

**Effect measurement of
Sustainable Supply Chain Governance - Systems:**

**The development and test of
a valid, reliable and manageable measurement system**

Lieneke Wieringa, 3234959

Master thesis, 45 ECTS

Supervisors: Walter Vermeulen, Carel Dieperink (second reader)

MSc in Sustainable Development – Environmental Policy and Management
Utrecht University, the Netherlands
June, 2010

SUMMARY

Since the year 2000 a rapid rise in interest for addressing sustainability issues is visible within supply chains. In the second generation of certification initiatives market and non-market actors cooperate in Sustainable Supply Chain Governance-systems (SSCG-systems) to realize sustainable production throughout the value chain. In order to control for the behaviour of different actors within these SSCG-systems control mechanisms are in place. Their focus is in general on controlling firm behaviour and the organisation of their productive activities (1st order effects). Direct ecological and socio-economic (2nd order effects) and structural and side effects (3rd order effects) hardly seem to be taken into account. Therefore, the focus in this thesis is on the development of a measurement system in order to measure these 2nd and 3rd order effects. The goal of the measurement system is to be valid, reliable and manageable. The characteristics that such a measurement system should possess are presented. Based on a literature study into scientific research approaches and lessons from practice the measurement system was created. Subsequently, it was tested on organic cotton certification in Paraguay. Based on these results the characteristics that a valid, reliable and manageable measurement system should have are as following.

Construct validity is in part realized through a strong theoretical foundation of the concepts used and multiple questions. Besides, profound testing and adjusting of the questionnaire to local perceptions and crosschecking during the interview further address this aspect. The characteristics to address internal validity are respondent validation, reflexive and rigour practice by the researcher, triangulation via key-person interviews and documented data, and the use of a target and control groups, before and after questions and causal explanation questions. External validity should be addressed via random sampling and thick description of contextual factors. Reliability is addressed via a well documented, consistent, systematic and logical research design in which cross-checking plays an important role. Besides, the data should be analyzed before filing into SPSS. The issue of manageability can be addressed by making use of likert-scale questions. The use of this type of questions makes the questionnaire easy to conduct, non-time consuming, offers the opportunity of asking many questions and clear representation because of the quantitative data acquired. Furthermore, structure is provided to the research through providing a six step research schedule. Nonetheless, the positive test-results from Paraguay the mayor shortcoming of the measurement system is the lack of statistical significance. Addressing this issue by increasing the sample size will have its negative impact on the manageability of the system. If the system remains manageable enough and how to exactly deal with this issue is material for future research.

ACKNOWLEDGEMENTS

The process of writing this thesis has been a huge challenge which I would not have been able to overcome without the help of some people. First of all I would like to thank Jeroen de Vries and Simon Ferrigno who were of great help for organizing my fieldwork-trip to Paraguay. My special thanks go out to Hebe Gonzalez, Mario Paredes, Wilfrido Ortíz and Lorenzo Zarza and all the other staff from Alter Vida whose help has been incredible and who have made my fieldwork possible. I would also like to thank all the respondents that were willing to participate in my research and received me in such a friendly manner. Furthermore, I would like to thank Tete for being such a great friend and making my stay in Paraguay also pleasant from a social perspective.

In Holland my supervisor Walter Vermeulen was there to assist me in the writing process and keeping the overall picture in mind. Many thanks go out to my friends Maaïke, Suzanne, Oscar and Kirsten who endured me, helped me to stay motivated and sometimes provided me with their critical insights. Special thanks go to Inda who has motivated and supported me throughout the whole process and especially during our “library-times”. And last but definitely not least I want to thank my brothers and especially my parents who have given me tremendous support and have stood up to all my complaints with incredible patience.

TABLE OF CONTENTS

1. INTRODUCTION: THE NEED FOR DEVELOPING A VALID, RELIABLE AND MANAGEABLE MEASUREMENT SYSTEM.....	1
1.1 Introduction.....	1
1.2 Research objective and relevance	2
1.3 Central research question.....	3
1.4 Research strategy.....	3
1.5 Thesis outline.....	3
2. EMERGING SSCG-SYSTEMS AND MEASURING THEIR IMPACTS.....	4
2.1 Supply chains and value chains.....	4
2.2 SSCG-systems	4
2.3 Goals (claims)	5
2.4 Monitoring systems	6
2.5 Challenges in designing a measurement system for SSCG-systems	7
2.5.1 General challenges in a developing world context.....	7
2.5.2 Specific challenges for SSCG-systems.....	7
2.5.3 Challenges for measurement of effects	7
3. THEORETICAL FOUNDATION: REVIEWING RESEARCH METHODS FOR IMPACT MEASUREMENT.....	8
3.1 Proving versus improving.....	8
3.2 Different methods	8
3.2.1 Quantitative research methods	8
3.2.2 Qualitative research methods	10
3.2.3 Participatory appraisal approaches.....	11
3.3 Finding a balance.....	12
4. PRACTICAL FOUNDATION: REVIEWING PRACTICES OF MEASURING IMPACTS IN DEVELOPING COUNTRIES.....	15
4.1 RISE	15
4.2 The COSA Project.....	16
4.3 Impact evaluation of the NFPP on rural household livelihoods.....	17
4.4 Livelihood Change in a Changing Natural Resource Management Context	18
4.5 Integrative Framework for assessing the impact of fair trade coffee	18
4.6 Organic Exchange	19

