

---

**The Sustainable Development Ambition of Fair Trade:**  
An Analysis of Theories and Practices for Evaluation of Sustainable  
Development

---



Picture from [www.oxfammagasinsdumonde.be](http://www.oxfammagasinsdumonde.be)

Jill de Bruijn  
Master International Development Studies (MID)  
March 2013



---

**The Sustainable Development Ambition of Fair Trade:**  
An Analysis of Theories and Practices for Evaluation of Sustainable  
Development

---

Master thesis by:

Jill de Bruijn

Reg. number: 710625-136-130

MSc programme: International Development Studies

Specialisation: Communication, Technology and Policy

Chair group: Law and Governance

Supervisors:

Dr. Ir. Otto Hospes

WUR Public Administration and Policy Group

Sophie Tack

Oxfam Magasins du Monde

Chairwoman of the Oxfam International Fair Trade Working Group



## Abstract

Starting off as a niche market in the 1980's, Fair Trade has now turned into a successful concept and many traditional trade actors have jumped at the market to claim a piece of this lucrative Fair Trade pie. Fair Trade organisations that run programmes in the South to support the activities of producer groups, and build their capacity, feel a strong urge to promote their unique approach to Fair Trade and firmly demonstrate their contribution to sustainable development and poverty reduction. Such an organisation is Oxfam. In this paper the question is raised if and how these organisations evaluate the impact of their programmes in the South; recognising that they try to accomplish change in a complex and context specific environment where a variety of actors, covering many sectors, are targeting the same 'multiple deprived' communities with development programmes. By means of a thorough literature study it becomes clear that the academic evaluation community is in disagreement. Some scholars argue that organisations only need to demonstrate 'what worked'; others on the other hand state that organisations need to evaluate 'what worked, for whom and under which circumstances', as this understanding will ultimately lead to improved interventions and opens the door for the reproduction of outcomes in similar programmes. These scholars also agree that evaluation in a complex environment can best be dealt with by using a theory-driven approach to evaluation. In practice however, four interviewed organisations (among which two Oxfam affiliates) indicate that the notion of complexity and use of theory for the purpose of evaluation are not in a developed stage yet; organisations currently face other challenges among which are dominant donors that decide on the rules and funding available for evaluations, or vague concepts such as sustainability, poverty and resilience that need further elaboration before changes can be evaluated. Moreover, the interviewed organisations fail to establish the causal connection between their interventions and poverty reduction and have not found an answer yet on how to evaluate these long-term objectives while being part of larger and more-inclusive system that drives change.

### Key words:

Change processes, Complexity, Evaluation, Fair Trade, Sustainability, Sustainable development, Theory-driven evaluation



## **Preface and acknowledgement**

When I lived for five years in Africa, working in the NGO sector, I was overwhelmed by the amount of actors all targeting the same local communities and authorities with similar development objectives. From a range of United Nations institutions to local and international NGOs, they all seem to be working next to each other rather than as a component of a wider development agenda. I witnessed how donors became more powerful switching from bilateral to multilateral funding, channelling funds through the United Nations system, setting priorities for development and leaving NGOs to compete and scramble for funding. Also, in an attempt to cut down on administration and reporting efforts, donors started to demand NGOs to work in consortia. Many NGOs saw themselves excluded from such consortia; others complained that coordination efforts in the consortia in both programmatic and financial areas had grown over their heads.

At the time, I already wondered how this variety of actors, all with their own agenda, could ever establish the outcomes from their intervention. And if they could, how on earth were they going to establish in this chaotic environment what caused the outcomes and if they had resulted in any structural changes?

Two years later and back in Europe working for Oxfam, I understood that Oxfam's Fair Trade Department was in the process of re-integrating their programmes in the South with producer groups in the wider development agenda from Oxfam. The success of Fair trade as a movement had gradually detached them from the Development and Humanitarian Department. While developing a strategy to position Fair trade within the context of Oxfam, it was agreed that Fair Trade activities in the South serve as a tool for sustainable development. As a consequence the question came up how to evaluate the impact of Fair Trade on development; which is where I saw an opportunity for finding answers on my previous thoughts and at the same time supporting Oxfam in gaining more knowledge about evaluation approaches to sustainable development.

This thesis would not have become what it is today without the enormous continuous support from Otto Hospes. Not only was it his article in the Broker on Line that inspired me to continue with this topic, he also managed throughout a 'pretty long time' to remain enthusiastic, encouraging and constructive about my work and progress. I would also like to thank Sophie Tack and the Oxfam Fair Trade Monitoring, Evaluation and Learning group for making me understand the wider development agenda of Fair Trade, its actual working within Oxfam

(especially within Oxfam-Wereldwinkels and Intermon Oxfam), and sharing numerous reports and information with me. Further I would like to mention two organisations that I interviewed; Vredeseilanden and Traidcraft: it was a real pleasure getting to know your organisation and your successes and concerns in evaluation. Without my reviewers this paper would have never reached a final stage: Tess Bresnan, Kim Roest and Jet Sneep, you were quick, honest and very inspiring with your review comments, thank you so much! Last but not least, I want to mention and thank my partner Filip Staes, for encouraging me throughout the process, putting up with all the weekend work and the stress and moreover, for doing all the shopping and cooking while I was working! I love you deeply for your support.

Berchem, 26 March 2012

Jill de Bruijn



# Table of contents

- Abstract.....5**
- Preface and acknowledgement.....7**
- 1 Introduction ..... 12**
  - 1.1 From development to sustainable development..... 12
  - 1.2 Modern complexity..... 12
  - 1.3 Evaluation of sustainable development impact..... 12
  - 1.4 Problem statement - successes and challenges of Fair trade..... 13
    - 1.4.1 The case of Oxfam ..... 15
  - 1.5 Aims of the research..... 16
  - 1.6 Ethical considerations ..... 19
  - 1.7 Structure of the report ..... 19
- 2 Literature study on evaluation approaches and complex environments ..... 22**
  - 2.1 Conventional evaluation approaches..... 23
  - 2.2 Theory-based evaluation approaches..... 24
    - 2.2.1 Programme theory deconstructed ..... 27
    - 2.2.2 Why theory-driven evaluation?..... 30
    - 2.2.3 Sources of theory..... 31
  - 2.3 Complexity..... 32
    - 2.3.1 Context dependency of outcomes..... 32
    - 2.3.2 Many actors and sectors..... 32
    - 2.3.3 Long-term change is difficult to monitor and measure..... 33
    - 2.3.4 Non-linear dynamics between activities and outcomes ..... 33
    - 2.3.5 Power of agency vs. social and institutional structures ..... 34
    - 2.3.6 Multi-level governance ..... 34
  - 2.4 Dominant evaluation approaches dealing with complexity ..... 34
    - 2.4.1 Theory of change ..... 35
    - 2.4.2 Realistic or realist evaluation..... 37
    - 2.4.3 Summary..... 38
- 3 Conceptualisation of sustainable development ..... 40**
  - 3.1 Our Common Future ..... 40
  - 3.2 Thick and thin interpretations..... 41
  - 3.3 The world as a system ..... 43
  - 3.4 Complexity of evaluating sustainable development programmes..... 43
- 4 Research methodology..... 45**
  - 4.1 Data collection and processing..... 45
  - 4.2 Selection of organisations ..... 47
    - 4.2.1 Vredeseilanden ..... 49
    - 4.2.2 Oxfam-Wereldwinkels..... 49

4.2.3 Traidcraft.....	50
4.2.4 Intermon Oxfam.....	51
<b>5 Evaluation approaches in practice: four case studies .....</b>	<b>52</b>
5.1 Organisation 1 – “Proving rather than improving” .....	52
5.2 Organisation 2 – “Evaluation belongs in the development department” .....	58
5.3 Organisation 3 – “Focus on our circle of influence” .....	61
5.4 Organisation 4 – “Long-term impact is inherent to Fair Trade” .....	67
<b>6 Research findings and observations.....</b>	<b>74</b>
<b>7 Conclusions.....</b>	<b>84</b>
7.1 Future research.....	86
<b>Literature .....</b>	<b>88</b>
<b>Appendix A – Interview questions.....</b>	<b>93</b>
<b>Appendix B – Methodological protocol.....</b>	<b>96</b>

# 1 Introduction

## 1.1 From development to sustainable development

At the Earth Summit in Rio de Janeiro 2012, one of the main topics was to develop an institutional framework for sustainable development (Earth Summit2012, 2012). Previously in 2000, the governments of 189 countries signed the United Nations Millennium Declaration, committing their nations to a new global partnership to reduce extreme poverty (UN, 2012); these 8 Millennium Development Goals (MDGs) range from ending poverty and hunger to improving child health and combating HIV. They are supposed to be reached by the end of 2015. All MDGs have targets that are expressed in measurable objectives. Though the 7<sup>th</sup> MDG strives to ensure environmental sustainability, critics claim that the current MDGs neglect scarcity and sustainability, and fail to implement equitable environmental solutions, like secure access to natural resources for people in poverty or action by rich countries to cut their consumption footprint (Green, 2012). At the Earth Summit, by 2012, it became clear that the global trend is to promote sustainable development goals (SDGs) instead of economic development goals. In the light of this evolution, it raises the question what does sustainable development entail, and above all, if (and how) can its effects be measured?

## 1.2 Modern complexity

Modern global issues, such as structural poverty and hunger, disasters, and conflicts, are caused by a complex mixture of causes like climate change, failure of government interventions, power abuse, conflict or war, inequality, trade and investment policies etc. Globalisation has added significantly to the complexity of humanitarian crises and suffering, involving a wide range of actors and stakeholders and creating complicated power dynamics, at national, regional and international levels. Moreover, contemporary development policies often promote sustainability and aim to generate 'multi-impact' on several disciplines and to strengthen 'multiple deprived communities' (Blamey and Mackenzie, 2007). They introduce for example new agricultural techniques that are climate neutral and respect the biodiversity of land. They advise poor people on their rights and support them to organise and form cooperatives or labour unions. They seek engagement from western governments and international institutions and advocate for 'development friendly' northern policies. At the same time these multi-impact programmes work on cross-cutting issues such as gender equality and HIV/AIDS.

## 1.3 Evaluation of sustainable development impact

Due to the increase in complexity, development agencies can no longer assume that their programme input will lead to a direct measurable outcome. The link between cause and

consequence does not follow a single or predictable logic. The complexity of development programmes and their widespread ambitions have led to ignorance and uncertainty about the impact on the target beneficiaries and their communities. Most current evaluation methods fall short in recognising complexity, they focus on measuring results and fail to explain why certain programmes work, for whom and under which circumstances (Pawson et al., 2005). Although alternative evaluation approaches are extensively discussed in contemporary literature, academics do not seem to agree on their purpose and application. Recognising that development unfolds in a more and more complex environment, another question comes to mind: do organisations that claim to deliver sustainable development measure their impact? If so, what evaluation approach do they prefer and how do they deal with complexity?

#### **1.4 Problem statement - successes and challenges of Fair trade**

Fair trade is an alternative approach to conventional trade and is based on a partnership between producers and consumers (Fairtrade International, 2011). A universally accepted definition of Fair Trade does not exist; however, the 'FINE' definition formulated by a consortium of leading Fairtrade organisations<sup>1</sup> has been widely adopted:

*"Fair Trade is a trading partnership, based on dialogue, transparency and respect, that seeks greater equity in international trade and contributes to sustainable development by offering better trading conditions to, and securing the rights of, marginalised producers and workers – especially in the South. Fair Trade Organisations, backed by consumers, are engaged in supporting producers, awareness raising and in campaigning for changes in the rules and practices of conventional international trade" (Fairtrade International, 2011).*

Fair Trade should not be confused with 'fairtrade' and 'fair trade'. Fairtrade can be understood as an element of Fair Trade; "it refers to the Fairtrade Mark, which is a particular label placed on products certified by the Fairtrade Labeling Organisations International (FLO) as being produced according to a FLO certification process which enshrines Fair Trade principles"(Tack and van der Veer, 2011). Fair trade on the other hand, refer to "the broader movement, which includes both active retailers of Fair Trade products, consumers, the volunteers in Northern Fair Trade shops the Southern co-operatives and workers, and the wider movement around global trade issues" "(Tack and van der Veer, 2011). In this study the emphasis will lie on Fair Trade

---

<sup>1</sup> Also referred to as the 'FINE' definition, marked out by FLO (Fair Trade Labeling Organisations International), IFAT (International Fair Trade Association), NEWS! (Network of European Worldshops) and EFTA (European Fair Trade Association)

programmes and Fair Trade producers in the South. Only occasionally there will be reference to fairtrade or fair trade activities.

For over three decades, the Fair trade movement has been working hard to address the issue of poverty among marginalised producers in the South and many studies show incredibly good results. In the North, Fair trade has led to increased consumer awareness about poor conditions of farmers in the South. In the North as well as in the South, sales of Fair Trade products have boomed as more and more markets, international, regional and national, embrace Fair Trade products. In the South, the number of Fair Trade producers is still on the rise. In particular, women are targeted for the recruitment of producers, providing them with an income and more power to take control over their lives and that of their children (Fair Trade Advocacy Office, 2010).

Although the success stories are many, there are also growing concerns that Fair trade may not cope with the consequences of its own success. Starting off as a niche market in the 1980's, Fair trade is bit by bit adapting to standard market conditions (Friedland, 2010). Although Northern consumers have easy access to Fair Trade products if they are sold in super markets, an identification problem has gradually emerged. Over the past ten years, the commercial branch has shown a proliferation of Fairtrade labels and consumers seem to have lost track of the difference between all the different suppliers. At the institutional level we see the same development: a proliferation of labelling, regulating, monitoring and controlling agencies have appeared. The niche market is slowly turning into a conventional market, deeply affecting national, international and transnational power structures among the Fair trade actors.

In a recent study for the French Platform for Fair Trade over 77 evaluation studies carried out between 1998 and 2009 were carefully examined. "Through this work, the French Platform for Fair Trade (PFCE) wanted a comprehensive overview of existing studies on the impacts of Fair Trade, *i.e.* one that presented some straightforward results, without hiding the many methodological issues encountered nor the great richness and variety of topics covered by the studies" (Vagneron and Roquigny, 2010, pp. 2). The most positive impact outcomes were the high differential prices compared to conventional trade and the higher and more secure income earned due to Fair Trade. In contrast, the studies showed very little knowledge about the impact of Fair Trade on the most vulnerable groups such as women and wage workers. Additionally they point out that the increase of inequality within local social structures remains unrecognised and under-theorised as an unexpected side-effect:

“the impact that fair trade may sometimes have on inequality: by supporting a group of beneficiaries, fair trade may sometimes become a source of conflicts, or exacerbate existing inequalities at the micro-level (between men and women, between farmers and wage workers, between temporary and permanent workers, etc.), the community level (horizontal inequalities), or within a value chain (vertical relationship)” (pp. 3).

Fair Trade offers the unique selling point of local economic development and initially the main objective of the Fair trade movement focussed on providing a better income for small and marginalised producers. For many years this differentiated approach allowed Fair Trade commodities to compete with heavily subsidized commodities produced in industrialised countries. However, recognising the current globalisation trend, Fair Trade has become an increasingly lucrative business. The neo-liberal agenda puts power and profits in the hands of Northern businesses rather than national governments or local producers, bypassing many essential conditions and ambitions for development. Due to this trend some Fair trade organisations wish to remain unique in the market and add value to the development agenda, such an organisation is Oxfam.

#### **1.4.1 The case of Oxfam**

Oxfam, a confederation of 17 organisations, runs fair trade projects as part of its sustainable development programme. Although not all 17 affiliates carry out fair trade activities, in March 2011, the Executive Directors of all Oxfam affiliates felt the need to have a thorough discussion around the fair trade activities. “They acknowledged the concerns of staff from fair trade Oxfams who felt that they had not been heard in recent discussions about the role of fair trade in Oxfam’s overall development strategy, including campaigns” (Tack and van der Veer, 2011).

As a response to these concerns the Oxfam Fair Trade Strategy Task Group developed a strategy paper, in which Oxfam commits to fair trade and the fact that it forms part of Oxfam’s aim to fight for economic justice.

There are three reasons why Oxfam wishes to support fair trade:

- 1) Fair trade is a sustainable development tool;
- 2) Fair trade is a valuable business practice; and
- 3) Fair trade is a movement of active citizens.

Oxfam recognises the need to re-integrate its fair trade section with its development, private sector and campaign programmes. The strategy paper communicates Oxfam’s cohesive vision

concerning fair trade and sustainable development and defines the way forward for re-incorporating the Fair Trade Section into the Oxfam programmatic structure.

During a strategic session in May 2011, in which all Oxfam affiliates with fair trade activities took part, a strengths, weaknesses, threats and opportunities (SWOT) analysis was developed. One of Oxfam's weaknesses refers to the lack of a conceptual framework for Fair Trade as a development tool with an agreed set of Oxfam criteria (i.e. empirical data, sound and solid research on impact on livelihoods of producers linked to wages/income, freedom of association etc.). Currently this framework is being developed by the Oxfam International Fair Trade Working Group (OIFTWG). Furthermore, Oxfam would like to have more in-depth knowledge about and better control over the expected contribution of its Fair Trade programmes to local sustainable development. The key question emerges how to evaluate the impact of Fair Trade on small scale producers of food and art crafts, their communities and environment. As a result a monitoring, evaluation and learning (MEL) group was established examining key questions such as in what areas Oxfam wants to make a difference and how Oxfam can make changes happen.

This paper refers to the specific case of Oxfam because Oxfam's ambition to integrate its Fair Trade programmes in its development programmes in the South motivated the author to study the link between Fair Trade and sustainable development. The author believes that the case of Oxfam is recognisable and maybe even representative for many other organisations with a sustainable development agenda. Oxfam was therefore approached by the author to take an active role in certain sections of this research; the Oxfam fair trade MEL group provided support and consultation during the literature study and the selection of organisations to interview. Furthermore, two Oxfam affiliates were assigned to be interviewed on their evaluation practices.

### **1.5 Aims of the research**

First of all, this thesis aims to verify and review if and how a selection of organisations evaluates sustainable development objectives given the complex environment in which they enrol their programmes. The specific aim is to learn how the concept of complexity is interpreted and valued in literature as well as in practice and to find out if and how organisations deal with multi-factorial, external and unpredictable programme circumstances in their evaluation approach to sustainable development.

Secondly, there is a special interest in the sustainability aspect embedded in current development programmes. According to many sources in literature, a sustainable approach demands a holistic and long-term vision on development which in practice must be challenging

considering the pressure to promptly deliver legitimate policies and accountability towards donors and taxpayers or, for nongovernmental organisations (NGOs) specifically, the public. Especially in times of financial crisis accountability and legitimacy becomes more urgent. This research therefore also wants to demonstrate how organisations characterise and evaluate the sustainability aspects within their programmes in the South, given the tension between desired long-term impacts versus imposed short-term performance requirements.

Finally, this research aims to give Oxfam's fair trade MEL working group insights into and understanding of evaluation approaches and practices that can contribute to measuring the sustainable development component of Oxfam's Fair Trade programmes. Oxfam, currently in the process of positioning and integrating its Fair Trade activities in its sustainable development work in the South, will serve as a case study to learn more about views and practices on evaluations for sustainable development as well as Fair Trade programmes.

Please note that this study will focus on evaluation approaches to measure the impact of Fair Trade programmes in the South, delivered through local partners. Outside the scope of this thesis is Fair trade work in the North, meaning:

- Outlets that sell Fair Trade products or commercial partnerships to import Fair Trade goods; and
- Lobby, advocacy, mobilisation and campaign activities.

To achieve the objectives, the following questions guide the research:

#### Research question 1

*In literature, which major evaluation approaches are considered suitable to measure results in a complex environment?*

This section of the thesis research will contain an overview from contemporary literature, exploring existing evaluation theories and methodologies and summarising the findings and opinions from of a variety of academics covering different social disciplines such as health, education, development, etc. The study will start with a broad overview and will consequently work toward a review of evaluation approaches that are more or less, suitable to deal with modern complexity.

#### Research question 2

*How can the evaluation approaches of a selection of organisations promoting Fair Trade and/or sustainable development be characterised?*

The answer(s) to this question will provide insights in the evaluating approaches the interviewed organisations use in order to measure the outcomes of their sustainable development programmes in the South.

*Sub-question1: Do the selected organisations take certain components of complexity into account in their evaluation approach? If so, how do they evaluate them?*

In order to better understand what complexity entails in theory and in practice, a literature study will be performed and a selection of organisations will be requested to describe what certain components of complexity means to them.

*Sub-question2: Do the selected organisations reflect certain components of sustainable development in their evaluation approach? If so, which ones and how do they evaluate them?*

In order to better understand what sustainable development is about, a literature study to conceptualise the construct will be performed and a selection of organisations will be requested to describe what sustainable development means to them.

