator associated with each change described in the programme causal model and . The document outlining the programme causal model includes indicators for each change in the causal model.

The measurements and calculations need to be clear, transparent, well justified and documented. Proxy indicators are acceptable as long as clear and reasonable justification is presented. All data must be disaggregated by gender.

6.3 Choosing Impact Assessment Methodologies

Methods that could be used to evaluate the impacts of interventions along the causal chains and ultimately on economic and social conditions could include as we have discussed in experiments with random assignments, quasi-experiments, non – experiments, participatory evaluation and other methods.

A particular note here on reflexive controls. It is often difficult to obtain data on SMEs that did not participate in the program. In this case the only option is to base the impact assessment on the information on participants before and after the intervention. In order the results to be valid a presumption need to be made that no other factor –other than the program- contributed to the outcome. Because this is seldom true, however, results from studies based exclusively on reflexive controls are often treated with substantial skepticism. That said, the approach may be sufficient when there is a clear and close relationship between the program and outcomes of interest. This could be the case for a number of types of UNIDO interventions, e.g. projects in agri-processing.

[77] http://www.sedonors.org/

Measurements need to be taken in close proximity- before and after the intervention.

Oldsman and Hallberg (2002) suggest a menu of 4 possible methodologies for evaluating impacts of BDS projects, see Table 4.

Table 4: Choice of method for impact evaluation

Question	Methodologies	Design
Did the initiative result in a sustainable market for BDS?	Experiment with random assignment	Randomly assign SMEs to control and treatment group and compare expenditure on outside service providers over time
	Quasi-experiment with constructed controls	Compare sales growth of service providers in the treatment and control group statistically controlling for selection and other extraneous factors
	Non –experiments with reflexive controls	Compare baseline and post-intervention sales of service providers in the treatment group
	Participant judgment and expert opinion	Ask SMEs and/or service providers whether they believe that the program has resulted in a sustainable market
Did the initiative result in stronger sales among the enterprises which received the services?	Experiment with random assignment	Randomly assign SMEs to control and treatment group and compare sales growth over time
	Quasi-experiment with constructed controls	Compare sales growth of SMEs in the treatment and control group statistically controlling for selection and other extraneous factors
	Non –experiments with reflexive controls	Compare baseline and post-intervention sales of SMEs in the treatment group
	Participant judgment and expert opinion	Ask SMEs in the treatment group whether the intervention has affected their sales
Did the initiative result in a lower unemployment?	Experiment with random assignment	Select comparable regions within the country and in which to initiate the program and compare unemployment rates over time
	Quasi-experiment with constructed controls	Compare unemployment rates in different regions of the country, statistically controlling for selection and other extraneous factors
	Non –experiments with reflexive controls	Compare baseline and post-intervention unemployment rates in regions in which the program operated
	Participant judgment and expert opinion	Ask experts whether they believe that unemployment rates declined as a result of the program

Source: Adapted from E. Oldsman and K.Hallberg,, "Framework for Evaluating the Impact of Small Enterprise Initiatives", Nexus Associates, Inc., 2002, Washington DC

6.4 Examples of applying TBE approach

6.4.1 Results chain development: example of Agribusiness development projects

UNIDO projects in the Component of Agri-processing could serve as a good example where relatively simply results chains could be developed and simplified quantitative evaluations conducted – reflexive controls. We have developed a template of such a results chain in Figure 16.

Note: we have not strictly followed the list of outcome indicators specified in the Programmes and Budgets document for the Component of Agri-processsing (see Figure 17). It is our understanding that under the current RBM/RRS exercise these lists might

be revised.

Whether or not reflexive controls method could be applied depends on the availability of baseline and monitoring data on the performance of enterprises immediately before and after the participation of enterprises in the project activities.

Quantitative evaluation would need to be complemented with by qualitative studies, and in particular with identification of changes in institutions and behaviors.