4.7 ISEAL Alliance	20
4.8 Summary	21
5. THE MEASUREMENT SYSTEM DESIGN: BALANCED IMPACT MEASUREMENT FOR SSCG-SYSTEMS	22
5.1 Addressing the measurement challenges	22
5.2 Conducting the measurement	25
6. SUSTAINABILITY ASPECTS FRAMEWORK	29
6.1 Theoretical foundation	30
6.1.1 Social pillar and Economic pillar	31
6.1.2 Environmental pillar	34
6.2 Comparison	36
6.2.1. Comparison economic pillar and social pillar	36
6.2.2. Comparison environmental pillar	38
7. ORGANIC COTTON CERTIFICATION	41
7.1. The cotton chain	41
7. 2 Indicator framework for the cotton chain.....	46
7.2.1. Environmental impacts	47
7.2.2. Social and economic impacts:	48
7.3. Final indicator framework for the cotton chain	49
7.4 The aspects used in the test-case	51
8. TEST RESULTS: EFFECTS FROM CERTIFICATION ON ORGANIC COTTON PRODUCERS.....	52
8.1 Organic cotton certification in Paraguay: Aratex.....	52
8.2 Research method for conducting the test in Paraguay	53
8.3 Main conclusions.....	56
8.4 Detailed Conclusions	58
8.4.1 Social and economic aspects	58
8.4.2. Environment	68
9. TEST RESULTS: METHODOLOGICAL CONSIDERATIONS TOWARDS THE BALANCED MEASUREMENT SYSTEM	74
9.1 Validity.....	74
9.2 Reliability.....	75
9.3 Manageability	76

9.4 System specific considerations	78
9.5 Question specific lessons.....	79
10. CONCLUSION & DISCUSSION	80
10.1 Conclusion	80
10.2 Discussion.....	81
REFERENCES	83
ANNEXES	88

FIGURES AND TABLES

<i>Number and name</i>	<i>Page</i>
Fig. 1 Modes of monitoring and orders of effects in SSCG-systems	6
Fig. 2 Chambers three connected circles	11
Fig. 3 Three ring Circus	30
Fig. 4 Russian Dolls	30
Fig. 5 Maslow's Hierarchy of Needs	32
Fig. 6 Daly's Triangle	32
Fig. 7 Indicator framework based on SLA	34
Fig. 8 Environmental LCA indicators	35
Fig. 9 UNCSD themes & Final themes	35
Fig. 10 Social aspect framework	39
Fig. 11 Economic aspect framework	39
Fig. 12 Environmental aspect framework	40
Fig. 13 Value Chain Map Organic Cotton	41
Fig. 14 World cotton production, by main countries, 1980/81 - 2012/13	42
Fig. 15 Cotton consumption, by main countries, 1980/81-2012/13	43
Fig. 16 Long-term price developments for cotton and world cotton stocks	44
Fig. 17 Monthly Average Cotton Prices (Sep '07-Mar '10)	44
Fig. 18 Environmental aspects for the cotton chain	50
Fig. 19 Social aspects for the cotton chain	50
Fig. 20 Economic aspects for the cotton chain	51
Fig. 21 Sustainability Polygon – Social and Economic effects	56
Fig. 22 Sustainability Polygon- Effects on the environment	57
Fig. 23 Production capacities and knowledge	58
Fig. 24 Cotton income	60
Fig. 25 Lack of food	61
Fig. 26 Investments within committee	64
Fig. 27 Mean equity	65
Fig. 28 Frequency of pesticide use	69
Fig. 29 Use of green manure	70
Fig. 30 Positive health change	72
Fig. 31 Time schedule for conducting the research	77
Table 1 Balance in achieving validity and reliability	14
Table 2 Manageability characteristics	23
Table 3 Validity characteristics	24
Table 4 Reliability characteristics	24
Table 5 Six steps research schedule	25
Table 6 Surface, production volume and yields	45
Table 7 Price relation of cotton between de international market and domestic market	46
Table 8 Six steps research schedule	54
Table 9 Change in mean scores of social and economic aspects 06-09	57
Table 10 Change in earnings	59
Table 11 Price changes - per kg of cotton in Guaranies	61
Table 12 Biodiversity measures – Change in means	72
Table 13 Final validity characteristics	75
Table 14 Final reliability characteristics	75
Table 15 Final manageability characteristics	76