This paper will bring to light the underlying assumptions, implicit and/or explicit, that organisations use for their programme design and evaluation theory. There are many institutions actively monitoring and evaluating the impact of their programmes and policies as well as governments, which perform evaluations on their development policies. Nevertheless, as this thesis attempts to stay close to the case of Oxfam, the research will focus on the practices of NGOs. Most NGOs currently evaluate their programme objectives. Measurable impact is often a funding requirement and a way to learn from best practices and adapt programmes to obtain better outcomes. This component of the research will illustrate in-depth the current practices and experiences among NGOs with impact evaluation. The organisations that will be interviewed are selected based on criteria previously established by the student with input from Oxfam fair trade MEL group.

## **1.6 Ethical considerations**

For the purpose of this research four organisations have been interviewed. These organisations will be exposed to “the risk of apparent breaches in good ethical practice” (Green and Thorogood, 2009, pp. 89). The author abides by the following ethical consideration: prevention from harm and damage.

This paper seeks to provide an in-depth analysis of evaluation approaches covered in literature and compare those results with modern practices from organisations. In order to prevent any harm or damage to the cooperating organisations, the author realises what implications the final results carry, regardless if they are true or not. “The potential for damage caused by the publication of research findings is not restricted to effects on what is publicly known or on the reputations of people or organisations. Also relevant is the use that may be made of the information” (Hammersley and Atkinson, 2007, pp. 216).

The purpose of this paper is to inform and has therefore no intention to deliver a value judgement on the way organisations are currently performing their evaluations. This paper does not aim to do any harm or damage to any of the interviewees, it respects the organisations’ integrity and autonomy. The author confirms that the interviewees are explicitly informed about the aforementioned intentions.

There is no intention to publish this paper for the public. However, i) the author will share the results of this research with university staff and fellow students; and ii) the author will share the highlights and conclusions of this research with the participating organisations. In both cases the author will present the interview findings in a consolidated and anonymous format, i.e. the interviewed organisations are mentioned in the paper but the actual data derived from the interviews will be presented unnamed. The author will verify that the interviewees explicitly confirm their approval of the aforementioned procedure.

## **1.7 Structure of the report**

In chapter one the author will explain the current issues that sustainable development organisations face in demonstrating the impact of their programme interventions. Fair Trade organisations with a development agenda in the South want to be able to demonstrate the impact of their activities on the lives of small scale producers as Fair Trade has been high jacked by conventional trade actors, who try to benefit from the enormous popularity of Fair Trade products with consumers. The question is raised if and how organisations that want to contribute to sustainable development evaluate their impact? How do they deal with modern

complexity and how do they interpret sustainability? The research will propose to find answers in contemporary literature as well as from (Fair Trade) organisations with sustainable development ambitions. As a case study, the Fair Trade and sustainable development activities and challenges from Oxfam are introduced.

Chapter two shows the results of the literature review. It will become clear that the academic evaluation community is in disagreement about the extent to which organisations should apply evaluations. Some argue that evaluations to establish or measure impact is sufficient; other argue that organisations need to be more intelligent and understand why certain interventions worked and for whom. Complexity is only a challenge for those academics that promote a more comprehensive approach to evaluation; they agree that evaluation in a complex environment can best be dealt with by using a theory-driven approach to evaluation.

As mentioned before, the concept of sustainable development is not entirely clear. In chapter three the author will sum up a variety of definitions as found in literature from specialised institutions such as the European Union and the United Nations. In addition the author categorises the abundant interpretations in thick and thin interpretations.

The research methodology in chapter four will introduce the approach which was used to find suitable organisations to interview for the purpose of gathering 'rich' information. The approach is qualitative and the selection method of purposive sampling allowed the author to select two organisations with similar strategy, activities, and objectives as Oxfam. Together with two Oxfam affiliates the author was able to perform a total of four in-depth interviews.

Chapter five gives a detailed overview of the organisation's experiences with and practices in conducting evaluations. Although theory-driven evaluation was promoted by the academic community in order to deal with complexity, in the day-to-day practices from the interviewed organisations the notion of complexity and the use of theory for the purpose of evaluation are not in a developed stage yet. Ambitions to change their evaluation approaches are high though. The concept of sustainability is in a more advanced stage, it must be noted however that sustainability has many aspects and embedding sustainability objectives in programme evaluation seems less straight forward.

In the second last chapter, chapter six, an overview and analysis are provided about the outcomes of the literature study and the interviews. Remarkably, the interviewed organisations struggle to overcome issues that are not on the scope of the academic community. These issues

have a rather internal focus unlike complexity, which has an external focus. The last chapter, seven, then continues with the conclusions. The interviewed organisations are all in the process of developing new or improving existing evaluation approaches. Together with donors they need to re-think their role in and contribution to the 'bigger system' that eventually will lead to long-term change, e.g. poverty reduction, which so many actors in the system pursue.

## **2 Literature study on evaluation approaches and complex environments**

This chapter provides insight into academic literature on evaluation approaches. Additionally, different aspects of the concept 'complexity' will be listed. In literature, complexity is often mentioned as one of the major challenges in the success of contemporary programme evaluation.

Along with the trend of growing rationalisation and accountability in modern society, evaluation practices became more commonly used since the 60's (Chen, 1990). The increased responsibility and contribution of states and non-profit organisations in systematically introducing new or improved social programmes created the need to know "whether or not public services and improvement efforts were succeeding" (Stufflebeam, 2001). Parallel to the practices, evaluation theories became a prominent subject for academic research and debate. "Evaluation theories describe and prescribe what evaluators should do when conducting evaluations. They define among others the role and responsibilities of the evaluator, the participants as well as the users, the use or purpose of the evaluation, and the choice for methodology" (Coryn et al., 2010, pp. 199). The information and conclusions derived from evaluation research should finally lead to more effective policies and programmes (Weiss, 1997; Pawson and Tilley, 2004). Remarkably it is the lack of substantial impact from evaluation research that has been subject of debate in the last few decades.

When reading literature on evaluation approaches it becomes clear that in general authors distinguish theory-driven/theory-based evaluation from traditional/conventional/rational or mainstream evaluation. The modern generation of programme evaluators embraces the formulation and testing of underlying programme theory whereas the conventional generation focuses on programme effects that can be observed by means of experiments, pilot or case studies, and real-time evaluations. All authors, read for the purpose of composing the theoretical framework of this paper, agree on one point: evaluation practices do not have the desired impact. All reviewed arguments, whether they are in favour of the conventional evaluation approach or the theory-driven approach, indicate that the outcomes of evaluation research are underutilised.

## 2.1 Conventional evaluation approaches

The conventional approach to evaluation can also comprise of qualitative information, but the primary focus will be on the collection of quantitative, measurable evidence. Commonly referred to as a ‘positivist approach’, it advocates for acquiring objective and value-free knowledge, based on evidence and obtained from quantitative research (Sanderson, 2000). To acquire evidence (empirical data) the evaluation is performed by means of an experiment or test. The programme intervention, for example a new drug, is given to a randomly selected group of people (experimental group). The results shown by the people within the experimental group are compared to a similar group of people that did not receive the intervention (the control group). These types of evaluations have been used successfully for years by biomedical scientists in finding new ways to cure or treat diseases, for example.

Undisputed features of conventional evaluations are that they are not based on programme theory; they assume the relationship between causes and effects is linear (see figure 1), and they are developed on a static chain of programme events: input, output, outcome and impact (Hospes, 2008). According to Sanderson (2000) what he calls the ‘rational model’ of evaluation is based on two major assumptions, namely “we know how to change social systems” and “we know what we want to achieve” (pp. 441).

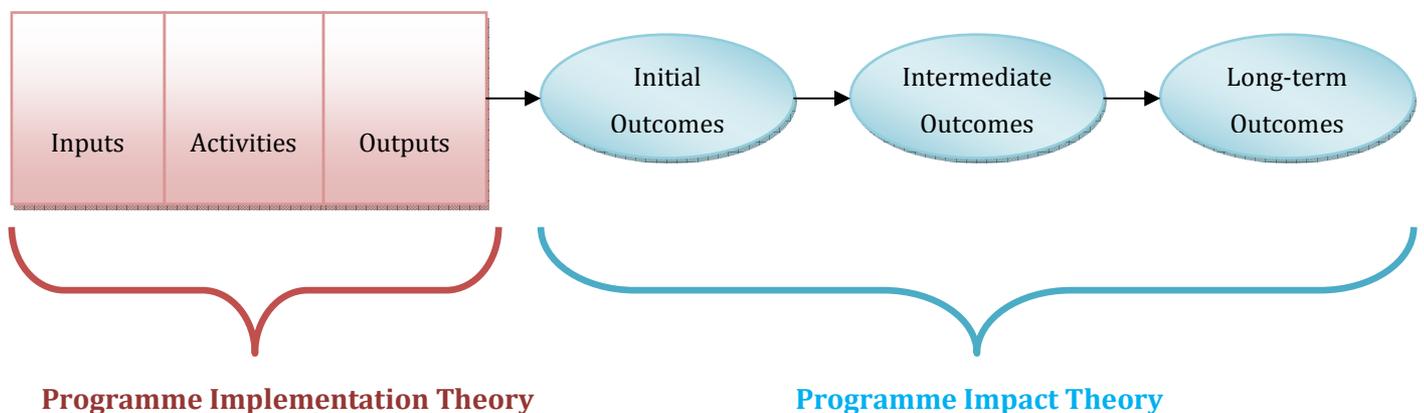


Figure 1 – Adapted Linear Program Theory Model, by Coryn et al (2010, pp. 201)

### Box 1:

In the linear model, see figure 1, there is a direct and traceable link between the inputs, activities, outputs, and outcomes. It must be said that the above model is a generalised example and that alternative models circulate in literature. Inputs in general refer to the financial, human and material resources used for the intervention (OECD, 2002). “Activities are the actions (e.g. training and service delivery) undertaken to bring about a desired end” (Coryn et al., 2010, pp. 202). A programme activity is also often referred to as an intervention (OECD, 2002). Outputs are the number of product or service units provided by the

programme, such as the number of parent education classes or the number of client contact hours (Rand Cooperation, 2004). Outcomes then are the anticipated changes as a result of these inputs, activities and outputs. Initial outcomes are usually expressed as changes in individual knowledge, skills, abilities, which may thereafter lead to behavioural changes (intermediate outcomes), which may lead to “long-term outcomes such as alleviation, reduction or prevention of specific social problems or meeting the needs of the programme’s target population” (Coryn et al., 2010, pp. 202).

The conventional evaluation approach is also referred to as method-driven; the evaluation theory is developed around a certain method which will often depend on the purpose (summative, formative, and developmental), data (quantitative, qualitative), design (naturalistic, experimental) and focus of the evaluation (process, outcomes, costs, impacts, etc), all with their specific advantages and disadvantages (Patton, 2000).

Although this type of evaluation may inform evaluators and programme designers about the results of the programme input, i.e. ‘what works’, it fails to analyse what works for whom under which circumstances (Pawson, 2005). According to Chen (2011), it therefore provides little information for programme improvement or the ability to generalise. Due to this deficit in policy and programme evaluation (Blamey and Mackenzie, 2007), new evaluation methods emerged that can be categorised under theory-based evaluation approaches.

## **2.2 Theory-based evaluation approaches**

Programme theory evaluation, theory-driven evaluation, theory-based evaluation, and theory-oriented evaluation are all references in literature for evaluations that are based on programme theory. In this paper the author will use ‘theory-driven evaluation’ but it may well be that in certain quotes other terms are used.

Although occasional earlier research on programme theories can be found in literature, “it was not until 1990 that theory-driven evaluation resonated more widely in the evaluation community with the publication of Chen’s seminal book *Theory-Driven Evaluations*” (Coryn et al, 2010, pp. 200). Since then a wide-scale movement in research on and practices in theory-driven programme evaluations unfolded. Chen states that “program stakeholders’ implicit and explicit assumptions on what actions are required to solve a problem and why the problem will respond to the actions play a central role in formulating the program theory” (Chen, 2011, pp. 6).

According to Coryn et al (2010, pp. 203), “at its core, theory-driven evaluation has two vital components. The first is conceptual, the second empirical. Conceptually, theory-driven

evaluations should explicate (unfold, elaborate) a program theory or model. Empirically, theory-driven evaluations seek to investigate how programs cause intended or observed outcomes”. In other words, theory-driven evaluation seeks to prove whether a programme has been effective or efficacious (i.e. causal description, proving **that** a causal relationship exists between A and B) and to explain the underlying causal mechanisms (i.e. causal explanation, showing **how** A causes B). “Theory-driven evaluation is any evaluation strategy or approach that explicitly integrates and uses stakeholder, social science, some combination of, or other types of theories in conceptualising, designing, conducting, interpreting, and applying an evaluation” (Coryn et al., 2010, pp. 201).

The key characteristic of theory-driven evaluation is that it goes beyond establishing whether programmes work; it also seeks an answer to how they work and why they fail (Coryn et al., 2010). “The aim (of programme evaluation) is to examine the extent to which program theories hold. The evaluation should show which of the assumptions underlying the programme break down, where they break down, and which of the several theories underlying the program are best supported by the evidence” (Weiss, 1995, pp. 67). Further, the majority of proponents of theory-driven evaluation believe that the contextual factors and complex environment in which the programme operates can influence the programme’s process and disrupt the prediction and attribution of outcomes as is suggested in the linear system. In a non-linear programme model there is no direct link from programme inputs to programme outcomes.

## ACTION MODEL (prescriptive assumptions)

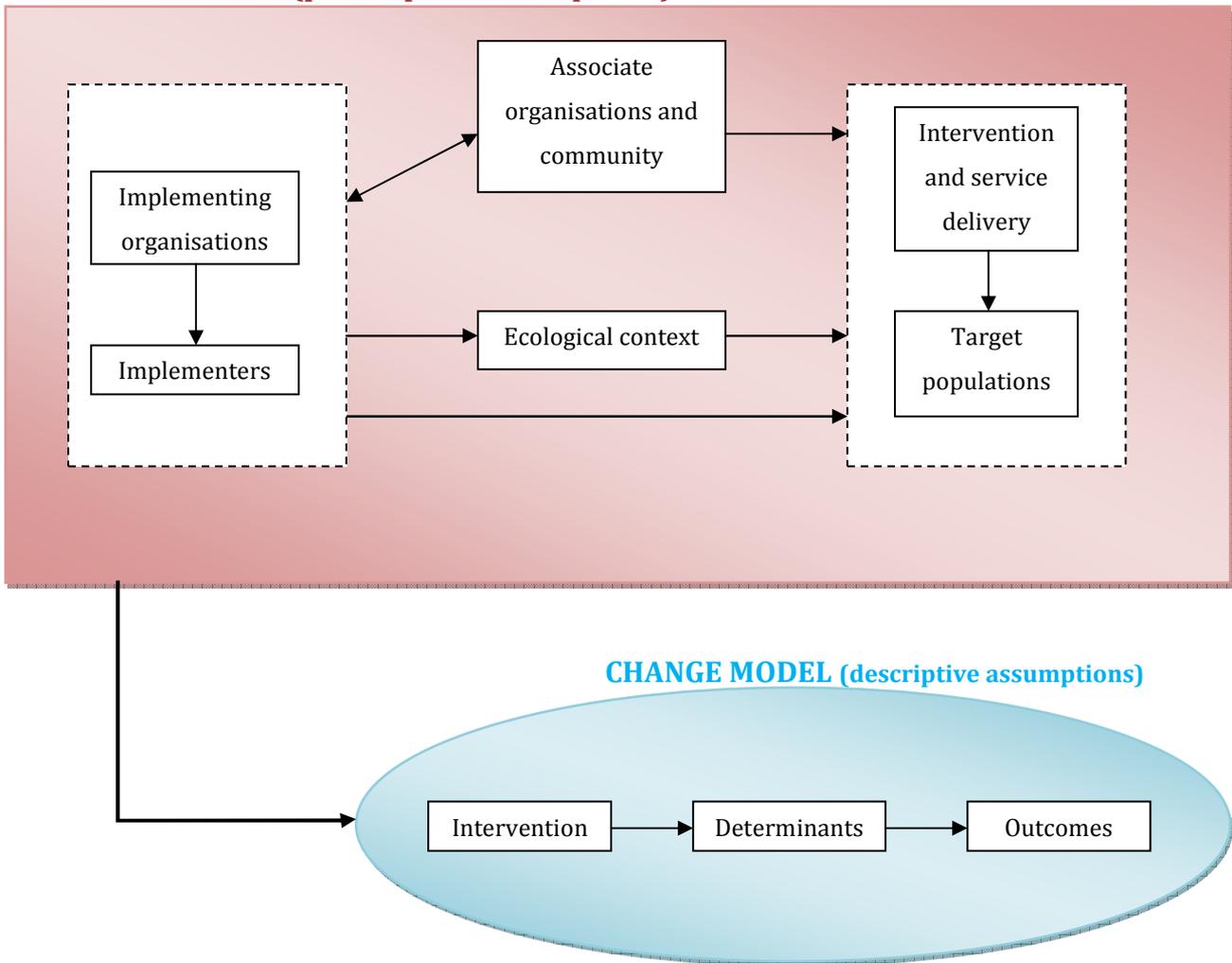


Figure 2 – Conceptual Framework of Program Theory by H. T. Chen (2005, pp. 31)

### Box 2:

Although figure 2 is just one of many 'simplified' models reflecting a non-linear system of change, it clearly illustrates a more complex nature of programmes and evaluations than the conventional linear approach does. Chen's model, for example, takes into account the determinants which refer to the 'mechanisms' later in this paper explained in the dominant theory-driven evaluation approaches by Weiss and Pawson and Tilley. Further the illustration shows that the action model, i.e. the implementation of the programme embeds its own dynamics and interactions between different actors and institutions. In reality complexity of dynamics, interaction, environment etc. within the context of social programming may go far beyond what is pictured by Chen.

According to Chen programme theory serves as a conceptual framework for evaluating effectiveness. He claims that theory-driven evaluation is method-neutral and promotes using multiple methods, a combination of both qualitative and quantitative. Chen calls the evaluation

of programme theory a 'holistic' exercise in which contextual factors and causal mechanisms are taken into consideration. The role of the evaluator comes down to facilitating stakeholders to clarify the contextual factors and mechanisms essential for what they believe will make the programme successful (Chen, 2011).

By performing a theory-driven programme evaluation, programme failures can come to light and be correctly attributed to underlying theories or assumptions that were incorrect. Weiss (1997) argues that to discover these incorrect theories is as valuable as identifying good theories.

### **2.2.1 Programme theory deconstructed**

Supporters of the theory-based approach promote the use of information acquired by both quantitative and qualitative research methods as well as the use of multiple and mixed methods. On the other hand, academics that defend the conventional approach seem less eager to include theory. Stufflebeam (2001) for example is not in favour of using theory as the base for programme evaluation. He states: "there is not much to recommend theory-based program evaluation, since doing it right is usually not feasible and since failed or misrepresented attempts can be highly counterproductive" (2001:39). Stufflebeam is sceptical about the competencies of the organisation to develop sound theory for their programmes as this "entails sound conceptualization of at least a context-dependent theory, formulation and rigorous testing of hypotheses derived from the theory, development of guidelines for practical implementation of the theory based on extensive field trials, and independent assessment of theory". Michael Scriven wrote an article entitled 'Minimalist Theory: The Least Theory That Practice Requires', in which he points out that defining programme theory would take time and money which is not needed to perform useful evaluation (Scriven, 1998).

The ongoing debate in the evaluation community unfortunately remains quite abstract as many writers about theory-driven evaluation focus on *what* to do but fail to explain *how* to do it (Weiss, 1995; Coryn et al., 2010).

A possible explanation why there is less interest to use programme theory in evaluations and work with measurable evidence only, may be found in the way programme theory is interpreted by different academics and their lack of consistency on how different theories are described (Blamey and Mackenzie, 2007). If one scrutinises literature to find out more about the meaning and application of programme theory within the context of evaluation, there seems to exist many different interpretations. Some scholars challenge the theory-driven approach to

evaluation. They state that in practice programme theory is not much more than a programme's logic (Stufflebeam, 2001; Scriven, 1998; Coryn, 2010) most likely in the form of a model or a chart. This 'tool' reflects the expectations of "how inputs are thought to be processed to produce outcome" (Stufflebeam, 2001:38). Pawson and Tilley (2004) claim that in a way all programmes are theory as they are built upon assumptions, values, ideas, experiences, studies etc. Weiss also prefers a straightforward definition as she claims that "the theory in question is the set of beliefs and assumptions that undergird programme activities" (Weiss, 1997, pp. 503). In general, proponents of programme theory prefer to look at it as a hypothesis that needs prior (social science) research and post-evaluation in order to be tested. They agree on the importance of context and complexity surrounding the programme environment. Still, amongst them dissension seems to rule; opinions differ considerably on how programme theory is constructed, how the evaluation should be designed, and who should take part in those processes.