Figure 17 Outcomes and outcome indicators specified for agri-related programmes in P&B 2008-2009 document

Outcomes	*Performance indicators*	*Sources of verification*
Support bodies (technology centres, extension services, private consulting firms, etc.) offer services on sustainable and demand basis	• Increased demand from SMEs for support services (trainings, consultancies, etc.) • Number of centres that offer services regularly	• Records and statistics of technology centres • Company based records • UNIDO sample surveys on performance of companies and employment • Statistics/records (population/income) of municipalities where support is provided
SMEs apply improved agro-processing technologies and offer improved products for accessing markets	• Number of enterprises with new/improved capacity	
Agro-value chain efficiency enhanced	• Number of cases where agriculture-industry-market linkages have been improved	

Source: Programme and Budget document, 2008-2009

Output Indicators would include, for example:

- Number of entrepreneurs trained;
- Number of enterprises consulted
- Number of intermediate agencies assisted, and so on.

The suggestions for Short-term Outcome Indicators are as follows:

- Number of entrepreneurs putting the business skills learnt and new technologies in practice;
- Reduced Costs by Participating Enterprises;
- BDS support bodies adopt provision of new services.

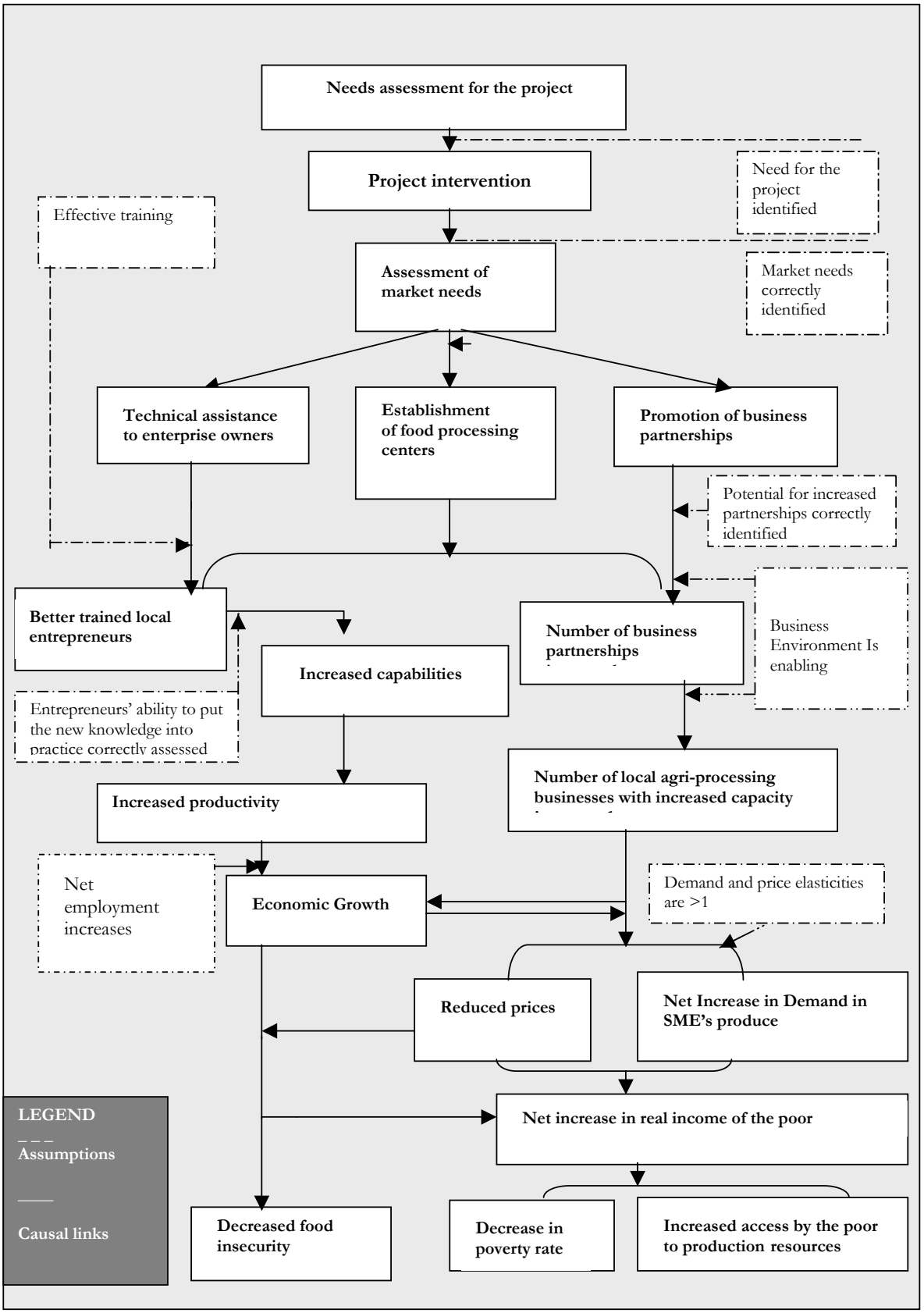
The suggestions for Long-term Outcome Indicators are as follows:

- Number of support bodies providing BDS services on a sustainable basis;
- No of SMEs (both participant and non-participant) adopting new technologies;
- No of business partnerships created and sustained .

Impact Indicators could be chosen as follows:

- Number of people lifted out of poverty in the project area
- Number of poor with access to productive resources
- Sales of participating enterprises
- Export of participating enterprises
- Number of de-novo SME created due the program
- No of jobs created /net/

Figure 16: Results Chain for Agri-processing projects



It is important to capture Outcome Indicators at various levels, reflecting the causal links in the results chain. If then the measurements of the indicators show that the anticipated positive developments did not occur or their magnitude was not in line with expected targets, then particular attention during the impact evaluation will be paid to those points in the chain and underlying assumptions.

It should be possible to use a not very expensive quasi experimental method for impact evaluation for such programs, since identifying control groups may not pose a significant challenge. If measurements for monitoring indicators are taken regularly, it should be possible to use also reflexive controls. Indeed, it is highly recommended that these are combined with RPAs.

6.4.2 Different complexities: Generic Model or Evaluating Impacts of SME Cluster Development Initiatives

It should be recognized that different Program components under a given Major Program Area might vary significantly in their complexity and hence in the complexities of designing the detailed results chains and developing impact evaluation models will suffice.

As an illustration we shall reflect here to the Program Component on 'Developing SME Clusters'. In comparison to agri-processing projects, these initiatives are rather complex , since they are by nature network project and combine building up business capacities with building up trust and interaction within the network.

While literature on institutional planning and development project evaluation is rich and extensive, only very little specifically related to monitoring and evaluating the performance of SME driven networks can be found. This refers to the fact that network evaluation is still a relatively new and highly complex field of work. In almost every case an appropriate evaluation model have to be developed to consider the specific situation and to fit the needs and requirements of the sponsor.