ABBREVIATIONS

BTC	Belgian Development Cooperation Agency
COSA	Committee On Sustainability Assessment
ICS	Internal Control System
IISD	International Institute for Sustainable Development
KPI	Key Performance Indicator
LCA	Life Cycle Assessment
NGO	Non Governmental Organisation
NFPP	Natural Forest Protection Program
NORAD	Norwegian Agency for Development Cooperation
OECD	Organisation for Economic Cooperation and Development
PLA	Participatory Learning and Action
PRA	Participatory Rural Appraisal
PSM	Propensity Score Matching
RISE	Response Inducing Sustainability Evaluation
RRA	Rapid Rural Appraisal
SSCG-system	Sustainable Supply Chain Governance – system
SLA	Sustainable Livelihoods Approach
UN	United Nations
UNCSD	United Nations Commission for Sustainable Development
USAID	United States Agency for International Development

1. Introduction: the need for developing a valid, reliable and manageable measurement system

1.1 Introduction

Over the last decade we have witnessed a rapid rise of voluntary certification and labelling initiatives addressing sustainable practices. Ever more companies show their interest in addressing sustainability within the supply chain (Labuschagne et al., 2005; Raynolds et al., 2007; Cramer, 2008; Vermeulen, 2008). Different internal and external pressures have caused this rise to occur. These pressures are coming from civil society, (international) governments as well as from within businesses themselves.

On the one hand global liberalisation has led to declining state regulation. Consequently, organizations try to fill this “regulatory vacuum” (Raynolds et al., 2007; Vermeulen, 2008). On the other hand are international institutions, governments, consumers, NGOs and management and employees within companies pressing organizations to behave in a more sustainable manner. Consumers become more knowledgeable and understand that not only the end product, but also the production chain is important. Organizations are held responsible for the sustainability of production of their suppliers and partners, both by governments and civil society (Raynolds et al., 2007; Cramer, 2008; Vermeulen, 2008). Finally, organizations move to more sustainable production processes out of economic incentives. Organizations use certification to acquire market shares and to enhance their control on the production chain for example by ensuring safety or quality standards (Cramer, 2007; Raynolds et al., 2008). Furthermore, production costs can be reduced and their reputation positively influenced, creating trust amongst their stakeholders and consumers. Finally, it can offer new market opportunities by staying ahead of global trends (WRI, 2008).

The first generation of certification initiatives, such as Fair Trade/ Max Havelaar, emerged in the 1980's and was often initiated by NGOs. After the year 2000 visible interest in addressing sustainability issues rose rapidly. This caused mainstream market suppliers to develop their own supplier control systems. These mainstream market firms are able to capture larger market shares compared to first generation certification initiatives. In the second generation of certification initiatives various market and non-market actors cooperate to achieve sustainable production practices throughout their value chain. These systems can be called ‘Sustainable Supply Chain Governance systems’ (SSCG-systems) (Vermeulen, 2008: 1). The SSCG- systems encompass both collective value creation and sustainable development. The aim is to create a sustainable system in which every actor in the chain will have to contribute in order to achieve sustainable production.

Three different types of SSCG-systems are prevailing; single firm approaches (1st generation), joint producer sector approaches (2nd generation) and cross sectoral approaches (3rd generation), all having their own control mechanisms and specific advantages and drawbacks. Nonetheless their differences, all three different approaches have in common that they rely on the basic assumption that business-to-business supply chain cooperation can realize more social and environmental sustainability (Vermeulen, 2008). However, this causality is not necessarily evident. Different positive as well as negative effects can occur.

In order to control for the behaviour of the different actors in the supply chain and assure realisation of the sustainable development goals different control mechanisms are constructed. These different control mechanisms are self-auditing, 3rd party audits conducted by an NGO or auditing agency and effectiveness evaluation also conducted through 3rd party audits. Each control mechanism measures different effects.

The question however is, what is the general focus of the control mechanisms currently in place. In general the focus seems to be on controlling firm behaviour and the organisation of its productive activities is measured, also called 1st order effects. Direct ecological and socio-economic impacts (2nd order effects) and structural and side effects (3rd order effects) hardly seem to be taken into account. Consequently, throughout the years 2nd and especially 3rd order effects are presumed but hardly ever actually measured (Labuschagne, 2003; Bitzer et al., 2008).

The focus of the control mechanisms currently in place is on analyzing the functioning and appropriateness of the systems and not the actual impacts that the systems have on realizing sustainable development (Vermeulen, 2008). Some indicator-sets are created in order to determine which issues should be covered by SSCG-systems (Kates et al., 2005; Labuschagne et al., 2003; Labuschagne et al., 2005; Piorr, 2003; Schmidt et al., 2004). However, methods to measure these indicators remain unspecified. In their attempt to create an indicator-set Labuschagne et al. (2005: 378) notice that there is currently “...no consensus on a consistent methodology to measure these [environmental and socio-economic] causes or effects”. Consequently, the actual success of SSCG-systems in achieving their sustainability goals cannot be measured and proven and thus remains debated.

In the non-academic world attempts have been undertaken to create a monitoring system to measure these 2nd and 3rd order effects. However, to isolate 2nd and 3rd order effects is a complicated task

due to methodological complications. Different contextual factors can influence the outcome such as changing national or local policies. Therefore, creating an academically underpinned methodology proves difficult, especially for NGOs that are less familiar with academic terms (Organic Exchange, 2008).