A good example of the mixed use of concepts among academics comes from Coryn et al. (2010, pp. 200). They performed a review of theory-driven evaluation practice covering the years 1990 to 2009. They write:

*...theory-driven evaluation is sometimes referred to as program-theory evaluation, theory-based evaluation, theory-guided evaluation, theory-of-action, theory-of-change, program logic, logical frameworks, outcomes hierarchies, realist or realistic evaluation (Mark, Henry, & Julnes, 1998; Pawson & Tilley, 1997), and, more recently, program theory-driven evaluation science (Donaldson, 2007), among many other (Rogers, 2000, 2008; Rogers et al., 2000; Stame, 2004) (2009:200).*

Birckmayer and Weiss (2000, pp: 409) outline the same dilemma:

*The theories in theory-based evaluations can come from various sources. Some writers suggest that they should be social science theories with a responsible scientific pedigree (Chen and Rossi, 1992). Other authors try simply to set forth the logic that must be involved if a program is going to work, and they may even name the program theory a "logic model" (Coffman 1999; McLaughlin and Jordan 1999).*

In their review of theory-driven evaluation practice, Coryn et al. (2010) notice that many authors distinguish between theories about a programme's 'process' and the theories about a programme's 'impact'. Weiss refers to programme theory as the combination of 'implementation theory' and 'programme theory' and she states that many authors of theory-driven evaluation

mix up these different kinds of theory (Weiss, 1997). Before programmes can have any effects they must be implemented, Weiss argues. Then, programme theory is based on the mechanisms that make things happen (Stame, 2004). These “mechanisms of change are not programme activities per se, but the response that the activities generate” (Weiss, 1997). Scriven (1998) also recognises that two types of theory exist; “‘internal theories’, which explain how an entity manages to produce its output and ‘external theory’ (also known as intervention theory) which explains how its outputs produce the effects they are used to produce, which we often refer to as outcomes by contrast with outputs” (pp. 60). Finally Chen (1990) states that programme theory consists of two parts: classifying the theories into ‘prescriptive theory’ and ‘descriptive theory’. Prescriptive theory deals with what the structure of a programme *should* be, including treatments, outcomes and implementation processes, and is also referred to by Chen as ‘normative theory’. The descriptive theory deals with what *are* the underlying causal mechanisms that link the relationships among programme treatment implementation processes and outcomes. This theory is also called ‘causative theory’ by Chen (1990). Chen shows with his non-linear programme theory model (refer to figure 2) an inclusive picture of links and mechanisms that create dependencies and dynamics between the different programme activities and outcomes. Chen’s *prescriptive* assumptions or theories, also referred to as action model, are similar to what was established as implementation theory by Weiss, process theory by Coryn and internal theories by Scriven. The same goes for Chen’s *descriptive* assumptions or theories, also referred to as the change model; they match what was previously defined as programme theory by Weiss, impact theory by Coryn and external theories by Scriven (Chen, 1990; Coryn, 2010; Scriven, 1998; Weiss, 1997). Chen comes back to two main questions that explain the rationale behind his model: First he mentions the ‘why question’: why does the intervention affect the outcomes? (change model); consequently he refers to the ‘how question’: how are the contextual factors and programme activities organised for implementing the intervention and supporting the change process? (action model) (Chen, 2011).

Although many terms circulate, two things are clear from abovementioned: i) many academics agree that programme theory consists of two elements; and ii) their definitions show a resemblance in referring to the action of implementation, meaning the operational and internal processes of programme execution, on the one hand and the causal mechanisms that lead to the programme impact on the other hand. For the remainder of this thesis the terms implementation theory for the first and impact theory for the second will be applied.

Most scholars recognise there is a difference in programme theories; however, they do not necessarily agree on their usefulness or even legitimacy. According to Weiss both

implementation and programme impact theory are useful for evaluation provided that their strengths are recognised. “Pursuing implementation theory is useful when the purpose is ongoing feedback to program staff about how the intervention is operating. It is what is required to translate objectives into ongoing service delivery and programme operation. When the purpose is understanding how program effects are (or are not) being realized, the evaluation has to follow the logic of program (impact) theory” (Weiss, 1997, pp. 506). Scriven (1998) on the other hand argues that it is possible to understand an external theory without knowing the internal theory. Thus, the internal theories are not genuine theories; they are rather matters of common sense, therefore often represented in a programme’s ‘logic framework’. As Scriven (1998) metaphorically compares: “Drivers can find out how well cars run without knowing anything about auto mechanics; and even so, they can tell what is wrong in a hundred cases without technical knowledge – the wheel fell off, the tire punctured, the lights went out, the gas run out...The primary task of an evaluator is to determine how well the car runs, and for that you don’t need the auto mechanics” (pp. 62).

### **2.2.2 Why theory-driven evaluation?**

The rationale for academics to advocate for evaluations driven by programme theory has its roots in the complexity of contemporary policy initiatives and the belief that no evaluation in social science can be value-free (Green and Thorogood, 2004; Sanderson, 2002). “Unlike atoms, human beings make sense in their place of the world, have views about researchers who are studying them, and behave in ways that are not determined in law-like ways”(Green and Thorogood, 2004: 13). Moreover, the current implementation of programmes aim for multi-level change. Individuals, families, communities and systems all have their role and responsibilities in engaging in the programme in order to achieve the desired change (Sanderson, 2000; Barnes, 2003). “The task of linking outcomes to action is much more complex than in the case of uni-dimensional interventions such as a new drug to treat breast cancer and hence an evaluation approach is called for that is capable of embracing these multi-level impacts” (Barnes, 2003, pp.266). Additionally, context is also considered by the theory-based approaches.

Weiss (1995) brings four main arguments forward as to why we should base our evaluation on programme theory. Her first argument is that it will concentrate evaluation attention and resources on key aspects of the programme. Secondly, she states that it facilitates aggregation of evaluation results into a broader base of theoretical and programme knowledge. Thirdly, she brings forward the need and advantages of programme practitioners to make their assumptions explicit and to reach consensus with colleagues about what they want to do and why. Finally, Weiss reckons that evaluations that refer to underlying theories and assumptions may have

more influence on policy and public opinion (Weiss, 1995). The recognition that contextual factors influence the implementation and outcome(s) of a programme is also an important argument for the promoters of theory-driven evaluation (Weiss, 1995; Weiss, 1997; Pawson and Tilley, 2004; Stame, 2004, Hospes 2008). Blamey and Mackenzie (2007) even take this argument a step further and argue that context itself is multifaceted and shaped by political, social, organisational, and individual dimensions. Flattening out these variations in context by performing a traditional trial or experiment type of evaluation would eliminate a key ingredient in the mix of factors determining the progress of a project. “In addition, for many such interventions, context is not simply an interesting backdrop but it is instead explicitly targeted for change” (Gambone in Blamey and Mackenzie, 2007, pp. 441)

### **2.2.3 Sources of theory**

Donaldson (in Coryn et al., 2010) identifies four *sources* of programme theory. He claims that programme theory can be based on 1) prior theory and research, 2) implicit theories of the ones close to the programme, 3) observations of the programme in operation, and 4) exploratory research to test critical assumptions in regard to a presumed programme theory. Weiss (1995, pp. 70) adds that “if good knowledge is already available on a particular point, then we can change its label from ‘hypothesis’ or ‘assumption’ to something closer to ‘fact’, and move along. However, where a central tenet of the program is still in doubt, or in contention, then it might represent a question for which evaluation is well suited”.

It is important to keep in mind the umbrella discussion about the social development paradigm, i.e. the framework containing all of the commonly accepted views about social development. This includes that the development paradigm used by western countries is conservative and not suitable to include modern complexity and contextual symptoms. Moreover, theories are not just formed by people’s views close to the programme: these views are shaped under the influence of people working at a different levels (e.g. donors and leading institutions such as United Nations Development Programme or UNPD) as well as coming from a different context than the programme or its objectives are supposed to enrol in (e.g. religion, culture, customs and tradition). Programmes and their underlying theories may be often designed because we are stuck to a certain development model that fits the dominant paradigm and we see no alternative or find no base of support to change this paradigm. To tackle those differences the theory-driven evaluation approach promotes the implication of local stakeholders in the programme theory design.

## **2.3 Complexity**

Almost all authors in favour of using programme theory share the opinion that theory-driven evaluation is especially an appropriate approach for programmes that are complex in nature (Barnes et al., 2003; Coryn et al., 2010; Sanderson 2000; Hospes, 2008). As discussed in the previous chapter, due to the complex environment surrounding development programmes the link between causes and effects is not necessarily clear and predictable, making the attribution of established development outcomes to a certain programme or actor a challenging exercise.

So what exactly makes programmes complex?

### **2.3.1 Context dependency of outcomes**

Programmes evolve in their own and very specific context. “This context is dynamic, complex and multi-layered, with many agencies and networks involved” (Hospes, 2008, pp. 2). The context dependency of how a programme develops makes the evaluation findings often hard to generalise to other context settings.

### **2.3.2 Many actors and sectors**

Related to the context are the so called ‘wicked issues’: those social problems that demand interventions from a range of bodies and agencies, and which often seem intractable (Barnes et al., 2003, pp. 207). Weiss (1995) refers to the term ‘community-based cross-systems reform’ indicating that modern reform deals with “multiple factors at the same time: education, training, child care, health care, housing, job creation, community building and so on to increase the desired effects” (pp. 69). Francis (2001) refers to the aim of sustainable development, with the objective to reach environmental, economic and social ‘harmony’. He calls for the use of integrated impact assessments and appraisals.

Barnes et al. (2003) confirm that there is “now widespread acceptance that exclusively experimental models are inappropriate for the evaluation of complex policy initiatives that seek multi-level change within individuals, families, communities and systems” (pp. 265). Based upon their research into evaluation of non-linear systems (in this case an integrated Health Initiative launched by the British government to create healthier communities), they conclude that the complexity of such programmes cannot be captured within one overarching theory.

“Involvement of stakeholders in theory generation instils ownership among stakeholders that extends beyond action to evaluation” (Sullivan and Stewart, 2006, pp. 180). According to the Aspen Institute, in a publication of 1997, potential stakeholders include: “staff, governors and directors, funders, technical assistance providers, residents, local business owners, civic leaders,

representatives of community-based organisations and public, private and non-profit organisations” (Sullivan and Stewart, 2006, pp. 181). Although the benefits of getting all stakeholders aligned and supportive of the programme objectives and activities may result in a clear picture of common beliefs and leaps of faith in a programme, the exercise to get there may be a lengthy and challenging one. Although Weiss says that some programme developers with whom she worked found this exercise as valuable a contribution to their thinking as the results of their actual evaluation, she also admits that it is almost impossible to spell out all the theories of change across the board of stakeholders. “They differ among themselves in emphasis, managerial structure and priorities” (Weiss, 1995, pp. 73).

### **2.3.3 Long-term change is difficult to monitor and measure**

Measurable improvements of social programmes can take many years to achieve (Barnes et al. 2003). This is contrary to the political agenda of donors, i.e. they seek legitimacy for the policies they develop and want to show the public results while they are still in office (Sanderson, 2000, Barnes et al., 2003). Often social programmes have a limit in time, although sometimes organisations run programmes for many years. In literature there is hardly any analysis on how to evaluate causal attribution over a long period of time, with or without the organisation still running their programmes. One can logically assume that over time, the cause of certain positive effects, for example poverty reduction, become more and more diluted when external actors (e.g. state policies) and the targeted individuals and communities create their own dynamics around the programme or programme results.

### **2.3.4 Non-linear dynamics between activities and outcomes**

Predictability of outcomes is difficult when dealing with complex programmes. Stame (2004, pp. 63) argues that “each aspect that may be examined and dealt with by a programme is multi-faceted”. Sanderson (2000) states that complexity has its roots in ‘chaos theory’. This is the study by natural scientists of non-linear systems in which they discovered that small changes can trigger huge consequences as the relationship between the variables is unstable, if not unknown. According to chaos theory, a non-linear system, upon the introduction of a new element or a change, will break up existing structures and move toward a new form of equilibrium. “Mature systems seem to evolve ‘naturally’ towards the ‘edge of chaos’” (Medd in Barnes et al., 2003, pp. 278).

In literature non-linear systems are also referred to as open or adaptive systems (Barnes et al., 2003; Hospes, 2008). Sanderson (2000) prefers the term ‘dissipative systems’. In this type of system energy and materials are exchanged with the environment of the system. This action

disturbs the equilibrium within the system upon which a new equilibrium will be formed. According to Sanderson, the system is very sensitive to the appropriate configuration of 'initial conditions' and is therefore non-replicable over time.

### **2.3.5 Power of agency vs. social and institutional structures**

Many social programmes seek to change behaviour of individuals. However, the social environment surrounding the individual is rather complex in itself. Rules, informal and formal, norms, values, interests, incentives, also referred to as institutions, all shape the rational choices of the individual, creating opportunities for behaviour change but also discouraging or blocking certain individual choices. The concept of new institutionalism focuses on the macro-social structures and processes on the one hand and the micro level of social agents (individual, households, firms, etc) on the other. Although institutions contribute to the complex environment programmes operate in, according to Sanderson (2000), they remain quite under-theorized in literature.

### **2.3.6 Multi-level governance**

Stame (2004) finds complexity in the fact that governments can hand over the delivery of many services to NGOs. She refers to the principle/agent theory in which the contracting party sets the overall objectives of a policy but the agent fills in the means, often prioritising what is relevant to its own agenda. The multi-level system makes attribution of a certain programme outcomes hard to trace. Donors (often big institutions such as the United Nations or the European Union) agree upon development objectives with states. States agree upon development objectives with their implementing partners. Stame argues that "subsidiarity is not a steady state, but a mutual relationship that needs to be actively implemented if one is to ... contribute to global goals" (pp. 65). However, Stame concludes, each level is self-contained; evaluation take place at different times, and are performed by different professionals with different methods (pp.66).

## **2.4 Dominant evaluation approaches dealing with complexity**

As mentioned before, almost all authors in favour of using programme theory agree that theory-driven evaluation is an ideally suitable approach for programmes that are implemented in a complex environment. Both 'theories of change' and 'realistic evaluation' are popular theory-driven evaluation approaches discussed in literature. Both approaches claim to be the answer to the shortcomings of conventional evaluation approaches that fail to deliver information and explanation about why certain programmes work and others do not (Pawson et al., 2005). Theories of change and realistic evaluation both attempt to unravel process and programme theories (Weiss, 1997; Blamey and Mackenzie, 2007) but Blamey and Mackenzie suggest that

each approach in reality only demonstrates strong performance in one of these theories. These evaluation approaches also have in common that the evaluation plan is designed based on the theories behind the intervention, implementation as well as programme impact theories. These theories also drive the ultimate selection of evaluation methods (Blamey and Mackenzie, 2007), both embracing the use of qualitative or quantitative methods or a combination of both.

The sections below outline the differences between the two approaches and what makes them unique and useful for which specific evaluation situation.

#### **2.4.1 Theory of change**

As explained before, theory-driven evaluation is often referred to as theory of change. Are they really the same or are they just poorly conceptualised in literature? According to the articles reviewed for the purpose of this paper, the term ‘theories of change’ was first introduced by Carol Weiss in 1995 when she elaborated, in cooperation with a variety of other writers, on new approaches to evaluating community initiatives. At that point in time she did not define its exact meaning but she did so in 1997, in her article entitled “How Can Theory-based Evaluation Make Greater Headway?” wherein she refuted many of the objections regarding the application of theory in the design of programmes and evaluations. In the chapter about the objection of ‘confusion about components of the theory’ she clearly states: “together, implementation theory and program (impact) theory represent the full theory of change” (pp. 506).

The most prominent characteristic of the theories of change evaluation approach is the (pro) active involvement of stakeholders (Blamey and Mackenzie, 2007) and its application to changes at the community level (Stame, 2004). The philosophy behind the theories of change is that change can only successfully happen if the people affected by the change can contribute to the actual process of it. However, the range of stakeholders extends further than to just the people affected; policy makers from local or national government institutions, local businesses, civil society organisations all form part of the process of programme and evaluation design, in order to anchor the foreseen change within both society and institutions. Apart from the ambition to design successful programme models and reveal causal mechanisms that, considering the context, contribute to the desired change, the principles of community engagement, capacity building and ownership seems to be a means on its own of this evaluation approach (Blamey and Mackenzie, 2007).

By defining the theories of change underlying a programme’s rationale the focus is on the implementation theory as well as on the impact theory. This often results in an exercise of

“mapping the nuts and bolts of the programme” (Auspos and Kubisch in Blamey and Mackenzie, 2007, pp. 445).

*A theory of change approach would seek the agreement from all stakeholders that, for example, activities A1, A2, and A3, if properly implemented (and with the ongoing) presence of contextual factors X1, X2 and X3, should lead to outcomes O1, O2 and O3; and if these activities, contextual supports and outcomes all occur more or less as expected, the outcomes will be attributable to the interventions (Connell and Kubisch in Blamey and Mackenzie, 2007, pp. 445).*

Due to the wide scope of evaluation and the many stakeholders involved, one of the most frequent criticisms of the theory of change evaluation approach is that it is a very intensive and complex effort (Weiss, 1997; Stufflebeam, 2001). The consequence of including a wide variety of stakeholders is that the programme evaluation may have to analyse several ‘routes’ of theory (multiple theories) that have been included in the evaluation design by means of negotiation and consensus. Weiss (1997) states that the theories of change evaluation is easy to apply when the programme design is based on an explicit theoretical foundation. Unfortunately this is rarely the case which is at least remarkable as the approach promotes an “explicitly cumulative approach to knowledge generating. This means that learning is believed to accrete slowly within and across evaluations rather than delivering big bang answers” (Blamey and Mackenzie, 2007, pp. 447-448). Apparently the knowledge generated so far by theory of change evaluations is not used or was not suitable to create a sound base of theory for new programmes to continue with. One of the reasons put forward by Weiss (in Blamey and Mackenzie, 2007) for this shortcoming is that the use of programme theory is still at a low stage of development. Another argument put forward by Blamey and Mackenzie (2007) is that there is a lack of evidence about how to deliver change at an individual and community level, especially when it comes to engaging and strengthening communities that are multiply deprived. However, this argument feels like a chicken and egg discussion as the theories of change evaluation approach was originally designed to fill this gap of knowledge. On top of the epistemological arguments there may also be procedural arguments, to be found in the way the theory of change evaluation is applied in practice. Blamey and Mackenzie (2007, pp. 445) for example “suspect that within the applications of the approach to date, evaluation practitioners have been predominantly engaged with explicating implementation theory”. Their presumption is shared by Weiss (1997), who states that “much of what goes under the label of TBE (theory-based evaluation) follows the chain of implementation” (pp. 506), meaning that the results in evaluations deliver knowledge

on the effectiveness of the programme and not about the underlying mechanisms that cause the actual change.

#### **2.4.2 Realistic or realist evaluation**

“The cornerstone of the realist project is a distinctive viewpoint on how intervention brings about change” (Pawson and Tilley, 2004, pp. 3). Promoters of realistic evaluation are of the opinion that programmes *are* theories and that they are embedded in several layers of social systems. They strongly believe that the effects of a programme can only be triggered by the active engagement of individuals (or group of individuals). Moreover, programmes are ‘open systems’ that cannot be regarded isolated from its context. Externalities can have an impact on the development and the implementation of a programme and vice versa, meaning the programme itself can “change the conditions that made it work in the first place” (Pawson and Tilley, 2004).

The realist approach to evaluation is based on the CMO configuration: context (C) and mechanism (M) deliver outcome (O) and about hypothesing and testing such CMO configurations (Pawson et al., 2005). The theory is concerned with the psychological and motivational responses leading to behaviour change (Blamey and Mackenzie, 2007). And “programmes are seen as the opportunities that an agent, situated inside structures and organisations, can choose to take, and the outcomes will depend on how the mechanism that is supposed to be at work will be enacted in a given context” (Stame, 2004, pp. 62). The *process* of how subjects interpret and act upon the intervention stratagem is known as the programme ‘mechanism’ and is as such at the centre of the evaluation. “Mechanisms describe what it is about programmes and interventions that bring about any effects” (Pawson and Tilley, 2004, pp. 6).

Like the theories of change approach, realistic evaluation also recognises the existence of multiple theories. By means of ‘theory-mapping’, a list of possible theories, i.e. a range of existing theories on a certain intervention, will become clear. These theories however may vary from macro, to meso, and micro theories representing a palette of ideas and beliefs in which the evaluator can set priorities and scope (Pawson et al., 2005). The first thing to do then for a realistic evaluator will be to survey these programme theories and decide which should be subject to testing. In this process the evaluator will have discussions with the programme designers and will additionally perform a documentary analysis, for example reviewing previous evaluation studies. Following these actions, the evaluator will search for data, qualitative as well as quantitative, that can match the theories and assumptions that have emerged during the initial stage. These theories and their matching data will form a ‘model’, consisting of a whole

package of CMO configuration hypotheses that will be tested in the next phase to see if it can “explain the complex footprint of outcomes left by the programme” (Pawson and Tilley, 2004, pp. 11). The final stage of the evaluation is the assessment and interpretation of the findings and the conclusive analysis whether the programme theories have worked or not.

An important feature brought forward by the proponents of realistic evaluation is that it “operates at ‘middle range’, using concepts that describe interventions at a level between the big policy ideas and the day-to-day realities of implementation” (Pawson and Tilley, 2004, pp. 18).