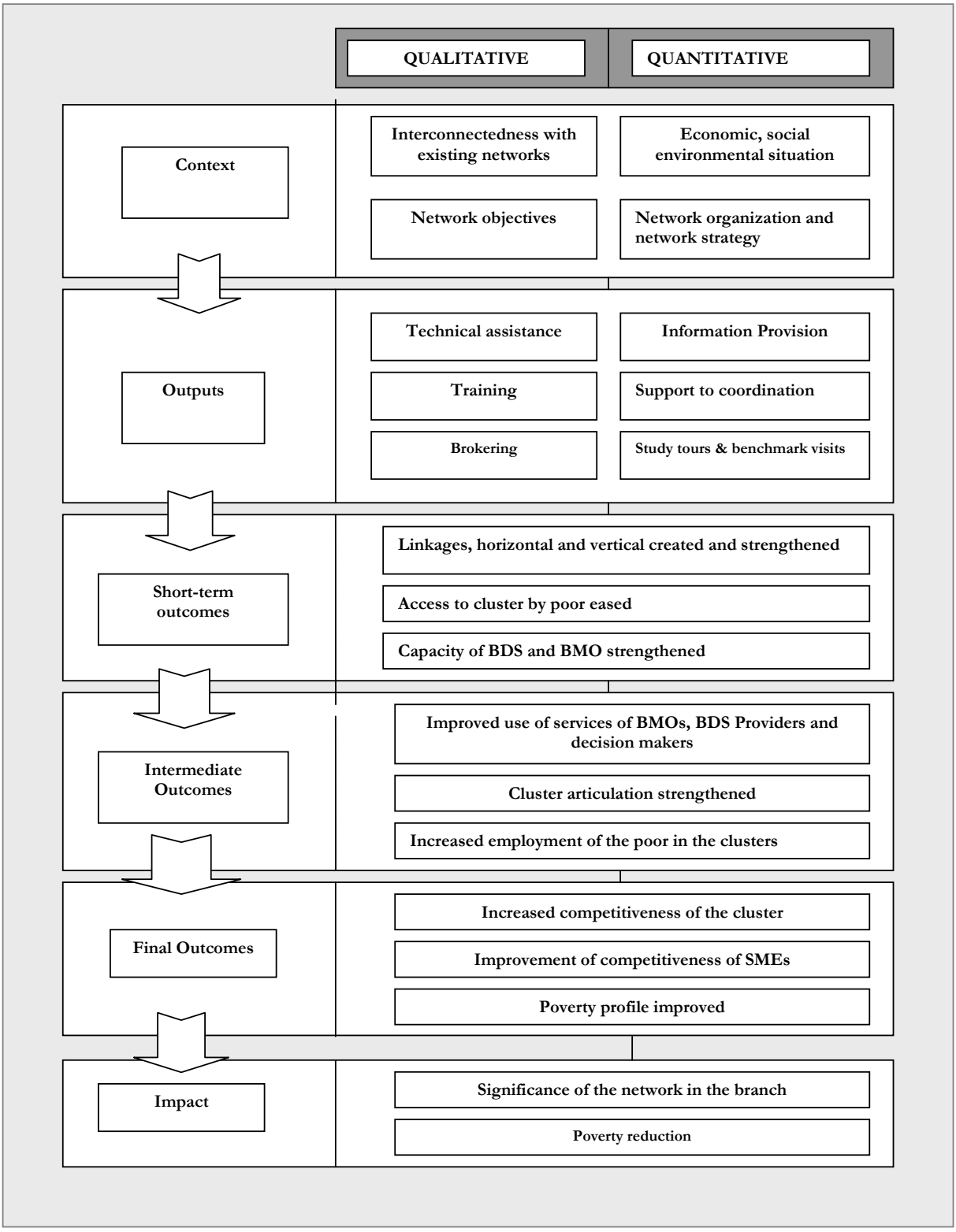
The SME Clusters development Department has developed a draft approach to impact evaluation for their projects, but it is still in its draft form. Without developing the results chain here, we would only suggest using the following generic model for impact evaluation of the cluster initiatives as in Figure 18.

The network evaluation should begin with the identification of the network potentials and the context of the network. On the basis of this background information the network processes can be analyzed. In later development stages the results of the network activities and the impacts for the region or the growth industries in a given region can be compared with the specific network's objectives.

Taking the complexity of SME driven networks into account a network evaluation is demanding and time consuming. This model allows a broad and comprehensive approach and appears to be an appropriate way of evaluating SME-driven networks.

Specific indicators can be identified which have to be operationalized and which can be measured or interpreted on a quantitative and qualitative basis

Figure 18 Generic model for impact evaluation of the cluster initiatives



7. SUMMARY OF RECOMMENDATIONS

What follows below is a summary of recommendations, which were made throughout the text on the actions needed to be taken to instutionalise the process and conduct of quality impact evaluations at UNIDO. The subtasks are presented in accordance with perceived priorities, and time frames (from short- to long-term).

1. Develop toolkits for Impact Evaluations for UNIDO TC thematic areas and main portfolios:

- Finalise the Toolkit for Impact Evaluations falling under the BDS thematic area and extend the approach to projects falling under other thematic areas;

- Improve the list of impact/outcome indicators according to the main program portfolios. Hopefully, this task will be achieved within the process of introducing RBM;

- Conduct a detailed review of all of UNIDO projects with the aim of refining the list of main thematic areas along, program portfolios and intervention instruments, along which a library of standardized results chains (with indicators) could be developed.

- Develop a Library of Intervention theories for the main program portfolios;

- Organise systematic learning on a number of categories of themes: targets, opportunities, actions, and actors for the main program portfolios.