In this research a monitoring system capable of measuring 2nd and 3rd order effects of SSCG-Systems is developed and tested on organic cotton production in Paraguay. The aim of the monitoring system was not only to be academically sound and therefore valid and reliable but also applicable in practice. In this manner, its use lies in the academic field as well as in practice where it can play a role as control mechanism for SSCG-Systems or as impact measurement systems at NGOs.

The objective is accomplished by conducting a literature review of different existing research methods and techniques in order to create a theoretical foundation for the monitoring system and analyzing already existing monitoring systems or attempts to create one to gain some experience from practice.

Amongst these already existing monitoring systems is the Organic Exchange monitoring system. Organic Exchange is a NGO focused on creating a sustainable cotton chain. With their monitoring system they try to determine the successes and drawbacks for farmers participating in an organic certification scheme. Their monitoring system has survived multiple trials in different nations, including Turkey, Senegal and India. Being one of the few organisations utilizing a 2nd and 3rd order effect monitoring system, accessing their information and experiences offers a unique opportunity to create a more practice based, thorough and sound monitoring system.

In the literature review different perspectives on conducting research are presented. These range from quantitative modes of conducting research with a highly academic background to more practice based Participatory Appraisal Approaches. The existing debate whether quantitative or qualitative methods for conducting research should be the way forward will be shortly touched upon and a synthesis within that debate in the light of this research is proposed.

1.2 Research objective and relevance

The objective of this research was to develop a theoretically and practically based valid, reliable and manageable monitoring system in order to measure 2nd and 3rd order effects of Sustainable Supply Chain Governance – Systems.

The value of creating a good monitoring system focused at measuring 2nd and 3rd order effects lies in different aspects. The knowledge obtained in the process of developing a monitoring system as well as the system itself and the knowledge which the use of it can produce has a great value for practice. Nevertheless, it also has great academic value, presented further on in the text.

Obviously being able to measure 2nd and 3rd order effects can assist in gaining insight in the actual impacts of SSCG-systems on the local social, economic and environmental situation of the target group. Besides, knowledge on the actual impacts also provides insights on the effectiveness of the system and sheds lights on its strengths and weaknesses.

Another value lies in the improved reliability, transparency and credibility provided through creating monitoring systems. An insight in the effectiveness of SSCG-systems positively influences the reliability of the systems which until now has not been questioned enough (Albersmeier et al., 2009). Raynolds et al. (2007) state that the second key factor that creates legitimacy of certification, next to moral authority by the initiating NGO, is the credibility of monitoring. In order to create credibility, reliability is a necessary aspect. Reliability in its turn can be strengthened through transparency. Knowledge on the actual impacts and effectiveness of SSCG-systems can positively influence transparency and along with that stakeholder- as well as consumer confidence. Moreover, transparency is a necessary tool to gain the confidence of all the different internal and external stakeholders for the companies' participation in a SSCG-system (Cramer, 2008). With adequate monitoring systems in place consumers will be capable of demanding monitoring on all orders of effects which in its turn can improve their confidence (Muradian et al., 2005). Transparency positively influences the reliability of a system.

Transparency also allows for comparison between different SSCG-systems. Stakeholders will be better capable of making a funded choice in which SSCG-system to participate or which products to buy. Moreover, this better informed position puts more pressure on the performance of the different SSCG-systems which is hoped to positively influence their results.

From an academic point of view the value in developing the monitoring system is found in the contribution it can make to answering questions concerning the effectiveness of the move in governance from state to market as posed by Raynolds et al. (2007). Are market actors capable of fulfilling a role as social actor, previously performed by the state? Furthermore, the in-depth study on research methods and its translation to a workable monitoring system can provide interesting insights and might shed light on the continued debate between quantitative and qualitative research perspectives.

1.3 Central research question

Determined by the research objective the following central research question is presented:

What characteristics should a monitoring system have in order to measure 2nd and 3rd order effects of Sustainable Supply Chain Governance-systems, functioning in a developing context, in a valid, reliable and manageable way?

1.4 Research strategy

In order to answer the research question as presented different steps were taken. The research consists of several parts.

Part one was the construction of the measurement system. First, literature on measuring methods was analyzed. The most important research methods are discussed and presented. This includes quantitative, qualitative and participatory appraisal methods. Next, different previously existing measurement systems were analyzed. Issues specific to measuring effects in SSCG-systems were used as a guideline to conduct this analysis. Based on this information from theory and practice an approach was determined for the measurement system. The systems specific issues as well as validity, reliability and manageability were addressed in the best possible manner. Besides the theoretical foundation a practical guideline is also provided along with the measurement system.

Part two entailed the depiction of the indicators to be included in the measurement system. To depict the different indicators a literature study was conducted. The indicators that were selected from this literature review were consequently compared to indicators present in already existing indicator frameworks, including that of Organic Exchange. To adjust the indicators to the specific case of organic cotton a literature study was conducted on frequently occurring social, economic and environmental impacts of cotton production. Finally, the indicators were presented to five key persons in the organic cotton field. Based on their opinions the final indicators were selected.