According to Pawson et al. (2005), the purpose of the realistic evaluation differs from the theory of change approach. They argue that the purpose of realistic evaluation is to review the same theory in comparative settings “in which it is assumed that programmes only work for certain participants in certain circumstances” (pp. 25). A review will include many studies of similar interventions in different settings and one of the tasks of the evaluator is trying to identify patterns of winner and losers. Theories of change, however, pursue reviewing to prove theory integrity and according to Pawson et al. (2005) this happens in a real-time setting, looking for ‘flows and blockages’ in the theories.

Pawson and Tilley, the founding fathers of the realistic evaluation approach, promote it because it draws its foundation from the methodology of natural science. “Realistic evaluation stresses theory and the scope for generalisation that comes from attention to explanatory theory – generalisation that is critical in moving progressively from one programme to another” (Pawson and Tilley, 2004, pp. 22). They also admit however, that their approach is intellectually enormously challenging. Extensive thinking is required to unravel programme theory, define expected outcome patterns and what sort of proof to look for. Additionally, all theories and outcome configurations need to be tested against existing theories and ideas, sprouting from a variety of sources. Pawson and Tilley say that “none of this is easy. It requires advanced theoretical understanding, abilities to design research and techniques for data analysis” (Pawson and Tilley, 2004, pp. 23).

### **2.4.3 Summary**

Although the above descriptions of dominant theory-driven approaches are not nearly complete, a broad analysis of the main characteristics and ways of working is sufficient to define the most outstanding differences and similarities between the two approaches:

Differences between the two approaches are:

- Theory of change evaluation is driven and owned by programme stakeholders;
- Realistic evaluation is driven and owned by the evaluator;
- Theory of change evaluation aims to define the underlying programme theories by means of negotiation and consensus among a broad base of stakeholders;
- In realistic evaluation the evaluator approaches sources, internal as well as external to the programme, for emerging programme theories;
- Theory of change includes evaluation of implementation theory. The programme activities, i.e. as listed in an action plan agreed upon by the stakeholders, is part of the evaluation plan;
- Realistic evaluation has a focus on impact theory, i.e. the mechanisms of the activities. They make use of implementation theory as well to increase their understanding of the process and rationale for design of the action plan;
- Community engagement, capacity building and ownership is a purpose of theory of change;
- The focus of theory of change evaluators is to test the internal validity of the theory/theories; their intent is not so much to render judgement on the particular initiative as to understand the viability of the theories on which the initiative is based; and
- The focus of realistic evaluators is to test the external validity of a programme, and specifically of those CMO configurations that are successful in a specific context for a specific individual or group of individuals. These configurations will be tested and refined during the evaluation until they are “sufficiently promising that they might be imported to other interventions that aim to change community-level behaviour” (Blamey and Mackenzie, 2007).

Similarities between the two approaches are:

- Both approaches design their evaluations based on programme theory;
- Both approaches recognise and include multiple theories;
- Both approaches evaluate changes within community behaviour, reached through intervention with individuals or groups of individuals;
- Both approaches claim to test causal mechanisms, the responses to the programme activities or intervention that will contribute to the desired change, taking into account contextual circumstances; and
- Both approaches recognise that, if applied correctly, they are time consuming and quite complex exercises.

### 3 Conceptualisation of sustainable development

In the previous chapter a comprehensive overview of factors that add to complexity was presented as well as suitable evaluation approaches that intend to incorporate these factors. Certain components of complexity appear to relate directly to the desired outcomes of sustainable development. Concrete examples are the long-term effects and multi-disciplinary and multi-sectorial dynamics in the area of implementation. However, there is no widely agreed framework for sustainable development or for the evaluation of its impact. In literature (academic as well as technical) different interpretations exist for sustainable development and what it is supposed to accomplish and how.

In this chapter an overview will be provided of the origin of the 'sustainable development' concept and the different interpretations that have evolved since its introduction.

#### 3.1 Our Common Future

The term sustainable development was introduced and extensively elaborated on in the report published by the World Commission on Environment and Development (WCED), in 1987. The report, *Our Common Future*, also known as the Brundtland report, defined sustainable development as: "development which meets the needs of the present without compromising the ability of future generations to meet their own needs" (UN, 1987, art. 27) The report was prepared as requested by the United Nations General Assembly, and the initial concept was solely to focus on "environmental issues". However, the report stated that the environment could not be addressed as an isolated area of policy or separated from development policy as "many critical survival issues are related to uneven development, poverty, and population growth. They all place unprecedented pressures on the planet's lands, waters, forests, and other natural resources, not least in the developing countries. The downward spiral of poverty and environmental degradation is a waste of opportunities and of resources" (UN, 1987, Chairman's Foreword). The WCED, headed by Gro Harlem Brundtland, thus established a clear link between poverty, inequality, and environmental degradation and called on nations for the return to multilateralism and a common approach and cooperation to address "complex problems bearing on our very survival: a warming globe, threats to the Earth's ozone layer, deserts consuming agricultural land" (UN, 1987, Chairman's Foreword). The Commission argued that environmental degradation, as a side effect of industrialisation, affects and threatens developing countries, referring to the "downward spiral of linked ecological and economic decline in which many of the poorest nations are trapped" (UN, 1987, Chairman's Foreword). The Brundtland report was adopted by the General Assembly and, in 1992, leaders set out the principles of

sustainable development at the United Nations Conference on Environment and Development in Rio de Janeiro, Brazil.

Today, 20 years later, the words 'sustainable' and 'development' are inseparable in the development discourse; they go hand-in-hand in the way we perceive and value development and how we want to approach it. Despite its successful integration in the development discourse, the exact meaning of sustainability remains vague (Francis, 2000; Parris and Kates, 2003; IISD, 2010) and "proponents of sustainable development differ in their emphasis on what is to be sustained, what is to be developed, how to link environment and development, and for how long a time" (Parris and Kates, 2003, pp. 560).

Despite the formulation and acceptance of a broad set of sustainable principles adopted in 1992, their implementation has proven difficult and often unsuccessful. According to a report by Drexhage and Murphy (2010) from the International Institution for Sustainable Development (IISD) this is due to the fact that the western paradigm for development still focuses on economic growth and sustainability often is perceived as an environmental issue. Furthermore, this way of thinking has been adopted and continued by those countries that managed to increase their development over the last 10 years (e.g. Brazil, Russia, India, China, South Africa and Mexico, i.e. BRICSAM countries). "The problem with such an approach is that natural resources are in imminent peril of being exhausted or their quality being compromised to an extent that threatens current biodiversity and natural environments" (Drexhage and Murphy, 2010, pp. 2). The authors not only call for a shift in paradigm, they also feel that western consumers should change their consumption pattern.

### **3.2 Thick and thin interpretations**

In reviewing literature on interpretations of sustainable development it becomes clear that some organisations apply only a 'thin interpretation' of what was described in the Brundtland report. Other organisations prefer a 'thick interpretation', including additional development outcomes on top of environmental, economic and social development, as well as recommendations for change to be attained and governed. For example in 1988 the Food and Agriculture Organization (FAO) Council called for technological and institutional change in order to ensure the attainment and continued satisfaction of human needs for present and future generations. Thus sustainable development is expected to (in the agriculture, forestry and fishery sectors) conserve land, water, plant and animal genetic resources, be environmentally non-degrading, technically appropriate, economically viable and socially acceptable (FAO, 2013). Another thick description is utilised by International Institute for Sustainable Development:

*“Despite an on-going debate on the actual meaning (of sustainable development), a few common principles tend to be emphasized. The first is a commitment to equity and fairness, in that priority should be given to the improving the conditions of the world’s poorest and decisions should account for the rights of future generations. The second is a long-term view that emphasizes the precautionary principle, i.e., “where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation” (Rio Declaration on Environment and Development, Principle 15). Third, sustainable development embodies integration, and understanding and acting on the complex interconnections that exist between the environment, economy, and society. This is not a balancing act or a playing of one issue against the other, but recognizing the interdependent nature of these three pillars” (Drexhage and Murphy, 2010, pp. 6).*

The United Nation Programme for Development (2011, pp. 2) builds on the work of Anand and Sen, defining “sustainable human development” as “the expansion of the substantive freedoms of people today while making reasonable efforts to avoid seriously compromising those of future generations.” The most comprehensive definition, and perhaps ambitious perception, of sustainable development comes from the European Council (2006). They state that sustainable development is about

*“safeguarding the earth’s capacity to support life in all its diversity and is based on the principles of democracy, gender equality, solidarity, the rule of law and respect for fundamental rights, including freedom and equal opportunities for all. It aims at the continuous improvement of the quality of life and well-being on Earth for present and future generations. To that end it promotes a dynamic economy with full employment and a high level of education, health protection, social and territorial cohesion and environmental protection in a peaceful and secure world, respecting cultural diversity” (European Council, 2006, pp. 2).*

RAND Corporation, a non-profit research organisation, is a typical example of an organisation that supports a thin interpretation of sustainable development. In their report ‘Getting to Outcomes’ (2004), they define sustainable programmes are about “the continuation of a program after the initial funding has ended. Programs are more likely to survive if they adapt themselves to fit the needs of the environment and the needs of their host organisations...” (Rand Cooperation, 2004, pp. 143). The United Nations Office of Internal Oversight Services (OIOS),

established in July 1994 to assist the Secretary-General in fulfilling his internal oversight responsibilities in respect of the resources and staff of the Organisation through monitoring, internal audit, inspection, evaluation and investigation, also has developed a much thinner interpretation as the basis for its work than initially intended by the UN Rio de Janeiro Declaration. UN OIOS (2006) defines sustainability as “the extent to which the impact of the programme or project will last after its termination; the probability of continued long-term benefits”. The Organisation for Economic Cooperation and Development (2010) sees sustainability as “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” (OECD, 2010, pp. 36).

### **3.3 The world as a system**

Besides the thin and thick interpretations, based on reports and glossaries, an interesting theoretical insight emerges. In the IISD report, Drexhage and Murphy (2010) argue that two major concepts drive development, namely the concept of needs, in particular the essential needs of the world's poor, to which overriding priority should be given; and, the idea of limitations imposed by the state of technology and social organisation on the environment's ability to meet present and future needs. They state that we need to see the world as a system “—a system that connects space; and a system that connects time” (Drexhage and Murphy, 2010, pp. 11). Moreover, Nootboom (2007) introduces a similar definition of systems as a key concept of sustainability:

*“Each system to survive in a larger, ever changing environment, must keep its wholeness as well as its ‘partness’ of a larger system in which it is embedded and on which it depends for survival and to which it gives a contribution and in which it competes for resources (e.g. Wilber, 2001; Teisman, 2005). If a system survives, it may be viewed as sustainable. A possible definition of sustainable development is therefore: ‘a system has a sustainable development if that development enables it to maintain its wholeness as an integral system, whilst also maintaining its role as part of a larger system on which it depends’ (Nootboom, 2007, pp. 646).*

Please refer to chapter 2.3.4 for elaboration on complexity and ‘adaptive systems’.

### **3.4 Complexity of evaluating sustainable development programmes**

Despite the lack of consensus on the interpretation and application of sustainable development, it is accepted in the majority that sustainable development calls for a well-balanced convergence

between economic development, social development and environmental protection (UN, 1987; Francis, 2000; Parris and Kates, 2003; Drexhage and Murphy, 2010; FAO, 2013) and that it aims for the probability of continued long-term benefits and to satisfy the needs of present and future generations (UN, 1992, principle. 3).

Complexity in sustainable development programmes is recognised by some scholars and institutions, but not by all. Some prefer to look at programmes and what they produce in reality, choosing not to unfold long-term effects or multi-disciplinary dynamics in the area of implementation. However, if complexity is recognised, its dimensions are diverse and all proponents agree that complexity adds a significant challenge to the evaluation of sustainable development programmes. It can be expected that the degree of complexity recognised within organisations relates to their interpretation of sustainable development; the thicker their interpretation of sustainable development, the more complex their experience within the programme implementation environment and vice versa.

## 4 Research methodology

The previous chapters illustrated in detail the various aspects, opinions and conceptions that different academics promote of the three main conceptual pillars of this paper: evaluation approaches, complexity and sustainable development. The outcome of this literature review forms the groundwork from which the interview questions for a selection of organisations were derived.

This paper presents a qualitative study, i.e. it seeks answers to the research questions focussing on the 'what', 'how' or 'why', rather than questions about 'how many' or 'how much' (Green and Thorogood, 2009). Within the field of evaluation approaches the data collection of this research will determine what type of evaluation approach the selected organisations utilise. Both conventional and theory-driven approaches will be considered during the interview. Those organisations that indicate use of theory-driven evaluations will be asked to elaborate on its application in order to define if they use implementation theory, impact theory or a combination of both. In order to form an idea of the level of complexity an organisation is dealing with in its evaluation approach, the study will be limited to and focussed on the organisation's capability and capacity to deal with i) context dependency of outcomes, ii) multiple actors and sectors, and iii) long-term outcomes. Finally, when discussing sustainable development, the 'thin' as well as the 'thick' interpretations will be used to frame the organisation's understanding of sustainable development. The category 'thick' contains all definitions provided by organisations that include economic, social and environmental aspects of development and foresee continued and long-term benefits.

In the findings of this paper (chapter six), the results of the literature research and the interviews will be compared. The author analyses the differences and agreements between theorists and practitioners with respect to the possibilities and limitations of evaluation approaches in capturing the dynamics caused by modern complexity.

### 4.1 Data collection and processing

Two different methods for data collection will be applied.

1. Documentation study:
  - a. Scientific literature covering theories of evaluation and evaluation approaches;
  - b. Specialist literature and/or glossaries by leading organisations in sustainable development and/or Fair Trade;

- c. Documents obtained from the interviewed organisations; and
- d. Documents obtained from Oxfam.

2. Oral and semi-structured interviews:

“In-depth and semi-structured interviews give the interviewer flexibility in how to ask questions, but very careful preparation is needed to think about what to ask in order to generate useful information” (Green and Thorogood, 2004: 113). During the interviews the author requested the interviewee, a member of staff responsible for evaluations and/or programme design, to elaborate extensively on the opinions and experiences that are (or have been) crucial for determining the current evaluation practices within the organisation. If needed, the author spent more time on questions that seemed relevant for the organisation and less time on the questions that they did not recognise or had no experience with.

Initially, the interview questions were presented in one main open question and two sub-questions. Each main and sub-question were supplemented by a series of closed questions that could optionally be used by the interviewer depending on the completeness and extensiveness of the answer provided to the main question. The main and sub-questions provided input to answering the research questions. Subsequently, the interviewer also put forward three reflective questions obtain better insight into how the specific evaluation practices and experiences are valued by the different organisations. Appendix A contains the list of interview questions.

In order to prepare for each interview, a decent amount of desk research needed to precede the interview. A good understanding of the organisation’s mission, change goals, programme activities and evaluation practices were required in order to probe the interviewee for additional information. The author developed a ‘methodological protocol’ to structure the required activities for the preparation for and elaboration of the interviews and define the information required to receive from each organisation prior to the interview. Furthermore, each stage contains an estimation of time. Please refer to appendix B for the methodological protocol.

Using both methods, i.e. reviewing documentation and conducting interviews, will allow the author to compose ‘information-rich cases for in-depth study’ and provide the author access not only to what people *say* but also what they *do* (Green and Thorogood, 2009).

The author will interview four organisations and hopes to retrieve as much relevant information as possible by posing open questions. Therefore, the interview questions do not need to be coded. The interview outcomes and documentation will be, per organisation, thoroughly analysed and compiled into a case study in which, by means of a rich narrative, the organisation's evaluation practices and experiences are extensively described. The case studies will be structured following the same line as the interviews.

#### **4.2 Selection of organisations**

This paper aims to provide an answer or technical advice to the question whether, and if so how, it is possible to evaluate the impact of Fair Trade programmes on sustainable development in the South. The case study is performed for the fair trade section of Oxfam and therefore two Oxfam affiliates, Oxfam Wereldwinkels and Intermon Oxfam that run Fair Trade programmes in the South will be interviewed and the remaining interviewees selected on characteristics similar or equal to those of Oxfam's Fair Trade work in the South.

In order to gain a thorough understanding of how organisations evaluate their sustainable development programmes, the author has selected two organisations, in addition to the two Oxfam affiliates, to be interviewed based on similar characteristics. For the selection of these organisations the author used the method of 'purposive sampling', which means: "explicitly selecting interviewees who are likely to generate appropriate and useful data" (Green and Thorogood, 2009, pp. 118). This method is deemed most suitable for obtaining extensive information on the underlying assumptions, implicit and/or explicit, that organisations use for their programme design and evaluation theory. According to Patton the overall aim of purposive, as opposed to probability, sampling is to include 'information rich cases for in-depth study' (Patton quoted in Green and Thorogood, 2009, pp. 118).

In order to apply the method of purposive sampling it needs to be clear what the sampling criteria are in order to search for suitable to interview. As was mentioned in chapter 1.5, the research attempts to work with organisations that have similar programmes and objectives as Oxfam. In order to understand the main characteristics of Oxfam's Fair Trade activities Oxfam's fair trade strategy paper was reviewed. In this paper retailing Fair Trade is defined as "a tool which responds to the failure of conventional trade to deliver sustainable livelihoods and development opportunities to people in the poorest countries of the world" (Tack and van der Veer, 2011, pp. 2). It seeks to increase *economic justice* in Southern developing countries for small scale farmers and crafts people. The main characteristics of Oxfam's Fair Trade activities that are presented in the paper are:

- Fair trade addresses the inequitable distribution of assets, income, opportunities and power for Southern small scale producers, crafts people and workers who are marginalised;
- Fair Trade is an alternative sustainable development tool for marginalised producer groups to develop secure and sustainable livelihoods;
- Fair Trade relationships builds on the following components:
  - Fair prices;
  - Pre-financing;
  - Long-lasting contracting obligations; and
  - Producer capacity building;
- Fair Trade aims to have a maximum impact on poverty while at the same time ensuring economic sustainability;
- Oxfam fair trades with partners. The common Oxfam values and criteria are the basis for partnerships;
- Fair Trade work completed includes:
  - Capacity building of small producer groups;
  - Supporting the establishment, function and credit-raising activities of Southern co-operatives; and
  - Supporting the establishment of Southern networks.

In short: ideally an organisation that qualifies to be interviewed runs programmes aiming for similar or equal objectives as Oxfam Fair Trade programmes, and promotes similar or equal standards and approaches (in the South). Such an organisation preferably should:

1. Reach beneficiaries that are small scale producers of agricultural or other products;
2. Work through local partnerships;
3. Aim for long-lasting commitment with partners;
4. Aim for sustainable development;
5. Offer and implement an alternative economic income model (e.g. trade offering fair prices and pre-financing) in addition to or instead of conventional income activities to develop secure and sustainable livelihoods;
6. Contribute to a more just economic system: promoting rights and increasing access to institutions and markets;
7. Aim for economic sustainability / independence; and
8. Build capacity of partners and beyond at local, regional, national and international level.

Based on above criteria and advice from the Oxfam fair trade MEL group, the following organisations were, next to Oxfam Wereldwinkels and Intermon Oxfam, found suitable for interviewing: Vredeseilanden and Traidcraft.

Below follows a short introduction from each organisation.

#### **4.2.1 Vredeseilanden**

Vredeseilanden is a non-profit organisation active within Sustainable Agricultural Chain Development. Vredeseilanden was established in 1958 by Dominique Pire and in 2001 the organisation obtained its current structure because of a merger with development organisations COOPIBO en FADO. The name Vredeseilanden was already well known with the public and the new structure therefore continued under this name. In 2007, Vredeseilanden made a shift from a food security agenda to one focused on market access for smallholder family farmers through inclusive value chain development. They strive for a world in which family farmers claim their rights to create sustainable livelihoods. These farmers develop themselves within sustainable agricultural activities, in a way that is economically profitable, beneficial for the environment, social processes and local culture. Vredeseilanden therefore invests in the capabilities of organisations of smallholder family farmers to encourage strong collective business organisations, while linking these organisations with other chains actors such as processors, traders, retailers, and consumers. Vredeseilanden is also active in setting up networks of farmer organisations and support these in their advocacy work with governments and private actors (Vredeseilanden, 2013).

**Mission:** Vredeseilanden is active within Sustainable Agricultural Chain Development. Their mission is to obtain better livelihoods for farmers. If these farmers manage to gain a decent income from sustainable farming then, i) they work themselves out of poverty, ii) they feed the world and iii) they diminish the pressure on our earth (Vredeseilanden, 2013).

#### **4.2.2 Oxfam-Wereldwinkels**

Oxfam-Wereldwinkels, founded in 1971, is the main Fair Trade organisation in Belgium and works with approximately 106 partners in Africa, Asia and Latin America.

Through national and international cooperation with producers and like-minded organisations, Oxfam-Wereldwinkels aims to create a large movement that strives for a just, sustainable world. In the South, Oxfam- Wereldwinkels collaborates mainly with disadvantaged producers who are organized in cooperatives, associations or other “groups”. These organisations must contribute

to the social and economic development of their members and the communities in which they live. In Flanders, Oxfam-Wereldwinkels has over 200 outlets where they sell Fair Trade products (Oxfam-Wereldwinkels, 2013).