2. Introduce/improve ex-ante assessments as part of project/program design:

- Review a selected list of around 10 projects to assess those in terms of their impact evaluability, including: availability of data and resources for conducting impact evaluations; availability and quality of logframes, problem trees and M&E indicators and targets, and so on;

- Take up measures aimed at introducing/improving ex-ante assessments. In particular, ensure that the following list below is included/addressed in ex-ante impact assessments:
  - Problem analyses and needs assessments;
  - Assessments of alternative delivery mechanisms;
  - Logframes and results chains as part of the project designs;
  - Clearly specified objectives with the corresponding list of activities;
  - Clearly specified, measurable and verifiable indicators at all levels (output, outcome, impact) with targets at output and outcome level, tied up with anticipated time-frames, as well as projections of impacts of the project at specified time horizons;
  - Reflections on the lessons from the past, i.e. discussion on which were the most important factors that proved to work and not work in similar projects in the past and how the proposed project is addressing the limitations to eliminate/mitigate the threat factors;
  - Plans for conducting regular monitoring during the project and evaluations (types, methods, time-frames).

3. Incorporate capturing early impacts in project completion evaluations/program evaluations:

- Review a sample of Project evaluations to assess whether and how early impacts could have been captured;

- Review a sample of IP evaluations to assess what impacts and how could have been captured.

4. Use cluster/thematic impact evaluations more and specifically for smaller projects. Develop guidelines for conducting cluster/thematic impact evaluations.

5. Commission/conduct 3-5 fully fledged impact evaluations in the coming 2 years.

- Pre-select a shortlist of around 10 projects. Criteria for such a selection are suggested below:
  - significance in UNIDO portfolio;
  - potentially growing area for UNIDO;
  - representing the major underlying theories of change in that area;
  - representing different thematic areas;
  - evaluability (at least some baseline or monitoring data available);
  - projects' base is still in the field (perhaps in the form of subsequent phases of the project);
  - ideally allowing for different methodologies to be employed.

- Conduct a theory-based analysis of these selected projects, based on the project documents, results reported in available terminal evaluations, program and cross-cutting studies;

- Prepare an overview of reported results against those intended and summary of previous analyses of performance, including reported deficiencies in data on impacts;

- Generate key questions concerning impacts and potential strategies for impact evaluation;

- Hold discussions with selected thematic area/sub-area teams to determine their potential collaboration in conducting pilot evaluations;

- Identify the list of around 5 Pilot impact evaluations to be carried out in the next 1-2 years out of the list above;

- Develop detailed impact evaluation strategies, in order to provide a meaningful exploration of outcomes/ impact for the selected 5 projects/programs;

- Conduct pilot evaluations. In conducting these pilot impact evaluations it is important to ensure assessing also:
  - Multiplier effects;
  - Unintended impacts;
  - Sustainability of the projects;

- Relation of impacts to program costs .

6. Based on the results of pilot impact evaluations, refine all the guides developed as support instruments for conducting impact evaluations.

REFERENCES

Abadie A, Diamond A, and Hainmueller J., Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program, July 2007, Harvard University, Cambridge, MA USA

AFD/ODI, "A Comparative Study of Evaluation Policies and Practices in Development Agencies', December 2007

Baker, Judy (2000) Evaluating the Poverty Impact of Projects World Bank.

Carvalho S and White H., Theory-Based Evaluation: The Case of Social Funds, the World Bank, Washington DC, 2004

Center for Global Development:" When Will we ever learn? Improving Lives Through Impact Evaluation", Report of the Evaluation GAP Working Group, Washington DC, 2006

Comparative review of lessons learned from 20 UNIDO Integrated Programmes, UNIDO, Evaluation Group, Vienna, 2007

EU Commission 'Ex ante evaluation a practical guide for preparing proposals for expenditure programmes", , 2001

GEF Evaluation Office/Foundations for Success, "GEF Impact Evaluation: Final Report on a Proposed Approach to GEF. Impact Evaluation. Impact Evaluation Information Document No. 2'

GEF (2006) "GEF Impact Evaluation – Final Report on a Proposed Approach", Working Document Prepared for the GEF Evaluation Office, Foundations of Success, Washington D.C.

GTZ 2004: Result-based Monitoring: Guidelines for Technical Cooperation Projects and Programmes Development Assistance Committee Working Party on Aid Evaluation (2002).

Haarhuis K., Leeuw C.M. and F.L. (2004) "Fighting Governmental corruption: The New World Bank Programme Evaluated", Journal of International Development 16: 547-561

International Labour Organization , World employment report 2004-05: Employment, productivity and poverty reduction , Geneva 2001, First published 2005

Leeuw, F.L. (2003) "Reconstructing Program Theories: Methods Available and Problems to be Solved", American Journal of Evaluation 24(1): 5-20.