A third part of the research consisted of analyzing the results from the tests. Both the actual results as well as the functioning of the measurement system were analyzed and scrutinized and possible adjustments suggested.

1.5 Thesis outline

In chapter two insights are presented in the functioning of SSCG-systems. Their goals will be explained and the functioning of monitoring systems is put forward in more detail. Moreover, the different challenges for measuring effects in SSCG-systems are presented. In chapters three and four a theoretical and practical foundation for the measurement is described. Theories on conducting quantitative, qualitative and participatory research are discussed and several examples from practice are analyzed. Based on this theory and the lessons learned in practice the manner of constructing the measurement system is proposed in chapter five. In chapter six the theoretical foundation for the indicator selection is presented from which a sustainability indicator framework is created. A cotton specific indicator framework is also put forward together with some general background information on the cotton chain in chapter 7. In chapter eight the results of the trial with the monitoring system in Paraguay are presented. The lessons learned during and from the trial are discussed and final adjustments are proposed. Finally, chapter nine contains the final conclusions, some discussion and propositions for further research.

2. Emerging SSCG-systems and measuring their impacts

2.1 Supply chains and value chains

When discussing the concept of supply chains it is important to also understand the concept of value chains, as there are two different perspectives of describing extended businesses. A value chain is "the whole series of activities business firms undertake to convert the raw materials or input resources to the goods and services required by a customer" (Porter, 1985). In value chains the primary focus is from the perspective of the customer. Value chains are concerned with what the customer is willing to pay for a product (Feller et al., 2006).

The term supply chain management emerged in the 1980's. A supply chain consists of at least three companies that are linked by an upstream flow of products, services, finances and information (Mentzer, 2002). "Supply chain is a term 'now commonly used internationally– to encompass every effort involved in producing and delivering a final product or service, from the supplier's supplier to the customer's customer'"(Feller et al., 2006: 4). The primary focus of supply chains is on the costs and efficiency of supply, contrary to value chains.

Value chains and supply chains can thus be seen as a different perspective towards the same chain of products. In a value chain the perspective is from the customer down to the supplier while to the supply chain the opposite perspective pertains. Therefore, one does not exclude the other.

In the case of SSCG-system the two perspectives also come to the forefront. On the one hand there is the goal assuring sustainable production of a product as a value for the supplier, while on the other hand creating a sustainable product can raise the value of a product as perceived by the customer. Thus, both perspectives can be considered important.

With regard to this research, the two concepts and their meaning will not further be discussed since it does not directly influence the research results because of the focus on the supply side. Nonetheless, it is important to have a little understanding of value and supply chain in order to better understand the basis of SSCG-systems.

2.2 SSCG-systems

SSCG-systems are a cooperative effort of various market and non-market actors aimed at achieving sustainable production practices throughout the value chain. Their objective is both collective value creation and sustainable development. Their aim is to create a sustainable system in which every actor in the chain will have to contribute in order to achieve sustainable production (Vermeulen, 2008).

The shape of a supply chain can differ per product and the same is true for SSCG-systems. Vermeulen (2008) distinguishes three types of SSCG-systems; single firm approaches (1st generation), joint producer sector approaches (2nd generation) and cross-sectoral approaches (3rd generation). In single firm approaches individual firms are the innovators in creating more socially and environmentally healthy supply chains. They analyze the areas of improvement, create feasible manners to realize these and control compliance to implementation of these improvements. This single firm approach can create first mover advantages for the initiating firms. However there is also the disadvantage of low level reliability of business-to-business self control systems. The joint producer sector approaches are mainly initiated by third parties representing state, market and civil society organisations. The state only plays a marginal role because most initiatives are initiated by market parties such as sector associations. Civil society is often only engaged through partnerships or consultation by market actors, although they can also take an initiating or pressuring role. The joint producer sector approaches rely on the control of suppliers through third party audits conducted by eco-label or other audit organisations. This provides more legitimacy and trust compared to business-to-business self control systems. The 3rd generation of SSCG-systems, the cross-sectoral approach, go beyond specific products and sectors. This approach is designed to function as a system that is widely applicable in a uniform way (Vermeulen, 2008).

The emergence of a whole range of different SSCG-systems is visible. This creates a wide range of new opportunities but also rules and demands for suppliers which they have to comply to. They have to start functioning in business-to-business governance systems in order to realize sustainable value chains.

2.3 Goals (claims)

Regardless of their differences, all three types of SSCG-systems have in common that they rely on the basic assumption that business-to-business supply chain cooperation can realize more social, economic and environmental sustainability (Vermeulen, 2008). They all claim to influence the supply chain in the three PPP-aspects: People Planet and Profit.