**Mission:**

Oxfam-Wereldwinkels is a democratic volunteer movement that strives for fair international trade and thus fights for everyone's right to a dignified existence (Oxfam-Wereldwinkels, 2013).

#### **4.2.3 Traidcraft**

Established in 1979, Traidcraft is UK's leading Fair trade company. Traidcraft envisions a world freed from the scandal of poverty, where trade is just and people and communities can flourish. Consequently they build lasting relationships with producers, support people to trade out of poverty, and work to bring about trade justice. The organisation acts as a trading company and a development charity at the same time. Traidcraft consists of three main organisations. The Traidcraft Foundation is the 'umbrella' foundation and has the objective to relieve poverty in the developing world and, in particular, to ensure that Traidcraft remains focused on its goal of fighting poverty through trade and stays true to its Foundation Principles. The trade and charity work are split into two different organisations, i.e. Traidcraft Plc and Traidcraft Exchange. The first entity is the trading company, owned by individual shareholders and responsible for selling the Fair Trade products in the UK market. The overseas development programmes are implemented by Traidcraft Exchange which is responsible for promoting trade justice and developing approaches to trade that supports small-scale producers and enterprises. The main objective is not to reduce poverty by providing producers access to international markets, rather this entity focuses on making trade fairer for the poor and the producers they work with are not necessarily certified by the Fair Trade Labelling Organisation (FLO) and are encouraged to sell their products in domestic markets (Traidcraft, 2013).

Traidcraft works with more than 100 producer groups in over 30 developing countries to provide 450 Fair Trade food, craft and textile products. They have 120,000 customers and a turnover approaching £20 million. In its pioneering role Traidcraft has, over the years, set up many initiatives in the UK as well as in Southern countries to promote, together with partners, Fair Trade and its products (Traidcraft, 2013).

**Mission:**

We fight poverty through trade, pioneering, practising and promoting approaches to trade that help poor people in developing countries transform their lives (Traidcraft, 2013).

#### **4.2.4 Intermon Oxfam**

Intermon Oxfam engages in fundraising, campaigns and fair trade. It is the largest fair trade organisation in Spain in terms of its volume. The mission of Intermon Oxfam is to help bring about change that will enable the sustainable development for countries in the South and to strive for achieving just social structures in relations between peoples, fostering a culture of solidarity. Intermon Oxfam's Fair Trade is based on three main pillars of action: commercial partnerships, strengthening organisations, and raising awareness with consumers in the North (Intermon Oxfam, 2013).

Intermon Oxfam currently supports 113 producer groups in Africa, Latin America and Asia, offering them the tools and technical assistance necessary to join the international market (Intermon Oxfam, 2013).

**Mission:**

Intermon Oxfam are people who fight, with and for disadvantaged populations, forming part of a wider global movement, to eradicate poverty and injustice, and to ensure that all human beings can exercise their rights to the full extent and enjoy a decent life (Intermon Oxfam, 2013).

## 5 Evaluation approaches in practice: four case studies

As described in the previous chapter, four organisations were selected based upon the nature of their activities, their programme approach and their objectives. Three organisations are active in Fair Trade and implement support activities for their partners in the South; one organisation focuses on improving production processes in the South and supports partners throughout the value chain. Each organisation was interviewed either in person or via Skype.

This chapter presents an overview, per organisation, about their current evaluation practices, challenges, successes and ambitions, followed by a summary. Each case study follows the line of the interview questions (see appendix A) and information obtained from the interviews was supplemented or verified with information retrieved from documents and websites (see appendix B). In order to respect the ethical considerations (see chapter 1.6), the case studies are anonymous; they appear in random order.

### 5.1 Organisation 1 – “Proving rather than improving”

#### *Evaluation practices*

Currently the organisation uses the logic framework as a methodology for programme and evaluation design. They indicate that they follow a very conventional approach conducting a baseline evaluation, midterm evaluation and final evaluation, mainly driven by donors who use the logic framework as a standard tool for programme proposal and evaluation purposes. The objectives they pursue with their trading and support activities are i) improved and sustainable incomes for poor people, and ii) greater dignity and opportunity for poor people; however, the organisation admits that they are not able to provide a ‘full and rounded’ picture of the impact of their work.

The programme logic is drawn up during the project design phase together with the identification of changes the programme intends to achieve and the best indications of change that they can identify for the outcome areas. The organisation uses (at least) two different logic framework models (used for two donors) which, as a rule, contain the overall objective and specific objectives, both linked to quantifiable indicators and targets. The expected results or outputs in the logistic framework tend to have a more qualitative character and define targets that are less concrete, e.g. increase in knowledge, capacity, skills or improved relationships or cooperation etc. The organisation tries to find a good balance in the definition of quantitative

and qualitative indicators but they point out that in general donors tend to be more interested in quantitative information.

The organisation works with local producer groups. Contrary to most organisations they do not refer to them as partners but as local key actors. In order to review the results and outcomes of their programmes, representatives from the producer groups are involved in the project review meetings (focus group discussions). In these discussions the results and the impacts are established and consequently cross-verified with other stakeholders such as local government bodies, other NGOs, or experts, etc., depending on the programme. The general rule is that the organisation always makes an effort to verify their data as well as the impact on the project participants.

In the terms of reference of their evaluations the organisation currently only requires the evaluation of their implementation (intervention) activities. The evaluation objectives are i) to assess whether the project implemented the right interventions, ii) to assess whether the project's organisation structure, management, planning and implementation process were effective and efficient and iii) to draw key strategic and programmatic lessons. Evaluations (including the terms of reference) are drawn up in a joint effort by the local field office and the head office; the project evaluation group includes members of the organisation's project management team and head of the country team, representatives from the producer groups, programme officer from the organisation's head office and the responsible evaluation manager. The organisation encourages the evaluators and field offices to apply participatory approaches for gathering data and information; however, their attempts show different degrees of success as the field staff needs more training on how to use these types of approaches.

#### *Use and evaluation of theory*

The organisation has not yet developed an overarching theory of change or change model and does not formulate or evaluate explicit or implicit theories or assumptions that underpin the choice for their activities and objectives.

One of the reviewed logic frameworks contains explicit theory under the column 'assumptions', e.g. farmer groups are willing to work together and government bodies recognise the associations as farmer representatives. Remarkably, when talking about the use of theory during the interview, the organisation did not mention these assumptions.

Although implicit theories, beliefs, or assumptions, are not taken into consideration for the purpose of evaluation, it is clear that they exist. While reviewing their evaluation documentation it became obvious that the organisation has had experiences in the past that they currently apply as underlying rationale for their chosen activities. A good example is the organisation's belief that by organising medium and small enterprises, through empowerment and capacity building, the enterprises will become negotiators, gain access to better information, become more aware of public policies, and will be perceived as a stronger representation. As there are no theories, assumptions, or mechanisms defined the question is why the organisation thinks that all these positive responses will be generated. When asked in the interview the organisation answered that they believe that producer group formation is a methodology for reaching bigger change and that it has brought good results in the past, strengthening business opportunities and advocacy efforts. Working with groups is a key area of learning and the benefits of this approach are not perceived as theory anymore. They know from experience that working with small holder farmers is an effective way of working and it has become a common practice.

Thus it can be concluded that the organisation has a focus on evaluation if a programme worked and only to a limited extent verifies why it has worked (or not worked). Due to their participatory approach, focus group discussion and interviews with beneficiaries, some information will be received about limiting or stimulating factors during the programme. There is no systematic approach of formulating theory in the programme or evaluation design that enables the organisation to discover or research why the foreseen outcomes were met or not.

Based on the organisation's strategic documents, it becomes clear that their intervention logic derives from the implicit assumption that a higher income will lead to increased well-being and finally to poverty reduction. This overarching model of change, is not explicitly recognised, used or evaluated. The organisation states that they mostly focus on their outcome indicators as defined in the logic framework, increased well-being of the people for example, and that they automatically assume that when those outcomes are achieved, the poverty levels will go down.

The organisation realises they need to make their implicit assumptions more explicit, but the assumptions are very context specific and in reality they focus only on the logic framework. In the project design phase the organisation has the tendency to use a modified 'problem tree' approach; they identify the key problems faced by the communities they target, and then perform an analysis to understand why these problems exist. Out of this exercise the objectives are defined. They realise there are limitations to such an approach; a wider understanding of how change happens in a specific community would be beneficial as well as the application of a

so called ‘appreciative inquiry approach’, focusing less on problems and more on understanding of potential opportunities.

### *Complexity*

The organisation tends to deal with complex factors from a pragmatic point of view. They did not indicate that a complex environment influences or hampers the attribution of programme outcomes to their organisation; neither did they state that they are looking for better ways to deal with complexity in their evaluation efforts.

Before the programme design phase the organisation performs an extensive analysis, also referred to as feasibility study or environmental scan, in order to create a better understanding of the dynamics of the context within which they will implement their programme. No evidence was found however, that the organisation takes the context and possible changes into account when evaluating the programme results. It is the organisation’s *hope* that evaluators take the context into account but they are not sure.

The organisation is in the process of developing an increased focus on flourishing communities and individual well-being. This new focus area will lead to closer and more diverse cooperation with other actors. Currently, with the focus still on enterprise and trading, they seek cooperation with actors in the same specific areas (e.g. with other NGOs that are capable of dealing with more technical aspects of increasing climate change resilience or with implementation partners). While these partnerships are subject to evaluation reflection and are even defined in the evaluation’s terms of reference, they may not always appear in the evaluation report.

The requirement for intervention from multiple sectors becomes visible via the cooperation of the organisation with child labour projects; to combat or regulate child labour they engage with authorities and families. They also work with big UK companies in order to address the negative impact of their activities on the populations in the South. Cooperation with the government is evaluated and that is only when the results (e.g. change in policy) is defined as an outcome of the programme. Increasing voice and influence of small-scale producers are a core element of the change the organisation aims to reach. To date, they collect anecdotal stories and case studies in order to demonstrate the results, but they are seeking a more systematic evaluation approach.

In regards to attribution of long-term impact, the organisation admits they lack the practice of assessing the ‘counterfactual’, i.e. would the change have happened anyway, without their intervention.

### *Sustainability*

There is no clear definition of sustainability in the organisation. The interpretation of sustainability is primarily that the benefits of the organisation's programme continue once the intervention is over. Secondly, depending on the programme, there is also an interpretation that environmental and social criteria need to be met. Financial and institutional sustainability may also apply. Since much of the organisation's work is about increasing the representation of the farmers, they want farmer structures to continue to exist when they have left the programme. Concurrently, as they are in trade it is also about building sustainable relationships between the supply chain actors, building the links between the small holder farmers and the other actors in the industry. They try to achieve win-win situations for all parties so that there is a motivation for all to sustain the outcome.

### *Strength and weaknesses of the current approach*

An obvious weakness is the lack of theories or assumptions that guide the evaluation about the expected impact / change. Finding good evaluators is also a problem. They admit that there is still a lot to be done, but since they are very self-critical they are going the right way with their evaluation efforts. They have become already more systematic and more focused in their search for evidence.

### *Contribution and added value*

The added value should be to learn from experiences and to improve future results.

### *Challenges*

In general the feeling is that the organisation spends more time in *proving* than *improving* their work. Currently, the certification of Fair Trade producers is heavily criticised and the organisation knows that they should be able to demonstrate their added value which is the development of producers and their communities. They work with certified (FLO) and non-certified producers and have become quite proficient in developing their own standards themselves (e.g. for cotton) which allows them to work with producers that do not qualify or are too small for FLO certification.

There is limitation to what the organisation's ambitions are; mainly due to the resources they have available for conducting evaluations and training staff. Other continuing challenges are in relation to capacity and willingness of the project field staff to gather the data. That mindset is

probably their greatest challenge in order to get staff enthusiastic about it rather than see it as extra work load.

#### *Ambitions*

In its strategic plan for 2011 to 2014, the organisation announces the activities it will undertake building 'flourishing' communities. 'From fair to flourishing' is a new way of looking at the well-being of people and communities in developing countries. Together with a well-known university the organisation is currently working to develop a way of measuring perceptions of well-being among their programme participants. This new method of evaluation will allow the organisation to verify if they are promoting human flourishing as foreseen in their vision and strategic plan. Until the new evaluation method is launched, the organisation remains 'very wedded to the programme logic framework'.

In the past years several evaluation methods were explored and utilised. Over the next six months, the person responsible for evaluation in the head office will analyse the evaluation experiences so far and work toward introducing a method that offers the possibility to evaluate theory of change. The organisation has already thought extensively about how their interventions can reach the greatest impact; they have identified key leverage points for development and analysed the roles of all the actors involved (trade and supply chain).

#### *Summary*

Though they are working hard on developing a new evaluation approach that fits their specific objectives and activities, currently this organisation does not formulate any sort of theory or assumptions for the purpose of evaluation. At present the organisation uses the logic framework for programme and evaluation design and they feel that they have a conventional approach to evaluation, which serves primarily for proving to the donor that they have reached results rather than for the organisation itself, to learn and improve their practices. They want to change this. Besides the long-term attribution challenges, there seems to be little awareness about how the work of other actors can affect the results and outcomes of their programmes, although from a more positive perspective, they do perform a screening in order to better understand what is going on in an area and seek cooperation if they feel this can add value to their work. There is no evidence that the organisation evaluates changes in the context that may have affected their work and no evidence that the organisation seeks broader input from other sectors in order to address issues that are out of their mandate. It must be noted that the organisation's ambition is to work with change in the context that they can bring about rather than understanding more widely what makes change happen. Their ambition is to use a thick interpretation of sustainability but without a clear definition, they need to further elaborate this concept and

integrate sustainability indicators systematically in their evaluations. Although they consider their evaluations very important, the organisation struggles with finding sufficient and good resources (time and people); overall they feel that a change in mentality within their organisation is necessary. Evaluations are too often seen as extra work and not as an integral part of the programme. There is still too little appreciation of and commitment to evaluation work.

## **5.2 Organisation 2 – “Evaluation belongs in the development department”**

### *Evaluation practices*

This organisation has not conducted an evaluation in over three years for their Fair Trade activities in the South. They reduced their ambitions from contributing to the partner’s capacity and improving farmers’ livelihoods to just delivering technical support to improve product quality. The organisation was obliged to drop its development ambitions due to a severe lack of funding.

Now, the Fair Trade activities have the objective to supply Fair Trade shops in the home country of the organisation’s head office with products that can be successfully sold and to raise awareness among the public on the positive impact of Fair Trade on farmers’ lives. To this end the organisation conducts impact assessments amongst the farmers they trade with, although not systematically and not based on a certain donor proposal or strategic plan with impact objectives. This impact assessment takes the traditional approach: first a livelihoods assessment is conducted to develop a baseline and then after a certain amount of time the assessment is performed again to establish the differences in the livelihoods situation between a group of Fair Trade producers and a control group of conventional trade producers.

In order to meet the new objective, the organisation adjusted its purchase policy that in the past used to favour development potential over market potential. Now, the policy no longer attempts to work toward a balance between market and development. They currently purchase Fair Trade goods using three ‘filters’, namely i) the products must be Fair Trade (FLO certified, or otherwise known products within the organisation); ii) the products must sell successfully in the home market of the head office; and iii) if there are two producers with similar products that fulfil the aforementioned criteria, the organisation will continue with the producer that shows potential to translate the profits into development opportunities. Indicators to measure the development potential are gender equality, environmental friendly/neutral, and poverty levels of the country.

Three years ago, the overall objective for the Fair Trade activities used to be to improve farmers' livelihoods and to reduce poverty. In order to better understand and respond to the needs of the producer groups, i.e. their partners, they recently developed a comprehensive tool for needs diagnosis based on the European Foundation for Quality Management (EFQM) methodology. Back then, the logic framework was used as a tool for programme design and there were no systematic evaluations performed to measure the programme's impact on farmers' livelihoods and poverty reduction.

#### *The use and evaluation of theory*

Currently the organisation does not have any development objectives and does not perform evaluations on their Fair Trade activities. The conclusion therefore is that the organisation does not use theory, in either evaluating their implementation processes or their programme impact.

Until recently, the organisation did refer to the overall objective of reducing poverty by means of Fair Trade activities for farmers. In an old logic framework the organisation also sets the objective of increasing the income of the producers. When asked if theories or change models were formulated in the past, the organisation admitted that in the Fair Trade Department they did not. However, the organisation feels that the Development Department should be able to perform proper evaluations based on theories and models of change. In the Fair Trade Department the focus solely is on the increase of income of producers; the organisation realises however, that more money does not automatically lead to development or poverty reduction.

#### *Complexity*

As the organisation currently does not conduct evaluation on programme development objectives they therefore do not (have to) deal with complexity issues. They show full awareness of the complexity of the environment in which they implement their programmes and hope that in the future they can re-introduce their needs assessments and evaluations.

#### *Sustainability*

The organisation does not use an academic definition of sustainability but in general the organisation interprets sustainability as: "continuity in the production and purchases from the producers without exhausting their land."

The commitment to partnerships with producer groups is considered a long-term intervention. This implies a certain ambition to work sustainably. However, there are no concrete exit criteria and the organisation states that long-term commitment is good, eternal commitment is not. The organisation needs to further elaborate on the exit criteria, at the moment they only know that a producer group is independent when they are less interested to sell their products to the organisation (meaning that the producer group has gained sufficient access to other markets).

#### *Strength and weaknesses of the current approach*

The strength of the producers' needs diagnosis was that it offered a detailed picture in the relevant context of the producer group. The strength of the current impact analysis is the opportunity it presents to communicate to the public in the organisation's home country and present the impact on farmers' livelihoods due to Fair Trade purchases. An identified weakness in the past was for the Fair Trade activities in craft products. These producer groups were not able to trigger changes in their community or generate opportunities to climb out of poverty. The organisation never really understood why this was the case and currently recognises that with implementation of a proper evaluation approach, a reason may have been identified.

#### *Contribution and added value*

If there was available funding for conducting evaluations the organisation is convinced it would contribute to more awareness, learning and understanding regarding opportunities for development. Evaluations would also enable the organisation to develop good products.

#### *Challenges*

Challenges at this moment are to regain funding opportunities and to maintain and increase the opportunities for Fair Trade in the home country. If the situation is to improve, the organisation will first need to decide how to continue with their development ambition and which department is best suitable to define objectives and activities and conduct evaluations.

#### *Ambitions*

Future opportunities for evaluating the impact of development objectives for the Fair Trade farmers could be accommodated by the Development Department of the organisation. It was noted that the Fair Trade Department has expertise and skills in international trade and certification, while the Development Department has expertise in defining and evaluating objectives for development. The organisation has not taken a decision yet concerning this proposal.

### *Summary*

Due to a lack of funding this organisation currently does not perform evaluations. They had to cancel their development objectives of Fair Trade and no longer support their partners in order to build their capacity. They still conduct impact assessments for the use of public awareness raising about the advantages of Fair Trade; these impact assessments have a conventional approach, they measure impact by means of observable outcomes. So far, the organisation has not developed an academic definition for sustainability. When asked though they indicated that they use a thin interpretation of sustainability, that is to say, in the context of continuity of their programmes' efforts and respect for the environment. In the future the organisation wants to re-start development and evaluation activities. They consider both important components of their work, however, which department in the organisation can best perform these activities is yet to be decided.

### **5.3 Organisation 3 – “Focus on our circle of influence”**

#### *Evaluation practices*

Over the last three years the organisation has spent a considerable amount of time and funds in order to re-think their sustainability ambition and implement a new evaluation methodology. They currently use the methodology of *outcome mapping* in which qualitative and quantitative data are collected for evaluation purposes and conventional reporting on programme results is combined with theory-driven reporting of results

Recognising a lack of data and knowledge about the current level of sustainable practices within their programmes and their environment as well as the concerns about the sustainability of their approach motivated this organisation to provide answers to the questions i) how to define sustainability in the context of a value chain development programme ii) how to make sustainability in its many dimensions actually happen and iii) how to measure progress. The organisation decided to embark on what they called a ‘sustainability trajectory’. First the organisation’s views on sustainable agriculture and value chains were modelled in a ‘vision’ in 2011, after which a road show was organised to reach out to overseas staff, partners and farmers to motivate them to adopt this new vision of sustainability and bring it into practice. In the last phase the organisation performed a meta-evaluation covering all overseas countries they operate in. The organisation asked the overseas offices, in cooperation with their partners, to perform a thorough assessment on the potentials of the chosen sub-sectors, the level of sustainability of the selected value chains, and the livelihood situation of the farmer members in

the cooperatives. This exercise provided the organisation with a solid baseline and enabled them to measure impact of their implemented programmes.

By making informed choices on agricultural sub-sectors and value chains and additionally understanding the needs and challenges of farmers to improve their livelihoods in a sustainable manner, the organisation prepared itself to take the next step and implement an evaluation methodology called *outcome mapping*.