Morra L, and Friedlander A., "Case Study Evaluations', WB/OED, 1999

NONIE, Report by Subgroup 2 on "Impact evaluation guidance", draft, January 08

OECD (2003) OECD Environmental Indicators: Development Measurement and Use, OECD, Paris.

OECD, Promoting Pro-Poor Growth: Harmonising ex ante poverty impact assessment", 2006

OECD-DAC, Glossary of Key Terms in Evaluation and Results Based Management 2002

Oldsman E. and Hallberg, K, Framework for Evaluating the Impact of Small Enterprise Initiatives, Nexus Associates, Inc., 2002, Washington DC

Oldsman, E. "A Guide to IFC TAAC Indicators, developed by E. Nexus Associates, 2005;

Oldsman, Eric: Assessing the Poverty Impact of Small Enterprise Initiatives, prepared for the Working Group for Impact Measurement and Performance, Committee of Donor Agencies for Small Enterprise Development, Nexus Associates, Inc. 25 December 2003, 24p.

Späth B, Current State of the Art in Impact Assessment: With A Special View on Small Enterprise

Development, SDC, 2004

UNIDO, RBM Implementation Plan with milestones for the biennium 2008-2009, November 2007

UNIDO, Independent Evaluation of UNIDO Investment and Technology Promotions Office in Italy

UNIDO, Evaluation Group, Comparative review of lessons learned from 20 UNIDO Integrated Programmes", Vienna, 2007

UNIDO, Programmes and Budgets, 2008-2009, Vienna

Vaessen J. and Todd D., Methodological Challenges in Impact Evaluation: The Case of the Global Environment Facility (GEF), University of Untwerp, Institute of Development Policy and Management, 2007

Weiss Carol H., (1998) Evaluation: Methods for Studying Programs & Policies 2nd edition. Prentice Hall

White H., Bamberger M., Flanagan A. and Sinha S., "Impact Evaluation in Official Development Agencies', 2008

White H., "Impact Evaluation – the Experience of the Independent Evaluation Group of the World Bank", IEG, WB, Washington DC, September 2006

White H., "Challenges in evaluating development effectiveness', IDS Working Paper 242, Institute of Development Studies, Brighton, UK, March 2005

White H., Impact evaluation: an overview and some issues for discussion, note prepared by the Independent Evaluation Group of the World Bank, in collaboration with the DAC Secretariat, 4th meeting of the DAC Network on Development Evaluation, 30 – 31 March 2006.

WB/OED, "Monitoring and Evaluation: Some Tools, Methods & Approaches', Washington DC, 2004

WB/IEG, Conducting quality impact evaluations under budget, time and data constraints, 2006

ANNEXES

Annex 1 TOR

Duties of the international expert:

The international expert will carry out an exploratory research study with a view to developing an improved methodology for impact evaluation in UNIDO. The study will be carried out in three stages:

1. Needs analysis including interviews at UNIDO HQ to identify the expectations and attitudes of UNIDO senior staff from different branches vis-a-vis impact evaluation;

2. Desk study (comparison of existing approaches and testing their specific matches with UNIDO requirements);

3. Outline of an improved UNIDO methodology for impact evaluation and recommendations for next steps.

The research work will be carried out from home base and through interviews at UNIDO HQ. In order to design and execute the research in a smooth and participative manner the expert shall be available for meetings with the UNIDO Evaluation Group and other staff at the UNIDO Headquarters as appropriate.

Qualifications

1. University degree in economics
2. Experience in impact evaluation of development projects
3. Experience in industrial development issues

Languages:

The consultant must be fluent in English and have excellent English writing skills