The extensiveness of the measures taken to realize more sustainability for People, Planet and Profit differs per system. Their goals can vary from having a narrow focus on just one of the three P's to having a very inclusive approach. Some of the often touched upon goals are preserving the environment, labour rights and fair profit distribution, each representing one of the P's. However, some SSCG-systems take on a much broader approach.

Utz Certified is one of these systems that take on a broader approach. They want their coffee growers to grow their coffee with care for the environment and their communities. With their system they want to empower growers with knowledge, realize school access for all children living on UTZ CERTIFIED farms, access to housing, healthcare and clean water and assure labour rights (UTZ CERTIFIED, 2010). The Marine Steward Council on the other hand has a more narrow approach with a focus on Planet. Their main principles are to achieve sustainable fish stocks, minimize environmental impacts and realize effective management (MSC, 2010). The Business Social Compliance Initiative is also one of the systems with a more narrow approach with their focus mainly on People. Their main goal is to assure protection of the worker's rights by assuring that their members live up to the International Labour Organisation Conventions and other important Declarations of the United Nations, the UN Global Compact and the OECD guidelines for multinationals (BSCI, 2010). Many systems exist and even more approaches are prevailing. The question however remains up to which level are these systems achieving their goals?

Previous studies have shown that SSCG-systems can bring about positive change in the livelihoods of producers. In Latin America sustainable certification processes have brought several benefits. Because of the certification producers received better prices for their coffee. In the case of Nicaragua farmers received up to 4,5 times more for their coffee than they did before joining the certification system. This large difference was also caused by the current coffee crisis which kept the coffee prices very low (Utting, 2009). Nonetheless, the resilience of coffee producers has improved (Utting, 2009; Kilian et al. 2006). Other positive effects that have been measured are improved access to credit, higher productivity of low-input agricultural systems, more entrepreneurial opportunities to differentiate their products in the market and increased competitiveness (Utting, 2009; Kilian et al., 2006; Kilian et al., 2004).

However, negative side-effects can also occur. Fair Trade organizations do not have the financial means to immediately pay for the product supplied. Sometimes it can take up to four months before payments are fulfilled. Consequently farmers will have to wait a long period before receiving their revenues, which creates the risk of them not honouring their contracts and selling their product to other buyers (Utting, 2009). Besides, paying a higher price only is not sufficient. Without changes in the quality of the products, the productivity, the export and value-adding processes real change in the region will be hardly realized (Kilian et al., 2006).

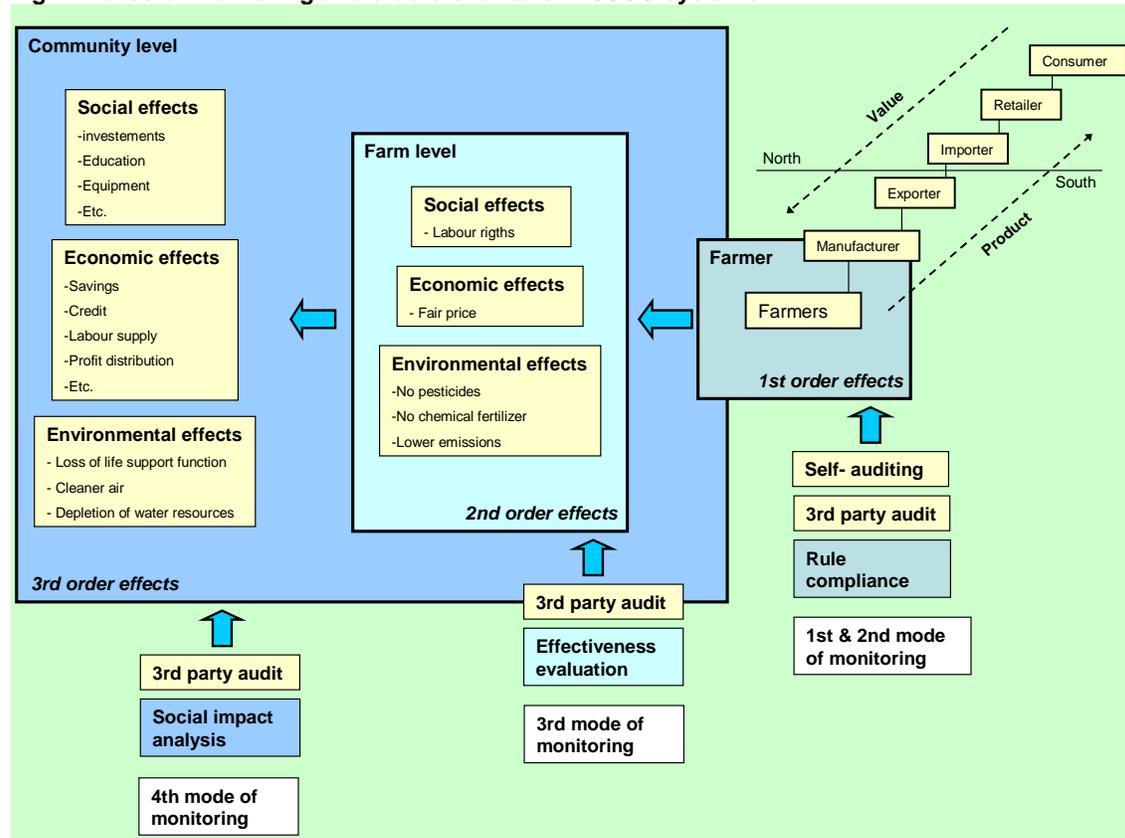
A different negative side-effect is the tendency of SSCG-systems to be mainly beneficial to the more affluent. The high initial costs for partaking in a SSCG-system are often unaffordable for small scale poor farmers. This causes the more affluent producers to reap the benefits while the poor often find themselves in more deteriorating situations. Consequently, the frequently intended goal of the empowerment of small scale farmers does not occur (Bitzer et al., 2008). It thus remains questionable who actually benefits from the new market opportunities offered by SSCG-systems (Kilian et al., 2006). Another negative structural effect can be caused by increasing food exports due to an increasing amount of farmers partaking in the SSCG-systems. Farmers who previously sold their products at local markets now opt for the more beneficial alternative provided through SSCG-systems. Consequently, local food supply decreases which raise prices and might necessitate import of food products.