The essence of the *outcome mapping* methodology is to identify partners that are (or can be) active in a value-chain and that have potential to change their behaviour in order to add value to the farmer's products, for example by means of stimulating sustainable agricultural techniques or to promote a (more) enabling environment that will eventually lead to the increased sustainable livelihoods of farmers. These partners are referred to as *boundary partners*; they have an active role in the change model and form an 'interface' between the beneficiaries and the organisation. Although organised family farmers form the core target group of the organisation's intervention, the overall purpose is that all *boundary partners* in the value chain benefit from the changes pursued in the programme as well. The organisation believes that the best approach to development is to work through local partners and focus on behaviour changes within their partners' practices. This way the organisation aims to work on and evaluate what lies in their circle of influence.

For each selected value chain (in some countries this organisation works in more than one sub-sector and / or value chain) the organisation draws up a Chain Intervention Framework (CIF). Besides an in-depth description of the sub-sector and the value chain, this framework identifies the target groups and the key issues these people face in their respective farming practices. The overall aim of the programme is specified in the specific objectives and (expected) results. Both the objectives and the results contain measurable indicators that can be compared with the baseline. In addition to the indicators, the intended results also contain an annual target to work toward.

The objectives and the results are directly linked to the activities of the *boundary partners*; they can change their behaviour and practices to reach the change necessary to increase the income from trade. The organisation's efforts therefore, focus on support, i.e. the *supporting strategies*, such as funding, training or coaching, thus enabling the *boundary partners* to adapt their behaviour and practices and work toward better results in the value chain. This necessary change from the partner is referred to as the *outcome challenge* and is established in a joint

effort between the organisation's overseas field office and the partner itself. The *outcome challenge* can be monitored through *progress markers* which establish the activities that the partner is supposed, and has agreed, to undertake with the support received from the organisation.

It must be noted that the organisation works in three year cycles, hence the CIF is a framework, laying out change and activities for a three period. Evaluations on the CIF are done twice a year, one midterm and one at the end of each year. The evaluations are documented through means of a Chain Intervention Report. in which the livelihood situation of the target group is assessed and reported, changes in context are reflected, progress on objectives and results indicators are monitored and reported, activities from the partners (*progress markers*) are screened, and adjustments for the programme are suggested for the time remaining.

Ownership of the evaluations lies with the value chain programme manager (of the overseas offices) and all the *boundary partners* active in that value chain.

#### *The use and evaluation and of theory*

Although the word *theory* did not come up much during the interview, the organisation referred to it as 'assumptions', we can conclude that this organisation evaluates if a programme worked and consequently uses impact theory to understand why a programme worked or not.

The organisation develops a change model that establishes their thoughts on how change happens by defining *outcomes challenges* together with their boundary partners. As mentioned before, the core thought behind the partners' *outcome challenges* is to create more benefits for all parties in the value chain by changing the *boundary partners'* practices. These *outcomes challenges* and their underlying activities, also referred to a *progress markers*, are the mechanisms (Pawson and Tilley, 2004) or determinants (Chen, 1990) in the change model, see figure 2, that will make the intervention lead to the change foreseen. In the CIR the *progress markers*, the context, intended and unintended achievements so far, and *supporting strategies* and, moreover, effectiveness of the organisation's support strategies are all reported on through a narrative methodology. Another part of the CIR consists of a quantitative evaluation, presenting data by means of collecting and measuring observable evidence and comparing it to a baseline. This evidence establishes whether the programme worked or not. Numbers and percentages form the core data shown in this section; nevertheless it is complemented with a section in which deviations from the targets are explained.

Whether this organisation evaluates implementation theory (how the organisation makes this change happen) is less obvious for some components. It is clear that the partners take a prominent role in the organisations' strategy and objectives and one would expect that the selection of *boundary partners* is a well thought-out decision, based on selection criteria and complemented with assumptions (or observations) as to why the partner fits in the programme and performs well. However, the rationale for the selection of the partners, including the organisation's assumptions that this partner would be a successful and well-motivated *boundary partner*, are not made explicit in the CIF and are not evaluated in the CIR. It is the same issue for the selection of *supporting strategies* and *progress markers*. Together with the partners these activities are identified, and in the CIR their progress is monitored. The selected *supporting strategies* (activities of the organisation) and *progress markers* (activities of the partners) are evaluated on whether they happened or not, however, they are not compared to any underlying prior assumptions. Therefore this evaluation highlights what was achieved by the organisation and their partners but provides little insight into whether their expectations were met.

### *Complexity*

In general it can be said that two out of three researched components of complexity, i.e. context and multiple actors / sectors, are recognised for their potential influence on project outcomes and are systematically evaluated for changes, challenges and opportunities. The *outcome mapping* approach facilitates evaluating context changes and multi-actor and -sector collaborations. Attribution of impact on the long-term remains a challenge for the organisation.

Before the design phase of a programme a thorough context analysis is performed by a local consultant or the field staff in order to establish if the context is supportive for what the organisation intends to do. *Boundary partners* and other actors for cooperation are identified (e.g. a university for technical training). In the context analysis they also identify other actors active in the same geographical and programmatic area. The organisation indicated they have a strong focus on including multiple stakeholders and are very keen on cooperation with strategic partners (e.g. government), technical partners (e.g. university) or other NGOs.

Since the organisation's approach includes the whole value chain (including farmers, factories, traders), it further suggests that they include multiple sectors in order to increase the impact of their interventions. Moreover, if the context analysis shows issues that are outside the scope of the organisation, they will seek collaboration with other parties to overcome these problems (i.e. they cooperated with a rehabilitation centre when they worked in an area where drug abuse was high).

The last component of complexity, attribution of programme effect on the long-term, remains unanswered in their evaluation approach. The organisation set the *outcome challenges* of their partner for a period of six years. After this period, no more evaluations will be performed. The organisation indicated that in the new strategy they want to reach structural changes in a sub-sector (e.g. rice) and they want to facilitate small farmers' access to modern markets in the longer-term. They intend to draft a strategy for evaluation of these specific objectives, but already stated that they do not have the ambition to work alone to reach these objectives. They feel the work could be partially done by other organisations, or by a consortium of partners. They gathered there will also be preconditions to reach this change; and these preconditions may very well not be in the hands of the organisation.

### *Sustainability*

As mentioned in the evaluation practices, the organisation developed its own vision on sustainability and conducted an extensive exercise with the purpose of understanding where they were with this vision and what should be done in order to implement their vision on sustainability.

The application of sustainable criteria and objectives is two pronged: they want to implement sustainable practices throughout their chain intervention activities and they want to reach sustainable impact outcomes for their target groups. First of all the organisation aims to promote and implement sustainable agricultural activities throughout the targeted value chain. Sustainable chain interventions should be economically viable, environmentally safe, socially fair, and culturally acceptable. The organisation also wants interventions and their impacts to “contribute to the development of sustainable livelihoods by family farmers, at scale and lasting”. More specifically the organisation wants their interventions to i) lead to empowerment and long-term economic gain for small-scale farmers, in ways that are ecologically sound, socially just and culturally adaptable and ii) contribute to an institutional environment that facilitates lasting impact and up-scaling.

In order to assess if these objectives were obtained, a separate impact assessment is performed, referred to as the Sustainable Livelihood Analysis (SLA). This tool examines sustainable livelihood indicators of the farmers such as their vulnerability context; their capital/ assets; policies, institutions and processes from other key actors; their livelihood strategies; and their livelihood outcomes.

### *Strength and weaknesses of the current approach*

Staff in the field appreciate the structure in the evaluation system, particularly for its planning and reliability aspects. It has enabled the organisation to implement its sustainability ambitions and appears to be a flexible evaluation tool as it grows with the programme and is easy to update (i.e. in making adjustments to the programme). A logic framework, they believe, is very static.

The organisation has a fulltime staff at its headquarters dedicated to Monitoring and Evaluation. Moreover, it was a deliberate choice to invest in identifying a proper methodology for evaluation.

### *Contribution and added value*

The organisation wants to continuously monitor the possibility for scaling-up their activities. As soon as they feel, and can demonstrate through means of proper evaluation efforts, that they have found a programme activity or approach that may work well and be replicated by other organisations (NGOs, government authorities etc) working in a similar context, they will seek ways of handing over and/or training these organisations in applying the same strategy.

Furthermore, they indicated that programme learning and adjustment is an enormous advantage of the methodology.

### *Challenges*

The organisation has now worked for three years with this approach and they have trained all the staff in the field. They believe that a challenge will remain to keep staff 'on board' and make sure they recognise evaluation is important and adds value to the programmes and their work.

An internal staff member at HQ has introduced the new evaluation approach. Their main donor has never accepted their choice for evaluation approach, despite efforts to present and train them on it. They have found a proper mechanism to translate their *outcome mapping* to the evaluation and reporting models required by the donor, but it requires extra resources to do this.

### *Ambition*

The organisation wants to be able to measure the long-term impact of their programmes. They are working on this.

### *Summary*

Thanks to an enormous effort in re-thinking their sustainability ambition and implementing a new and suitable evaluation methodology, *outcome mapping*, this organisation has made steps forward in their evaluation efforts. They collect qualitative and quantitative data and combine conventional with theory-driven evaluation on outcomes. The method *outcome mapping* shows similarities with the realistic evaluation approach. The organisation has a strong emphasis on impact theory and they implicitly look for successful context, mechanism and outcome formulations so they can upscale their activities through other actors. The organisation deals with complexity in a pro-active manner. They perform thorough baseline environment analyses and seek collaboration with other actors, in their own sector as well as in other relevant sectors to support and increase their programme's impact. The organisation recognises their programmes are influenced by many factors, which can support or diminish long-term change objectives; this has led them to seek an evaluation methodology that can incorporate measuring their programmes' impact in the long run. Regarding sustainability, the organisation uses a thick and well thought-out interpretation. The sustainable objectives however, seem to be integrated in quite a complex mix of activities and outcomes, targeting all stakeholders of the programme. With no identified evaluation methodology, the issue of how to measure the long-term impact of some important links remain under-theorised, e.g. the link if sustainable family farming leads to sustainable livelihood incomes or the link if improved livelihood incomes lead to poverty reduction.

#### **5.4 Organisation 4 – “Long-term impact is inherent to Fair Trade”**

##### *Evaluation practices*

In this organisation there is no systematic evaluation approach. Existing monitoring practices primarily focus on the Fair Trade working of the producers, i.e. do the producers meet the Fair Trade standards as set by the FLO or by the organisation itself (for small producers). Only since recently the organisation implemented partner support activities. Recognising that the Fair Trade market gradually had changed from a pioneer market to a conventional trade market, allowing many commercially-driven players to respond to the growing demand in Fair Trade products, the organisation felt it needed to reconfirm the exclusiveness of its approach to Fair Trade. This claim of ‘uniqueness’ demands a new and more comprehensive approach to evaluation.

Additionally, the organisation's choice to aim for small and disadvantaged producers made them realise that the income received from the Fair Trade activities would not cover all investments necessary to enhance this long-term development. Additional funding from donors was sought

to fund gaps in agricultural and production knowledge, tools and materials, quality of land and plantations and so on. Moreover, external funding is also used to enable the partner (e.g. the cooperative) to perform such activities with their member producers. Donors that are willing to contribute funds want to receive extensive reporting on development indicators and therefore also demand a more inclusive approach to evaluation.

In short, the organisation is at the very beginning of setting up, professionalising and improving the use of evaluations that go beyond just Fair Trade performance. So far the organisation has gone through one complete evaluation cycle covering the producer groups support activities and development objectives. It is too soon for this organisation to draw rigorous conclusions on the implementation or the rationale behind their evaluation approach and practices; subsequent chapters of this paper will primarily recapitulate the evaluation 'intentions' the organisation has had so far.

Unlike the evaluation of the impact of their programmes on development, the organisation has a long history in monitoring and evaluation the performance of their partners. The organisation indicated that they monitor a set of fixed indicators, namely i) the amount of Fair Trade products imported into the European Union (this indicator establishes that the standards of the products conform to European standards and that the producers have (developed) the capacity to export) and ii) the amount of farmers and communities that join the cooperatives (this indicator is based on the belief that a successful and well-performing cooperative will attract more members). These two indicators are monitored on a periodic basis and when any irregularities are found, further research into the underlying causes is performed.

Furthermore, the organisation uses the monitoring database of the European Fair Trade Association (EFTA) in order to share and/or retrieve performance information on cooperatives. Another external monitoring tool is the annual report that FLO-certified cooperatives are required to prepare for their members. This report provides useful information for the organisation when they want to start up programmes with a new certified partner.

The single evaluation that was conducted to measure development outcomes, was based on a logic framework with one overall and several specific objectives formulated as well the expected results. The indicators for measuring progress on objectives and results form a mix of quantitative and qualitative information. The evaluation exercise that followed the programme shows an interactive and participatory approach; prior to the evaluation, the external evaluator spoke to the producers in order to better understand their needs, issues and challenges. Based

on this input, the evaluator continued to review the progress made during the programme cycle. Both quantitative and qualitative data were gathered, and deviations from the project results and foreseen impact were explained based on the inputs from a wide variety of stakeholders, including the partner, the beneficiaries (producers) as well as community leaders and government officials.

#### *The use and evaluation of theory*

Using a highly interactive and participatory approach, the organisation establishes in its evaluation if a programme worked and why a programme worked. However, the fact that the organisation seeks to clarify why their programmes did or did not work does not imply that the organisation has adapted a theory- driven approach to evaluation. No evidence was found that the evaluator attempted to surface implicit theories or assumptions for the purpose of evaluation, meaning that the organisation's evaluation approach is conventional, measuring concrete results rather than testing theories.

In the terms of reference lining out the evaluation assignment, the main objective is “to evaluate the outcome of Fairtrade and financial support on sustainable agriculture production to improve economic, environmental and social conditions of farmers in Africa” and consequently continues with “...the study will evaluate the coherency and efficiency of the implementation of activities planned by each partner to reach the results from the project”. The terms of reference and the actual evaluation suggest there is a strong interest in evaluating the effectiveness of the *implementation activities* and *impact experienced / observed by programme stakeholders*. The impact on farmers' livelihoods was evaluated but the evaluation outcome states that “impact on the farmer's livelihoods was difficult to measure, as at the start of the programme no baseline survey was conducted, hence there was no data to compare possible progress with”. So while the organisation does include the evaluation of implementation processes and impact on producers' livelihoods and social conditions of the communities, the evaluation remains limited to establishing declared and observed impacts and does not include the assessment of any underlying theories.

It must be noted though that the terms of reference requests to assess whether or not the theory of change is clear and coherent considering the organisation's specific contribution. In the actual evaluation there is no mention of a theory of change, and again, no evidence was found in the evaluation that theories or assumptions were formulated for the purpose of the evaluation.

The fact that there is no explicit theory formulation does not mean the organisation does not use theory at all. Based on the organisation's interpretation of what Fair Trade is about and how actors behave, it can be deduced that the organisation builds its programmes around numerous theories and assumptions. As mentioned earlier, the organisation sees Fair Trade as a long-term development strategy, i.e. they consider trade as a key lever for the development of disadvantaged producer groups. During the interview it became clear that the organisation's change model (or intervention logic) of Fair Trade is based on their belief that fair access to international markets will help producers climb out of poverty; the higher income and the protected Fair Trade market will facilitate the producers to become more socially resilient. The organisation is convinced that the Fair Trade premium, not to be mistaken with the guaranteed minimum price they pay the producers, will be invested by the cooperatives in the social requirements of their member communities. The only criteria that the organisation requires are that the spending of the premium is decided upon in a democratic manner and benefits the whole community. The generic implicit theory behind the Fair Trade change model is that higher incomes for farmers will lead to the increased social resilience of their communities, which will result in poverty reduction. The organisation has adopted this specific Fair Trade change model, including assumptions about the intervention, determinants and expected change. In the evaluation however, only the intervention is evaluated; the determinant (Chen, 1990), i.e. the spending of the Fair Trade premium, is not subject to evaluation.

### *Complexity*

A complex environment is not perceived as a threat or a limitation to demonstrate the impact of the organisation's programmes. Regarding the organisation's approach to complexity, they tend to be pragmatic and seek cooperation in order to increase the impact of programmes or to deal with practical issues.

The organisation always produces a context analysis prior to initiating a programme. The context analysis entails the political, economic and agricultural situation, historical background, trade relations and limitations, and the level of political and social movements. These context analyses are said to be pretty standard; the United Nations' ranking list of Least Developed Countries is also considered. In the evaluation however, the organisation does not report on changes in context that may have affected the outcome of the programme. The organisation indicates that the evaluation is primarily about the partner's work and not as much about the organisation's role and the context.

The organisation has no policy or system to analyse the role and activities of other actors in their segment of cooperation. On the other hand, the organisation cooperates with other actors if that will increase their impact. An example of such a cooperation agreement is with an alternative bank that can provide the farmers with credit. The organisation states that this cooperation sometimes leads to issues around the end goal of the programme and about the process of how to get there.

In general the organisation seeks cooperation with other sectors when they conclude that certain Fair Trade standards are not met or are weak, e.g. they introduced a child registration system in order to control child labour.

The desire to reach long-term impact is inherent to Fair Trade; the partner cooperation is set for long-term commitment, from both sides. However, the organisation admits that their exit strategy is not very clear, despite their internal criterion to build the capacity of partners so that they can function independently. The organisation realises that they will have to define in more detail what their role is in the long-term with their partners once objectives are attained. Currently the reality is that the organisation only ends partnerships when a product is no longer marketable, in other words, the support to the partner is linked to the tradability of the product. This is not in alignment with the organisation's mission. As the organisation has only recently begun evaluating the partner support activities and their impact on sustainable development, it is too soon to draw conclusions about the organisation's way of handling the complexity concerning attribution of change in the long-term.

### *Sustainability*

Apparently sustainability seems to be embedded in two processes: the production process and the development process. Sustainability is grounded in two important documents the organisation uses. The first is the standard matrix from the Fair Trade Labelling Organisation (FLO) for small producer organisations. These standards contain environmental and social criteria regarding the Fair Trade production process.

Additionally the organisation has adopted the definition and principles of sustainable development as internationally accepted by the United Nations Conference on Environment and Development (UNCED), also referred to as the World Summit on Sustainable Development held in Rio de Janeiro in 1992. The organisation's view on sustainable development is that it implies an integral economic, social and ecological development. They feel sustainable development is a

never-ending process without a finishing point and constant weighing up of economic, social, ecological / participative elements.

Regarding their evaluation practices the organisation indicates that i) ecological objectives are evaluated but the input is a donor driven rather than the organisation's internal sustainability definition as mentioned above, ii) economic objectives are evaluated as they have been made quantitative and are included in the donor logic framework, iii) social welfare objectives, e.g. poverty rates, are not defined and not evaluated, and iv) long-term objectives are not evaluated as such although the cooperation with and commitment to partners have a long-term focus.

#### *Strength and weaknesses of the current approach*

There is no systematic approach, which offers the advantage to conduct evaluations that are completely adapted to the context.

#### *Contribution and added value*

Evaluations are used for learning, adjusting programmes midway and accountability towards the donor.

#### *Challenges*

The organisation tries to find a balance between a good evaluation approach and the work load that comes along with this approach; they have not yet resolved how to implement proper data management. Moreover, the exchange of information is not well facilitated by an established system which results in difficulties when analysing trends among partners. The organisation also indicates that they are not adopting or utilising new technologies (e.g. the use of GIS to map schools, water points, roads, urban areas, etc., and understand how the presence or absence of basic facilities and services contributes to / hampers development). The organisation points out that currently there is little overview of the ensemble of monitoring and evaluation activities.

#### *Ambitions*

In 2013 the organisation aims to commence a thorough exercise identifying all existing relevant monitoring and evaluation practices and analyse their strengths and weaknesses in order to find the best evaluation approach for the organisation's core activities (trade and partner support) and objectives (higher income, sustainable development and poverty reduction). They indicate that they would prefer to switch to a systematic evaluation approach that leaves room for the interpretation of specific contextual factors.

### *Conclusions*

Until today, this organisation has only performed one evaluation of support activities to their partner and sustainable development objectives. In this evaluation the approach was participatory; many stakeholders were interviewed including farmers, community leaders and authorities. However, the evaluation approach falls short in surveying theory and assumptions that underlie the model of change and the intervention logic. Two handicaps come to light that may explain why there is no evaluation of theory; i) the organisation does not use indicators for measuring poverty or resilience which may indicate that these concepts have not yet been clearly defined, ii) the organisation does not set a baseline for the farmers' livelihood situation when they start a programme. The changes sought in livelihoods, social resilience, and poverty reduction remain at the free interpretation of the farmers and their communities as no baseline and no specific targets are defined. Thus, the evaluation only focuses on the impact that beneficiaries' self-report and, without objective evidence, that is attributed to the organisation's programme activities. The conclusion is not that a lack of objective data is wrong per se, but rather that the organisation should continue with this practice if it is the result of a well-thought out strategic decision. With respect to complexity, the organisation does not anticipate on possible issues in measuring their programme impact, they are practical and cooperate with other actors if they believe it will increase their impact. Sustainability is well-defined in the organisation and could be labelled as a *thick interpretation*. Confusion in future evaluation efforts could arise as two different processes ask for a sustainable approach (trade and production and development). In addition to the observation that there are no clear definitions of key concepts such as poverty and resilience, the links between a higher income, social resilience, strong(er) position in the economic context, sustainable economic development and poverty reduction are under-theorised.