	Tasks	Duration	Venue	Deliverables
1	Establishment of a detailed research plan with milestones (first meeting at UNIDO HQ)			Research plan with milestones and detailed description of deliverables
2	Needs analysis; interviews with UNIDO senior staff from different branches to identify their expectations and attitudes			Notes
2	Needs analysis; desk study	3 days	Home	Analytical paper establishing the specific needs and constraints for impact evaluation in UNIDO
3	Desk study of UNIDO's specific requirements	7 days	Home	Analytical paper comparing

	for impact evaluation and of existing approaches to evaluating the impact of development projects			existing approaches with a view to their potential match with UNIDO needs
4	Presentation and discussion of results (second meeting at UNIDO HQ)	3 days	HQ	Draft questionnaires
5	Preparing an outline for an improved impact evaluation methodology for UNIDO	6 days	Home	Outline paper with recommendations for next steps
6	Final presentation and discussion (third meeting at UNIDO HQ)	2 days	HQ	Slides
7	Preparation of final report	2 days	Home	Final report
	Total	**28 days**		

Annex 2. Estimated Costs from Selected Impact Evaluations

Programs	Methods and Sample size	Evaluation Costs		
		Data Collection (% of total)	Analysis (% of total)	Total USD
Nicaragua School-based management	Matched comparison of schools, two follow-up surveys and focus group surveys (242 schools, 400 teachers, 3000 students)	**35**	**65**	**495000**
El Salvador School-based management	Random assignments (200 schools, 2000 students)	**59**	**41**	**443.000**
Columbia Voucher Program	Matched comparison of schools (150 schools, 2000 students)	**69**	**31**	**226.000**
Armenia Social Fund	Matched comparison group using LSMS and facility survey (2260 households and 53 facilities)	**18**	**82**	**111.000**
Bolivia Social Fund	Random Assignment and matched comparison with baseline and follow-up surveys (7000 individuals and 380 facilities)	**69**	**31**	**878.000**
Honduras Social Fund	Matched comparison with pipeline communities, special household and facilities surveys (2320 households and 81 facilities)	**32**	**68**	**263.000**
Nicaragua Social Fund	Matched comparison using LSMS and facilities survey (2000 households and 400 facilities)	**56**	**41**	**449.000**
Trinidad and Tobago Youth Training	Tracer surveys, sample from existing national data (2500)	**63**	**37**	**238.000**
Argentina TRABAJAR workfare	Matched comparison using national survey and follow-up (2800 individuals, 120 projects)	**90**	**10**	**390.000**
Bangladesh Food for Education	Unmatched comparison using statistical controls from national data (3625 households)	**36**	**54**	**140.000**
Czech Republic Active Labor Market Program	Matched comparison group using national data and follow-up survey (9477)	**20**	**80**	**250.000**
Mexico PROGRESA	Random assignment of localities, and also non-experimental methods (24407 households)	**17**	**83**	**2.415.000**
Zambia Project Recovery Project	Matched comparison with pipeline communities, using national survey with special module follow-up plus facility survey (2950 and 100 facilities)	**76**	**24**	**174000**
Average from Firms		**85**	**15**	**350.000**

Source: John Blomquist, "Impact Evaluation of Social Programs: A Policy Perspective', Social Protection Human Development Network, 2003

Annex 3. Mapping Corporate Outcome for Major Programme C according to the three intervention modes

	Government and business environment	Interventions aiming at strengthening BDS providers, BMOs and other support structures	Direct support to pilot companies
Impact (=positive and negative long-term changes from using output, intended or unintended)	Better public service provided by government agencies New/improved industrial processes, services, products, business models lead to desirable economic, social, environmental impact	Better economic/environmental performance of support bodies' clients (companies, cooperatives...) New/improved industrial processes, services, products, business models lead to desirable economic, social, environmental impact	New/improved industrial processes, services, products, business models lead to desirable economic, social, environmental impact
Outcome (=positive & short or medium-term changes from using output: Why do it? What are the effects?)	Better performance of the governance system (efficiency and economic/social/environmental effectiveness) and favorable business environment	New/improved support bodies (SB) and their services: SB more need-oriented (competence; specialization; accessibility); SB are accredited/recognized; their services are more need-oriented (quality; range; cost; access; etc); new financing schemes are operational;	Enhanced performance of pilot companies (by applying what works and learning from what does not work)
	Favorable business environment (thanks to government adoption of recommendations, new laws, regulations; creation/reorganization of administrative structures; adoption of new monitoring/statistics system or officials applying enhanced competence in their work)	Private and public stakeholders use improved diagnostic and analytical tools for product and sector analysis of competitiveness at sector and product level	Private and public stakeholders use improved diagnostic and analytical tools for product and sector analysis of competitiveness at sector and product level
	Monitoring and assessment of industrial performance and competitiveness at national, regional and global levels institutionalized	Quality and productivity centres are established or deliver improved services to enterprises	Pilot enterprises institutionalize a quality culture using quality management systems for continuous improvement
	Public and private sector institutions involved in the formulation, implementation and monitoring of industrial strategies, policies and programmes apply analytical	National upgrading schemes are operational and pilot enterprises are upgraded	Joint initiatives undertaken within cluster