Some information on the positive and negative effects of SSCG-systems is available. However, a clear picture of the actual on the ground impacts of these systems is often missing. There is a lack of empirical evidence of the impacts SSCG-systems cause. At this moment for many SSCG-systems it is unclear whether they are actually achieving their sustainability goals. Most studies focus on analysis of outputs and outcomes but they fail to make a connection to impacts. Clear insight into the costs and benefits for producer compliance is lacking. So far there is no consolidated framework to conduct critical, comparable science-based research about the impacts of SSCG-systems. Farmers, consumers, policy-makers and consumers lack objective information on what the actual impacts and side-effects are. Meanwhile the concern of the stakeholders about the effectiveness of SSCG-systems is growing. Filling this gap is not only important for taking away confusion for consumers and policy-makers but it is also crucial for the farmers to show that the standards actually deliver what they promise and moreover which type of sustainability initiative most fulfils their needs (Albersmeier, 2009; Utting, 2009; Studer, 2009; Giovannucci and Potts, 2008; Gonzales 2008; ISEAL Alliance 2009; ISEAL Alliance 2008; Kilian et al., 2004; Nelson et al., 2002;).

2.4 Monitoring systems

In order to control for the behaviour of the different actors in the supply chain and assure realisation of the PPP-goals different control mechanisms are constructed. These different control mechanisms are self-auditing, 3rd party audits conducted by an NGO or auditing agency and effectiveness evaluation also conducted through 3rd party audits. Each control mechanism measures different effects.

Fig. 1 Modes of monitoring and orders of effects in SSCG-systems



The first mode of monitoring is a self assessment executed by the farmer (supplier). The supplier checks whether he is complying with the rules as imposed by the SSCG-system. The effects measured can be called 1st order effects. The second mode of monitoring is a validation by 3rd party audits. The 3rd party controls whether the self-assessment is executed in the correct manner. This could be executed by a NGO or an auditing agency. The agency measures the compliance of the supplier to the rules as agreed upon in the SSCG-system and thus measures 1st order effects. The 3rd mode of monitoring is effectiveness evaluation. Effectiveness evaluation is partly conducted by 3rd party audits but also by academic researchers, universities and research institutes. This mode of monitoring is not only focused on analyzing the compliance of firms to the rules as put down by the SSCG-system but also monitors direct ecological and socio-economic impacts which can be called 2nd order effects. They measure for example the change in emission levels or the amounts of pesticides used. The 4th mode of monitoring can also be executed by 3rd party audits or through research. In 4th mode of monitoring an impact analysis is conducted through which 3rd order effects are measured. Through social impact analysis structural and side effects of the SSCG-systems are monitored. In this case the measurement is no longer on firm level ($n=1$), but on the wider context a firm is acting in ($n>1$). Together with 2nd order effects this third order effects express the effectiveness of the actual aim of SSCG-systems; sustainable development (see the figure 1).

In this research the focus will be on developing a 3rd and 4th mode of monitoring system in order the measure 2nd and 3rd order effects.

2.5 Challenges in designing a measurement system for SSCG-systems

When trying to design a measurement system for SSCG-systems there are different challenges that should be overcome in order to design a solid measurement system. These challenges can be attributed to the development context, the supply chain system context it has to function in, and the difficulty of measuring effects in general. The different challenges will be discussed in this order.

2.5.1 General challenges in a developing world context

Doing research in a development context is very different from doing research in a developed world context. The main difference lies in data availability and access to information. In western countries many basic statistics and reports concerning the population and their environment are available at governments or via local organisations. This offers the opportunity to easily gather some basic information about the research area. This assists in creating a first impression of the area and later in contextualizing the information. In a developing world context this information is often lacking and when it is available it is often in hard copy format and only available at specific places. Practice shows that it is often unclear where information can exactly be found. Local institutions do not have well organised systems to save their data. Even if a research has been conducted on a certain subject the reports are in many cases not saved or impossible to find. Thus one cannot depend on the availability of basic statistics as a source of information for research.