## 6 Research findings and observations

This thesis aimed to answer two main questions. The first question was: *In an academic literature review, which major evaluation approaches are considered suitable to measure results in a complex environment?* In chapter two an extensive overview is provided of this literature research. In short the academic community did not seem to agree on whether or not organisations should be willing to explain why their programmes had impact or not.

The second research question attempted to answer to the following: *How can the evaluation approaches of a selection of organisations promoting Fair Trade and/or sustainable development be characterised? Do the selected organisations take certain components of complexity into account in their evaluation approach? If so, how do they evaluate them? Do the selected organisations reflect certain components of sustainable development in their evaluation approach? If so, which ones and how do they evaluate them?*

In chapter five, the individual stories of the organisations that were interviewed provided a ‘rich’ overview of their evaluation practices, successes and challenges.

In this chapter the general conclusions from the case studies will be connected to the discussion of the literature, beginning with a short summary of the literature study and then discussing the results of the interviews comparing them to the literature study outcomes.

It must be noted that academics and practitioners do not use the same evaluation terminologies. The interviewed organisations showed very little recognition of common academic terms such as theory-based or theory-driven evaluation, programme theory (implicit or explicit), mechanisms or determinants, or realistic evaluation. On the other hand, common practitioners’ terminology was not often found in the academic literature, such as: intervention logic, indicators, means of verification, etc. Moreover, some terms are used interchangeably or with different meanings in different organisations, e.g. theory of change versus change model (or intervention logic).

In literature some academics promoted the use of conventional evaluation approaches, focusing on establishing whether the programme had reached its intended outcomes or not. Other academics promote a more extensive form of evaluation, stating that the conventional method of evaluation fails to deliver information for programme improvement and generalisability of outcomes. As an answer to finding out *why* programmes did or did not have the intended effects these academics promote the use of theory-driven evaluation. They distinguish in formulating

theories on the action of implementation, meaning the operational and internal processes of programme execution on the one hand, and theory for the causal mechanisms that will lead to the programme impact on the other hand. In the literature reviewed, two evaluation approaches were often subject to discussion: theories of change and realistic evaluation. Both evaluation methods are theory-driven and have emerged to fill the deficit in conventional programme evaluation (Blamey and MacKenzie, 2007). Critics of the theory-driven evaluation approaches counter that working with theory is too complex for organisations, i.e. they do not have the expertise, and it will demand an unreasonable amount of time and resources (Scriven 1998, Stufflebeam 2001). Other authors also question it is necessary that organisations investigate the nuts and bolts of a programme trying to understand why their programmes have produced intended effects or not (Scriven, 1998).

In the real world of the interviewed NGOs however, discussions about evaluation approaches and complexity are not as advanced as in the academic world. There is agreement that programme evaluation is necessary to create an understanding if and why the organisation's programme produced (or not) the expected outcomes. However, the use of theory to gain this understanding is minimal and complexity in general is not perceived as a hindrance in attributing programme outcomes to the organisation's efforts. An important given is that three out of four organisations are still in an exploratory phase about the use of appropriate evaluation approaches.

It was noted that all four organisations built their programmes on a similar 'overarching' model of change:

*The interviewed organisations aim for improved livelihoods of small scale, deprived, farmers by promoting sustainable agriculture, farmers' cooperatives, ownership of and/or partnerships in the production value chain and access to local, regional or international markets<sup>2</sup>. The improved livelihoods will lead to reduced poverty rates.*

The interviewed organisations believe in general that the increase of income (from better quality products and (Fair) trade activities) will lead to improved livelihood. They also assume that the increase of income and improved livelihood will benefit the communities of the farmers. All in all, in the long-term poverty rates among these deprived communities should go down due

---

<sup>2</sup> Vredeseilanden focuses on improving the working of the value chain rather than facilitating access to markets.

to their respective interventions. The programme activities are intended to support partners in solving issues or shortcomings that hamper the promotion of sustainable agriculture, farmers' cooperatives, ownership of and/or partnerships in the production value chain and access to local, regional or international markets.

While the umbrella objectives stated above suggest the organisations incorporate a foundation for programmatic design and evaluations based on theory, only one organisation interviewed for this paper reported utilising explicit theory for programme design and evaluation. They have a strong focus on constructing theory to measure their impact; the use of theory for evaluating their implementation was less obvious though. This organisation did not 'label' their evaluation approach as a theory-driven evaluation. Their methodology of *outcome mapping* uses a distinctive vocabulary with words such as *boundary partners*, *outcome challenges*, *progress markers*, etc. When asked, they did not 'recognise' the use of theory in their evaluation approach, preferring to call it 'assumptions'.

The other three organisations currently use a conventional approach to evaluation, i.e. measuring results, outputs and outcomes rather than testing theories. All of them show ambition to develop and implement a more comprehensive evaluation approach in order to better understand the outcomes of their programmes. None indicated that they want to use a theory-driven form of evaluation. While the academic community assumed that evaluators would surface implicit theories as part of the evaluation exercise, in reality none of the interviewed organisations has assigned this task to an evaluator.

The obvious presence of implicit theories however, doesn't mean that organisations cannot (or should not) evaluate these theories. Weiss and many other scholars assume that it is normal practice that evaluators start their evaluation assignment with the exercise to establish which theories underlie the programme rationale, implicit and explicit theories. The evaluation will then focus on these theories and try to establish whether they were good theories or not. Birckmayer and Weiss (2000, pp. 409) clarify:

*"The theories in theory-based evaluation can come from various sources. Some writers suggest that they should be social science theories with a reasonable scientific pedigree (Chen and Rossi, 1992). Other authors try simply to set forth the logic that must be involved if a program is going to work, and they may even name the program theory a "logic model" (Coffman 1999; McLaughlin and Jordan 1999). For example, a program that aims to increase safety in a public housing project by giving tenants representation on the managing board*

*must assume something like the following: If tenants are members of the managing board, they will (a) demand greater security precautions and, because they are on the board, will be able to get additional security staff; or (b) keep their eyes on the property and residents and interlopers to forestall crime and vandalism; or (c) encourage their neighbors to take more responsibility for security. Nothing fancy or high falutin in the way of theory is involved here”.*

It must also be noted how little organisations use prior research from other organisations, institutions, universities, or even their own experiences in order to declare certain theories ‘proven’. As was mentioned in chapter two, Weiss suggests that “if good knowledge is already available on a particular point, then we change its label from ‘hypothesis’ or ‘assumption’ to something closer to ‘fact’ and move along” (Weiss, 1995, pp. 70). Organisations would first have to do conduct an internal exercise to find out what they believe can be considered facts based on their extensive experience.

Two of the interviewed organisations evaluate the improvement of livelihoods. Both organisations use a different method for this livelihood assessment than their own evaluation approach; it is used as a separate tool (in both cases it was the Sustainable Livelihood Approach as developed by DFID<sup>3</sup>) and in fact, neither of these two organisations initially mentioned in the interview that they do this type of assessment. Only through further questioning, was it revealed that this impact assessment was carried out. Another organisation indicated that they are currently cooperating with a university that performs research about the ‘well-being’ of people in order to make this concept more concrete. The outcome of this study will eventually be integrated in their programme and evaluation targets.

To date, no organisation evaluates the impact of improved livelihoods on poverty reduction. In fact, none of the organisations has a strategy, policy or tool for evaluating long-term effects of their programme interventions.

The lack of knowledge about or experience with the use of theory and theory-driven evaluations does not mean that the interviewed organisations are not exploring what would be a suitable evaluation approach for their programmes. On the contrary, all the organisations confirmed they consider evaluations paramount to proving the impact of their work; moreover, each organisation indicated they were in the process (or recently had been in the process) of

---

<sup>3</sup> United Kingdom Government’s Department for International Development

developing a systematic evaluation approach that would, one way or another, define a clear intervention logic (change model) and clear(er) indicators for change. In their evaluation discourse practitioners do not distinguish conventional from theory-driven evaluation approaches, nor do they seem to internally debate if and how to formulate theories or assumptions for the purpose of evaluation. The observations below may explain why the academic and the practical world are not aligned in their consideration of evaluation approaches:

- i) Only one organisation has the objective to generalise successful programme formula in order to upscale its impact. It is the same organisation that only recently started to evaluate impact theory. Other interviewed organisations did not indicate their objective is to know and predict successful programme activities in order to ‘export’ them to other actors or other settings.
- ii) The interviewed organisations mentioned certain limitations they experience in the fulfilment of programme and evaluation activities. These challenges do not seem to appear on the radar of academics.

Donors have a lot of say in how organisations perform their evaluations. They ‘dictate’ the rules of the evaluation process, timing and formats. Apart from the discussion whether evaluations should be theory-based or not, the interviewed organisations all indicated that they struggle financing their evaluation efforts. To implement a suitable evaluation approach, get all stakeholders on board and conduct evaluations on a regular basis (at least once a year) requires a considerable investment in time and resources. Thus donors decide on the evaluation requirements while simultaneously not providing funds to conduct the required evaluations.

Furthermore, the majority of the programme managers and programme staff do not seem to appreciate the potential added value of evaluations. Evaluations are considered extra work in order to convince the donor their money was ‘well’ spent.

In addition, after conducting the interviews and reviewing the documents shared by the organisations, it became clear that complexity for these organisations goes far beyond what is discussed in academic literature. Examples of this expanded perspective of the ‘internal environment complexity’ that organisations have to resolve are:

- Proliferation of standards to adhere to (certification standards by for example the Fair Trade Labeling Organisation or the internal standards for producers that

are not FLO-certified, sustainability standards applied to development, agriculture, production / value chain, partnerships, cooperatives / producer groups, business environment, trade relation). These standards cover ecological, social, economic and long-term objectives;

- Multiple donor strategies that with different reporting requirements;
- Pursued outcomes that are not within the expertise (multiple deprived communities). Organisations end up with a variety of partnerships for programme execution because they lack capacity themselves (context analysis, programme evaluation, impact assessment on livelihoods);
- Different target groups for evaluation results. Evaluation is not only about accountability toward donors, it also informs beneficiaries and the wider public about programme outcomes;
- Vague and undefined concepts that have become dominant in the discourse for development, such as sustainability, poverty reduction and resilient communities. 'Development' itself is also an ambiguous term.

iii) Only recently have the interviewed organisations realised they need to change their evaluation practices, whereas in the academic community this notion came about in the late '80s / early'90s. An explanation for this phenomenon falls outside the scope of this research, but the interviewed organisations referred to possible underlying causes:

- Monitoring and evaluation is still perceived as an activity imposed and used by donors;
- Bearing the financial crisis in mind, less public and private money is available for development. There is pressure for development organisations to demonstrate impact and innovation;
- For Fair Trade activities specifically, the market has gradually converted from a pioneer to a conventional market over the past few years. Those Fair Trade organisations that have a development agenda need to be able to demonstrate and convince the public and donors that their approach adds more value to the beneficiaries in the South than the approach of the conventional Fair Trade actors.

As stated before, the organisations interviewed for this research formulate their programme activities based on implicit theories/assumptions and none surface these theories for the purpose of evaluation. Only one organisation was found to develop explicit theories for both programme impact well as evaluation design.

In the literature review, a trend was spotted among academics to recognise the complex composition of the environment where development organisations implement their programme activities. They promote the use of theory-driven evaluations in order to learn about programme impact in complex settings (Barnes et al., 2003; Coryn et al., 2010; Sanderson, 2000; Hospes, 2008). Considering the limited use of theories for evaluation by the interviewed organisations, the question if and how these organisations deal with evaluating their impact in a complex environment becomes even more interesting. According to academic reading, there are many factors that can add to the complexity of a programme environment in development. The organisations were asked to elaborate on their experiences in and tactics for dealing with the following components of complexity i) context, ii) wicked-issues, i.e. multiple actors and sectors and iii) long-term attribution of programme outcomes.

In the interview it became clear that the organisations do not perceive modern complexity a challenge for the attribution of programme impact or evaluation. They do tend to adjust their practices to be able to operate more efficiently and effectively in a context in which many actors are active.

All the organisations perform some sort of context analysis when they start a programme. Only one organisation performs context analysis for the purpose of evaluation, comparing it to the baseline context analysis and screen for changes that may be relevant for the programme implementation. Context analysis is seen as a standard practice by the organisations, in order to better understand local policies, needs of local communities, and potential opportunities for cooperation. One organisation also uses the context analysis to assess if a context is conducive to attaining the change they pursue. It is this organisation that has a scaling-up strategy for their activities, once they have found to have a successful formula. The other organisations have not reached the stage of generalising programme outcomes yet, hence they do not use context evaluation for this purpose as is suggested in the literature.

Regarding the recognition of ‘wicked issues’, i.e. social problems that demand interventions from a range of bodies and agencies, and which often seem intractable (Barnes et al., 2003), the interviewees all state that they focus their intervention on their area of expertise (farmers’ ownership in the production value chain and Fair Trade). Again, almost all organisations will team up with other actors in their programmatic region if they feel this will increase the impact of their programmes. While two of the organisations systematically evaluate the fruitfulness of these partnerships, only one organisation actually develops and evaluates impact theories on the outcomes of the partnerships. There is no evidence that the interviewed organisations take an

integral and holistic approach to development or that they are encouraged to do so by other actors (local authorities, donors, UN, etc.), and, as a result, in their approach to evaluation, the phenomenon that multiple actors and sectors are required to support the advancement of communities is not dealt with.

The involvement of a broad range of stakeholders in the definition of the programme objectives, activities and the evaluation plan, in order to create ownership, alignment and support for the project is generally not seen as a strategy to deal with complexity. As described in chapter 2.3.2, “potential stakeholders include: staff, governors and directors, funders, technical assistance providers, residents, local business owners civic leaders, representative of community-based organisations and public, private and non-profit organisations” (Sullivan and Stewart, 2006, pp.181). The interviewees indicated that they seek cooperation with other local actors in order to tackle practical issues. However, no evidence was found that organisations try to create ownership, mutual interests, or supportive attitudes with a wide variety of stakeholders in order to reach agreement on the programme outcomes. Regarding the evaluation approach, again, no evidence was found that the organisations involve other stakeholders than their partners for setting the ground rules for the evaluation or to explore theories or assumptions about expected implementation or the impact of the programme.

The attribution of programme effects after a longer period of time was only pro-actively identified as a serious issue by one organisation. Other organisations confirmed it as a weakness in their evaluation strategy but only after being specifically asked about it. All the organisations indicated they have not found an answer in how to include the long-term effectiveness of their programmes in their evaluation approach.

Besides a very practical and pragmatic approach to complexity, primarily by setting up partnerships in order to increase the impact of their activities, the organisations did not indicate they consider complexity a specific challenge to their programme implementation and evaluation activities. Their notion of complexity cannot be compared to the in-depth analysis in literature. They are not pro-actively looking for evaluation approaches that will help them tackle issues of complexity in finding and measuring their programme impacts.

From the interviews a couple of observations were made that may explain why the notion of complexity is not considered relevant and why complexity is not (yet) seen as an issue for impact attribution in evaluations:

- i) None of the interviewed organisations evaluates long-term impact of their programmes. They have no strategy to monitor their areas of intervention after the intervention is finished;
- ii) Three organisations indicated they have no clear or any exit criteria. They have no objectives or indicators for when they consider the cooperation with their partners a success;
- iii) None of the organisations operates as part of a 'bigger system', recognising that they are part of open or adaptive systems in which the link between activities and outcomes is non-linear (see chapter 2.3.4). Evaluation efforts are limited to activities that are in their expertise of work and stakeholder involvement is limited;
- iv) Overall there is a poor definition (or no definition at all) of key concepts such as poverty, sustainable development, or resilience, etc. These are not simple questions to answer, and require a more inclusive approach as suggested by Parris and Kates (2003). They reviewed 12 efforts by international institutions to characterise and measure sustainable development and still come to the conclusion that there are no indicator sets that are universally accepted, which according to them is "due to the ambiguity of sustainable development, the plurality of purpose in characterizing and measuring sustainable development, and the confusion of terminology, data, and methods and measurement" (pp. 559). Using theory-driven evaluation could support an organisation in fine-tuning such concept definitions; however, it will not solve the lack of a universal agreement on and adoption of their indicators.

As mentioned above one of the key concepts that are poorly defined in literature is the concept of sustainable development itself. It was difficult to find an unambiguous definition of sustainable development and it was established that this lack of clarity considerably adds to the complexity of setting clear programme objectives and defining an appropriate approach for evaluation. In the literature review the following assumption was made: 'it can be expected that the degree of complexity recognised within organisations relates to their interpretation of sustainable development; the thicker their interpretation of sustainable development, the more complex their experience within the programme implementation environment, and vice versa'.

Two interviewed organisations had developed a paper to define sustainability in order to integrate sustainability targets in their programmes. It was obvious they adopted a thick interpretation of sustainability, meaning they included targets for economic and social development, environmental protection and aimed for the probability of continued long-term benefits. Only one of these organisations had conducted a thorough exercise to define the

concept of sustainability in such detail that they were able to perform a 'sustainability screening exercise' in all their current programmes. Only then they really understood what the issues and challenges were in the sectors and value chains they were targeting for change. Consequently they adapted their interventions in order to respond to those challenges. It is this organisation that also uses theory to evaluate the impact of their programmes.

The other three organisations perform evaluation on sustainability aspects to a certain extent. They indicate these evaluation efforts are not systematically conducted and/or do not include all aspects of sustainability; thus it can be concluded that the concept of sustainability remains a challenge to i) define and make concrete (targets and indicators) and/or ii) integrate sustainability targets in the programme and evaluate their effects.

All the organisations struggled with ability to evaluate the impact of their programmes in the long-term. Although all aim for development that is *lasting* and respects the needs of *future generations*, they have not found a way, affordable and realistic, to measure the impact of their specific programmes over a considerable amount of time. Therefore, it remains ambiguous if the selected organisations can objectively claim that their activities in the South trigger, stimulate or lead to the sustainable development of the communities they are serving. It is most likely that their intervention lead to some positive developments for the small scale farmers they target; however, when it comes to demonstrating what impact on development was actually reached in a broader context and if this development led to poverty reduction or other positive outcomes in the longer-term, they fall short. Programme design and evaluation are not adapted yet to delivering proof of these objectives. Moreover, only one organisation had the ambition to use their evaluations to upscale successful interventions and conducted systematic evaluations to understand *why* certain activities were successful or not. The Fair Trade organisations that want to be able to show that their approach and impact on farmers in the South is different from the conventional Fair Trade actors, still have some work to do in order to understand (and reproduce) the impact of their programmes. Since their programmes also aim to change (government, business, institutional) trade policies, evaluations can provide excellent input to promote successful activities and evidence of development in order to support their advocacy work.

## 7 Conclusions

All the interviewed organisations evaluate their programmatic objectives, in some manner, and confirm the importance of conducting appropriate evaluations in order to understand the impact of their interventions in achieving their definition of sustainable development. The extent to which they apply these evaluations (systematically or not, fixed approach or not, using theory or not) varies considerably. Aside from one organisation that had recently undergone an extensive exercise to implement a new evaluation methodology, each organisation reported seeking further development of their approaches to evaluation. While in researching their evaluation approaches, they do not specifically refer to theory-driven evaluations as a solution to dealing with complexity and providing evidence to replicate their work, they do, to some extent, recognise that theory-driven evaluation will ground their planned programmatic objectives and activities. Currently, most of the theories they are utilising are implicit theories, i.e. perceived as the organisations' belief or convictions and are not verified and confirmed through evaluation. Overall, there is limited knowledge of and experience in the use of theory or the use of theory-driven evaluations.

The academic evaluation community and the NGO practitioners do not have much in common when discussing these issues in evaluations. In addition to different use in terminology, the selected NGOs also indicated that they have to deal with a variety of challenges that can seriously impact their evaluation dynamics. First of all donors are key decision makers in the way organisations conduct their evaluations. They dictate the content, the formats and the rules in general. Moreover, it remains unclear how donors contribute to the evaluation process directly aside from dictating the parameters. Donors are not clear about whether development impact indicators are set at the 'macro level' and whether the NGO evaluations form part of a wider system of evaluation in which development is seen as a joint effort by many stakeholders and actors. The interviewed NGOs by all means do not feel they are part of a larger or inclusive system. They have no strategy, policy, or methodology to evaluate overarching objectives such as the reduction of poverty rates or objectives that could establish impact over a longer-term period. Moreover, the NGOs all complained that they struggled to finance their evaluation efforts, implying that donors do not support innovative approaches to evaluation. Secondly, the mind-set amongst programme staff and local stakeholders was raised as an area of concern. Complaints about the enormous amount of time and resources to conduct evaluations are ostensibly wide-spread and the evaluation people that were interviewed indicated that they were not sure how they were going to keep the staff, the partner and the beneficiaries 'on board' if they were to expand their evaluation activities. Demonstrating added value from evaluation efforts was seen as a possible solution to motivate people in the field, but how to do this was

unclear. A final challenge is that the selected NGOs seemed to be entangled in a web of complexity, not of the academic theorised version of complexity, but rather the 'internal' environment of the organisations. For example, they reported struggling with a proliferation of standards, some imposed by institutions, others self-imposed, in the evolution that comprises the modern discourse of development. Some concepts have been embraced by the NGOs, such as sustainability; however their incorporation of this concept, in programme and evaluation design varies considerably from one organisation to another. Other concepts remain vague such as resilience and poverty while others, e.g. improved sustainable livelihoods, are evaluated through a specific evaluation method which requires a different area of expertise. Due to the sustainable development ambition organisations enter an arena of very diverse activities and outcomes; in general organisations solve this challenge by seeking partnerships with other actors. The aforementioned challenges have not been recognised yet by the academic literature.