	and diagnostic tools		
	Public and private sector institutions develop and implement SME development strategies, policies and support programmes	Standards: National Standards Bodies (NSBs) enable enterprises to comply with WTO agreements, especially on TBT/SPS, and take into account private sector/exporter and consumer needs	Improved competitiveness of SMEs in supported clusters and networks
	Administrative institutions supportive of entrepreneurial initiatives, particularly of the youth and women	Metrology/calibration and product testing: Effective local capacity, operating according to international best practices and standards, provides services to local testers and producers/exporters and for consumer protection	SMEs apply improved agroprocessing technologies and offer improved products for accessing markets
	Government and private institutions (support bodies) implement cluster and networking development initiatives	Accreditation: Local or regional accreditation schemes able to assess the performance of the local laboratory, inspection, certification body activities	Export-oriented agro-industries use good hygiene practices, hazard analysis and critical control point systems introduced by UNIDO supported service providers
	Agro-value chain efficiency enhanced	Certification: Internationally recognized certification services, available for export and local enterprises for international and private buyer and retailer standards in areas such as quality, environment, social accountability, food safety, traceability, etc.	SMEs adopt the export consortia approach
	Policy institutions adopt the export consortia approach to promote SMEs	Consumer affairs: Consumer associations defend and promote consumer rights based on national policies in line with international best practices	SMEs assisted by UNIDO introduce corporate social responsibility standards
	Countries that have been assisted comply with their obligations under the Montreal Protocol	Public and private sector institutions involved in the formulation, implementation and monitoring of industrial strategies, policies and programmes apply analytical and diagnostic tools	Enterprises/farms meet international market standards owing to adoption of non-ODS-based technologies
	Increased capacity of recipient countries to reduce stockpiles of POPs	Public and private sector institutions develop and implement SME development strategies, policies and support programmes	Countries supported by UNIDO use best environmental practices (BEP) and best available technologies (BAT)
	Countries supported by UNIDO use best environmental practices (BEP) and best available technologies (BAT)	Sustainable BDS, information services and ICT support offered by private and public institutions	
		Improved effectiveness of public-private consultative mechanisms	
		Institutions offering training and support services for people in rural areas to build entrepreneurial capabilities on	

		sustainable basis Administrative institutions supportive of entrepreneurial initiatives, particularly of the youth and women	
	Support to Government (training of officials; conducting policy studies and advice; raising awareness; organizing workshops & study tours; etc)	Methodologies and best practices to promote rural, women and youth entrepreneurship development used by human resource development and support institutions Government and private institutions (support bodies) implement cluster and networking development initiatives Institutions and enterprises that support clusters use UNIDO methodologies and/or tools for pro-poor economic development Support bodies (technology centres, extension services, private consulting firms, etc.) offer services on sustainable and demand basis Industry support organizations provide support services to export oriented companies on a sustainable basis Institutions/service providers inform SMEs about CSR standards and their implementation at company level on a regular basis	
Output (=products, goods or services after completing activities).	Concrete knowledge; drafts law & proposals to government; recommendations; administrative reorganization plans/proposals; new monitoring system demonstrated on a pilot scale	Support to support bodies(training staff; providing equipment & advice; conducting studies; raising awareness; etc)	Experiences and lessons learnt from demonstration/experimentation of new or improved industrial processes, services, products, business models
Activities (=actions taken using inputs)	Consulting and training services to the addressing Government bodies	New/improved support bodies (SB) and their services: SB more needs oriented (competence; specialization; accessibility); SB are accredited/recognized; their services are more needs oriented (quality; range; cost; access; etc); new financing schemes are operational;	Support to pilot companies(raising awareness; training staff; providing pilot financing schemes; conducting benchmark studies;etc)

Source: Adapted, UNIDO EVA