The complexity that is subject to academic research does not seem relevant to the interviewed organisations with respect to attribution of programme outcomes to their interventions. Contextual circumstances are recognised by the interviewed organisations but not seen as ingredients for programme successes or failures. Rather than monitoring and analysing if contextual changes promote or undermine possible programme outcomes, organisations choose to adapt to the context, seeking pragmatic and practical forms of cooperation to overcome negative contextual factors. Their interpretation of sustainable development has an influence on how much complexity is experienced. Though the organisations all wanted to implement the thick interpretation of sustainability, in reality only one organisation actually managed to define it properly, train all staff and integrate it in its programme and evaluation design. Others are still working on elaborating and rolling-out the concept in its full form.

All in all it could be said that organisations apply a 'tunnel vision' to the evaluation of their impact, i.e. they oversee their own activities and measure their own impact. There is no recognition they are part of a larger system in which contextual changes and actions by other actors can also trigger change or influence their outcomes. Improving people's lives structurally and support them climbing out of poverty is not likely to happen without the support of many other stakeholders and a favourable context. The organisations seem to circumvent having to deal with these 'external' components of complexity (not consciously, but due to the overwhelming nature of their internal environmental complexity); since they do not include the overall goal, i.e. to reduce poverty, in their evaluation efforts, they do not have to 'prove' they contribute positively to a system that develops over time and achieves sustainable development. The interviewed organisations could re-think their role in and contribution to the bigger system,

and make deliberate choices about the actual change they want to achieve considering the constraints they face with their donors, financing, evaluations, staff mentality, and many other internal and external factors that make their work complex. The author is of the opinion that it should not be necessary for all sustainable development organisations to contribute directly to the accomplishment of changes at the (quite chaotic) macro level, especially not for organisations that work at grass-root level and without universally accepted indicators about key-concepts such as poverty. The aspiration to contribute to sustainable development and poverty reduction is surrounded by vagueness and ambiguity and sustainable development organisations should feel free to opt for 'smaller' ambitions that are more concrete and more straightforward to measure. Poverty reduction has become a goal, a popular word in the development discourse, but no one really knows how to measure it. Organisations should dare to formulate their end goal otherwise.

It is important to note that the only organisation interviewed that had performed a systematic review on sustainable development, reported greater understanding of where they could have impact, how their activities affected people's lives, how to monitor and steer progress of the programme's desired outcomes, and how to recognise a successful formula for the purpose of expansion or replication. This is probably the best practical example for promoting theory-driven evaluations as opposed to the purely academic arguments for such evaluations.

## **7.1 Future research**

This thesis provided useful insights in how organisations evaluate the impact of their programme activities in the South and how they deal with the complexity of the environments in which they operate. While certain answers to the proposed questions were identified, this process also instigated many questions not previously foreseen. The dominance of donors for example; they play a major role in setting the development agenda in general and evaluation requirements in specific. Stame's (2004) article about the vertical complexity of multi-level governance states that "the evaluation deficit refers to the unsatisfactory situation in which most evaluations, conducted at local and other sub-national levels, provide the kind of information (on output) that does not immediately inform an analysis of effects and impacts at higher levels, i.e. whether global objectives have been met" (pp. 59). It would be interesting to study the motives of donors and policy makers in pursuing particular outcomes and promoting certain evaluation methods. Sanderson (2000) argued that 'cross-cutting' social problems need to be addressed through 'joined-up' policy initiatives and evaluations need to deliver substantial information that can lead to policy change. Can the work of NGOs be perceived as part of a 'joined-up' development policy? What methods do policy makers (often donors as well) pursue

and promote to evaluate *their* targets for sustainable development and how do NGO targets fit in?

The role of donors and global objectives of development raises questions regarding, the ambition of sustainable development NGOs. Should they seek to be part of a larger system, trying to influence its internal working, dynamics and outcomes or is the impact in the niche they operate the only realistic outcome that can be expected? By desiring a multitude of sustainable outcomes, it almost becomes impossible for an NGO to work alone as they exceed their internal capacity and knowledge. NGOs have entangled themselves in a web of complexity, in their internal environment as well as in their external environment. Are there ways to de-construct this complexity for example by introducing a model for development that is divided into layers of influence (with activities and outcomes on macro, meso and micro levels) and maps actors that can reach change within each layer? As became clear from the literature study, change in a complex environment cannot be covered by one overarching theory, let alone by one 'overarching' actor.

A last observation that deserves further attention is the mind-set around evaluations. The organisations indicated that they struggle to find time and resources for developing and implementing a suitable approach to evaluation, and then to conduct them in a systemic manner. Evaluation is still not perceived by many as an integral part of programme cycle; moreover, and detrimentally, it is perceived as additional work, on top of the already busy work load expected of staff. Would evaluation be more successful if all staff members and stakeholders embraced its outcomes? How can evaluation, from the very start of the programme or even already in the exploration phase, be embedded in the design of the activities? Do we need a different profile for staff? Do we need a different approach to evaluation that is more inclusive and less elaborate? What if we delegated evaluation activities to other actors and not integrate it at all in the programme efforts? Stufflebeam (2001) made a point that when staff responsible for implementing programmes are also performing evaluations off their work, there may be a conflict of interest. These are all considerations that are valuable for further research, during the interviews it became clear that evaluation has not yet found its place and acceptance in the organisations.

\*\*\*\*\*

## Literature

Barnes M., E. Matka and H. Sullivan (2003) 'Evidence, Understanding and Complexity: Evaluation in Non-Linear Systems', *Evaluation*, 2003 9: 265 – 284

Birkmayer J.D. and Weiss, C.H. (2000) 'Theory-Based Evaluation in Practice: What Do We Learn?', *Evaluation Review*, 24: 407 – 430

Blamey A. and MacKenzie, M. (2007) 'Theories of Change and Realistic Evaluation: Peas in a Pod or Apples and Oranges?', *Evaluation*, 2007 13: 439 – 454

Chen H.T. and Rossi, P.H. (1980) 'The Multi-goal, theory-driven Approach to Evaluation: A Model Linking Basic and Applied Social Science', *Social Forces*, Vol. 59, no. 1: 106 – 122

Chen, H.T. (1989) 'Issues in the Theory-Driven Perspective', *Evaluation and Programme Planning*, Vol. 12: 299 – 306

Chen, H.T. (1990) *Theory-Driven Evaluations*, SAGE Publications, Inc., London

Chen, H.T. 'Practical Program Evaluation: Assessing and Improving Planning, Implementation, and Effectiveness', SAGE Publications, Inc., London

Chen, H.T. (2011) 'Theory-Driven Evaluation: Conceptual Framework, Methodology, and Application', in *Evaluation von Programmen und Projekten zur Förderung einer pluralistischen und demokratischen Kultur, Bielefeld 16-18 February 2011*, Proval Gesellschaft für Sozialwissenschaftliche Analyse, downloaded at [http://www.proval-services.eu/cms1/downloads.php?cat\\_id=10](http://www.proval-services.eu/cms1/downloads.php?cat_id=10), viewed at 24 September 2012

Coryn C.L.S., L. A. Noakes, C.D. Westine, D.C. Scröter (2010) 'A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009', *American Journal of Evaluation*, 2011 32: 199, <http://aje.sagepub.com/content/32/2/199>

Depuis Maurin, M. (2011), 'Etude comparee avec d'autres labels ethiques', *Plate-Forme pour le Commerce Equitable*, [http://www.commerceequitable.org/images/pdf/garanties/guide\\_labels\\_ce\\_2011.pdf](http://www.commerceequitable.org/images/pdf/garanties/guide_labels_ce_2011.pdf)

Drexhage J. and Murphy, D (2010), International Institute for Sustainable Development 'Sustainable Development – From Brundtland to Rio 2012', United Nations Headquarters, New York, September 2010

Earth Summit2012 (2012), Stakeholder Forum, <http://www.earthsummit2012.org>, viewed at 20 August 2012

Eyben R., T. Kiddler, J. Rowlands, A. Bronstein (2008) 'Thinking about Change for Development Practice: A Case Study from Oxfam GB', *Development in Practice*, 18:2, 201 – 212

European Council (2006), Review of the EU Sustainable Development Strategy (EU SDS) – Renewed Strategy -, 10917/06, 26 June 2006

European Fair Trade Association (2011) *EFTA Fair Trade Assessment Guidelines*, Schin op Geul

Fair Trade Advocacy Office (2010), Fair Trade Advocacy Office, <http://www.fairtrade-advocacy.org>, viewed at 20 August 2012

FAO - Food and Agriculture Organisation of the United Nations (2013), Fisheries and Aquaculture department, <http://www.fao.org/fishery/topic/13297/en>, viewed at 10 October 2012

Francis, P. (2001) 'Integrated Impact Assessment for Sustainable Development: A case study approach', *World Development*, vol. 29, No.6, pp. 1011-1024

Friedland, W. H. (2010), 'Hutchens, Anna: Changing Big Business: The Globalisation of the Fair Trade Movement', *Journal of Agricultural and Environmental Ethics*, 24: 71 – 74, [Online] DOI 10.1007/s10806-010-9238-6 (Accessed: 11 August 2012)

Green (2012), 'Oxfam', <http://www.oxfamblogs.org/fp2p/?p=8551>, viewed at 20 August 2012

Green J. and Thorogood, N. (2009) '*Qualitative Methods for Health Research*', 2<sup>nd</sup> edition. SAGE Publications, Inc., London

Hammersley M. And Atkinson, P. (2007) '*Ethnography*', Principles in Practice third edition, Routledge, Oxon

Holm Pederson, L. and Rieper O. (2008) 'Is Realist Evaluation a Realistic Approach for Complex Reforms?', *Evaluation*, 2008 14: 271, <http://evi.sagepub.com/content/14/3/271>

Hospes, O. (2008) 'Evaluation evolution?' *The Broker Online*,  
<http://www.thebrokeronline.eu/Articles/Evaluation-evolution> (Accessed: 28 December 2011)

Intermon Oxfam (2013), Intermon Oxfam, <http://www.intermonoxfam.org/es/quienes-somos/presentacion/mision-vision-valores>, viewed on 4 November 2012

Johnson, R.B. (1998) 'Toward a Theoretical Model of Evaluation Utilization', *Evaluation and Program Planning*, 21: 93 – 110

Lee K.S. and Chavis, D.M. (2011) 'Cross-case Methodology: Bringing Rigour to Community and Systems Change Research and Evaluation', *Journal of Community & Applied Social Psychology*, [Online] DOI 10.1002/casp.1131

Nooteboom, S. (2007) 'Impact Assessment Procedures for Sustainable Development: A Complexity Theory Perspective', *Environment Impact Assessment Review*, 27, 645 – 665

Parris T.M. and Kates, R.W. (2003) 'Characterizing and Measuring Sustainable Development', *Annu. Rev. Environ. Resour.*, 28: 559 - 86

OECD – Organisation for Economic Cooperation and Development (2010) 'Glossary of Key Terms in Evaluation and Results Based Management', <http://www.oecd.org/development/peer-reviews/2754804.pdf>, viewed on 11 October 2012

Oxfam-Wereldwinkels (2013), Oxfam-Wereldwinkels,  
<http://www.oxfamwereldwinkels.be/missie>, viewed on 5 November 2012

Pawson R., T. Greenhalgh, G. Harvey, K. Walshe (2005) 'Realsit Review – A New Method of Systematic Review Designed for Complex Policy Interventions', *Journal of Health Services Research & Policy*, Vol. 10 suppl. 1: 21 – 34

Pawson R. and Tilley, N. (2004) 'Realist Evaluation', *Paper published for the British Cabinet Office*, based on Pawson R. And Tilley, N. 'Realistic Evaluation', Sage, 1997

Rogers, P.J., A. Petrosino, T.A. Huebner, T.A. Hacsí (2000), Program Theory Evaluation: Practice, Promise and Problems, *New Directions for Evaluation*, no. 87: 5 – 13

Sanderson, I. (2000) 'Evaluation in Complex Policy Systems', *Evaluation* 6: 433 - 454

Sanderson, I. (2002) 'Evaluation, Policy Learning and Evidence-based Policy Making', *Public Administration*, Vol. 80 No.1: 1 – 22

Scriven, M. (1996) 'Types of Evaluation and Types of Evaluator', *Evaluation Practice*, Vol. 17, No. 2: 151 – 161

Scriven, M. (1998) 'Minimalist Theory: The Least Theory That Practice Requires', *American Journal of Evaluation*, 19: 57 – 70

Stame, N. (2004) 'Theory-Based Evaluation and Types of Complexity', *Evaluation*, 2004 10: 58, <http://evi.sagepub.com/content/10/1/58>

Stufflebeam, D. L. (2001) 'Evaluation Models', *New Directions for Evaluations*, no. 89: 7 – 98

Sullivan H. and Stewart, M. (2006) 'Who owns the Theory of Change?', *Evaluation*, 12: 179 – 198

Tack S. and van der Veer, T. (2011) 'Fair Trade Strategy', paper presented to Oxfam International Executive Directors meeting 27 – 29 September 2011, Paris

Traidcraft (2013), Traidcraft, <http://www.traidcraft.co.uk/about-traidcraft>, viewed on 4 November 2012

Patton, M.Q. (2000) 'Utilization-focused Evaluation', Evaluation Models, *Evaluation in Education and Human Services*, Vol. 49, V, 425 – 438

Trochim, W.M. (1998) 'An Evaluation of Michael Scriven's "Minimalist Theory: The Least Theory that Practice Requires"', *American Journal of Evaluation*, Vol. 19, No. 2: 243 – 249

United Nations (1987) 'Report of the World Commission on Environment and Development: Our Common Future', Transmitted to the General Assembly as an Annex to document A/42/427

United Nations (1992) 'Report of the United Nations Conference on Environment and Development', United Nations General Assembly, A/CONF.151/26 (Vol. I)

UNDP – United Nations Development Programme (2011) 'Human Development Report 2011 Sustainability and Equity – A Better Future for All', United Nations Development Programme, New York

United Nations (2012), 'United Nations Millennium Development Goals', <http://www.un.org/millenniumgoals/bkgd.shtml>, viewed at 20 August 2012

Vagneron, I. and Roquigny, S. (2010) 'What do we really know about the impact of fair trade? A Synthesis', Paris: PFCE, 25 p.

Vredeseilanden (2013), Vredeseilanden, <http://www.vredeseilanden.be/over-vredeseilanden/onze-missie>, viewed at 4 November 2012.

Weiss, C.H. (1995) 'Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community-Based Initiatives for Children and Families', in J.P. Connell, A.C. Kubisch, L.B. Schorr and C.H. Weiss (eds) *New Approaches to Evaluating Community Initiatives, vol. 1, Concepts, Methods and Contexts*, Aspen Institute: Washington, D.C.

Weiss, C.H. (1997) 'How Can Theory-Based Evaluation Make Greater Headway?', *Evaluation Review*, 21: 501 – 524

Weiss, C.H. (1998) 'Analyzing and Interpreting the Data', in: C.H. Weiss, *Evaluation – Methods for Studying Programs and Policies*, 2<sup>nd</sup> edition. Prentice Hall: New Jersey 12: 271 – 290

\*\*\*\*\*

## Appendix A – Interview questions

### (Q1) MAIN QUESTION

**How would you characterise the evaluation approach you currently use?**

- Has your organisation adopted a specific evaluation method or model?  
If so, which?
- With your evaluations, do you establish:
  - o If the programme worked? (conventional approach). If so, how?
  - o Why the programme worked? (theory-driven approach). If so, how?
  - o Both. If so, how?
- Do you use theory as a base for evaluation?
- Do you develop theory based on:
  - o How change happens - mechanisms? (realistic evaluation approach)  
Can you elaborate how you construct the theory?
  - o How your organisation makes this change happen - actions? (theory of change evaluation approach)  
Can you elaborate on how you construct the theory?
  - o Both  
Can you elaborate on how you construct the theory?
- Who provides input for the development of evaluation theory:
  - o Evaluator
  - o Programme staff. If so, which ones?
  - o Stakeholders. If so, which ones?
  - o Other

### (Q2) SUB-QUESTION 1

***In your evaluation approach, (how) do you deal with the effects of:***

- A) Specific circumstances related to the context you are working in
  - o Do you perform prior (social) research?
  - o Do you use documentation / studies?
  - o Do you consult local stakeholders? Which ones?
  - o Do you use knowledge based upon experience from your organisation?
  - o Other?
- B) Multiple actors trying to achieve change within the same environment / within the same communities as your organisation

- Do you cooperate, besides your partners, with other local, regional, international actors? Which ones?
  - Can you describe this cooperation and its objectives?
  - Do you evaluate this cooperation?
- C) Outcomes that require change within or commitment from multiple sectors
- Do you aim to achieve and evaluate objectives that cover multiple sectors?
  - If so, which sectors and how do you deal with change that needs support / input from other sectors than your own?
  - Do you evaluate this support / input from other sectors?
  - Do you evaluate possible side-effects of your programmes?
- D) Attribution of your programme activities to long-term change / impact
- Do you evaluate initial outcomes (changes in individual knowledge, skills, and abilities)? If so, how?
- Do you evaluate intermediate outcomes (changes in behaviour)? If so, how?
  - Do you evaluate long-term outcomes (alleviation, reduction, prevention of specific social problems or meeting the needs of the target population)? If so, how?

***(Q3) SUB-QUESTION 2***

***In your evaluation approach, (how) do you characterise and evaluate the level of sustainability of your programmes?***

- How would you define 'sustainability' within in the context of your programmes in the South?
- Do you base your sustainable programme goals on the principles or guidelines of a certain authority/ institution?
  - Brundtland report (UN 1997)
  - UN World Summit on Sustainable Development (Rio de Janeiro 1992)
  - Other
- What sustainable programme impact does your organisation aim to reach?
  - Long-term objectives? If so, for what period of time and (how) do you evaluate them?
  - Ecological objectives? Which ones and (how) do you evaluate them?
  - Economic objectives? If so, which ones and (how) do you evaluate them?
  - Social welfare objectives? If so, which ones and (how) do you evaluate them?
- Which of your programmes do you evaluate on its sustainable results?

- Fair Trade / Alternative livelihood programmes
- Development programmes
- Advocacy programmes
- Other programmes
- All programmes

### **REFLECTIVE QUESTIONS**

(Q4) – Can you elaborate on the strengths and weaknesses of the approach you use?

(Q5) – According to you, what contribution / added value do the evaluations your organisation performs deliver to the future working of your organisation, and for whom?

(Q6) – Is your organisation pleased with the current approach? If not can you explain which challenges remain to overcome in the area of programme evaluation?

\*\*\*\*\*

## Appendix B – Methodological protocol

Step	Activity	Sources	Learning outcome	Timing	Duration (est.)
1	Examine mission	Internet	Profile of the organisation	Data collection, prior to the interview	1 hour
2	Examine strategic plan, including MEL	Documentation > strategic plan; Internet	Change goals, working methods, partnership working, including MEL strategy and purpose	Data collection, prior to the interview	3 hours
3	Examine sustainable development programme(s) in the South	Documentation > representative example of a programme proposal, including logic framework	Intervention logic, programme activities, expected outputs and outcomes (short-, medium- and long-term)	Data collection, prior to the interview	3 hours
4	Examine evaluation plan(s)	Documentation > representative example of an evaluation design, proposal, terms of reference or otherwise	Approach used for evaluation (method, qualitative/quantitative data collection, use of theory, timing > ex ante, midterm, ex post), purpose of evaluation	Data collection, prior to the interview	2 hours
5	Examine evaluation report(s)	Documentation > representative example of an evaluation report	Outcome of evaluation, recommendations	Data collection, prior to the interview	2 hours
6	Perform interview	MEL responsible/ programme manager/ evaluator	How is the evaluation plan constructed, whose input is required, how is theory used, what is done with the evaluation results, etc. > see Interview Sheet	Data collection	2 hours + 3 hours travel time = 5 hours
7	Transcribe interview	-	Correct recording and reflecting on interviewees' answers, to be annexed to final thesis report	Data processing, after the interview	8 hours
8	Analyse and process data acquired	Interview transcriptions and information acquired during steps 1 to 5	Answer to research question 2: <i>What are practices and experiences from organisations with evaluation methods for sustainable development or Fair Trade programmes</i>	Data processing, after the interview	16 hours
<b>Total estimated time to prepare for, undertake, and process an interview</b>					<b>40 hours,</b> or 5 working days of 8 hours

\*\*\*\*\*