Besides the availability of data, reaching the target group often poses a difficulty. Conducting postal surveys is almost impossible. People are often inaccessible via postal services and moreover many people in developing countries are still illiterate. Consequently conducting postal surveys is not an option. This forces the researcher to conduct face to face interviews. This in its turn influences the amount of participants that can be taken up in the target group. Interviewing is a very time consuming activity and thus the n is likely to be small in comparison to postal surveys. Besides, discovering which people are part of the target group might be difficult since there is a lack of basic statistics. This can also cause problems for drawing a random sample. Finally, the accessibility of the target group might also be an issue. Target groups can live in remote areas and proper infrastructure is often lacking which makes not only postal surveys almost impossible but also interviewing difficult. Assuring a large sample is complicated.

2.5.2 Specific challenges for SSCG-systems

SSCG-system is a fairly new concept. The importance of conducting analysis of the effects of these systems is only in recent years receiving more attention. Therefore, not many measurement systems have been created so far. Besides, the difficulties in measuring the effects of SSCG-systems might have slowed down the process of creating measurement systems. Because not much research is done yet in this area, no well-established method is available and there are not many experiences a researcher can build upon.

Besides, there are some specific requirements for creating a measurement system of SSCG-systems that complicate the development a bit more. SSCG-systems operate on a global scale. To make the gathered data as valuable as possible assuring comparability of the data is important. Therefore, the measurement system should be applicable in a global context while on the other hand not overlooking context specific issues. This poses a mayor challenge.

Secondly, the population from which the sample can be drawn is likely to be relatively small. Although certification is an upcoming market, the amount of producers enrolled in these schemes is relatively small. Thus finding a sample might be difficult and help from local organisations could be necessary. Besides, the size of the sample is also likely to be small, because of infrastructural challenges and the time span available for conducting the research.

Because the measurement system will function in a market environment, costs are preferably kept low. The costs of enrolling in a certification scheme are high enough in itself for all stakeholders thus additional costs should be kept as low as possible. This also puts constraints on the amount of interviews that can be conducted.

2.5.3 Challenges for measurement of effects

Finally there are also some challenges which are specific for the measurement of effects. The measurement system is not designed for measuring a static point in time, but it is supposed to measure change over the years. In this case it is aimed to measure change caused by the certification scheme. The difficulty lies in being able to attribute the change measured to producers entering the certification scheme. Determining who the agent of change is can be a big challenge. Although certain changes can be measured, assuring that these changes are caused by the certification scheme is very difficult. There are always many contextual factors that might have caused the measured changes. Correctly measuring this causality proves difficult.

3. Theoretical foundation: reviewing research methods for impact measurement

3.1 Proving versus improving

In measuring 2nd and 3rd order effects two different approaches can be taken. While traditionally the focus has been on proving effects, currently more attention is being drawn to improvement. Whereas in the proving approach the focus is merely on providing results, in the improvement approach the focus is on learning from the results obtained and using this to improve the situation (ISEAL Alliance, 2009; Barrientos, 2003; Nelson et al., 2002). This improvement or learning approach enables one to study which aspects of a certification scheme are the most effective and why. It also shows how positive effects can be enhanced and negative effects minimized (ISEAL Alliance, 2009; Barrientos, 2003). It is important that not only proof of the results is presented to the different stakeholders but that information is provided to them that can lead to empowerment and change (Bryman, 2004). It is important to create awareness and better understanding in order to be able to stimulate change. In doing so it is important that all stakeholders are being heard (Bryman, 2004; Barrientos, 2003)

The objective of the research determines the approach, the type of data needed and the way it is analyzed. When the focus of the research is to prove effects, a proving approach should be taken whilst when the focus is on improving a situation, a learning approach should be taken. In proving an effect the focus is external, to prove results to the stakeholders. Precision and objectivity in this approach are important. SSCG-systems often bring along large costs for different stakeholders participating in it. Therefore, there is an increasing demand for results based management. Different stakeholders are demanding transparency and reliability in order to gain insight in what is actually being achieved (ISEAL Alliance 2009).

When taking on the improving approach the focus is more internal, the aim is to understand the process in order to improve the effects (ISEAL Alliance, 2009). Within the development community there is a shift visible towards a focus on the learning aspects of evaluations. The focus should be on how to improve the process.

The focus in this research is a learning approach but the importance of proving the qualities of certification schemes should not be denied either. In SSCG-systems external actors play an important role and proving the actual effects of SSCG-systems is therefore very important. Not only to convene direct stakeholders but also to convince the consumer about the effectiveness and functionality of SSCG-systems. Thus a combination will be sought in which proving the effects plays an important role, but at the same time an attempt will be made to reveal the successes and deficiencies in the system in order to be capable of improving it. Sharing this knowledge with all relevant stakeholders is important to realize the learning approach at all layers of the SSCG-system.

3.2 Different methods

When conducting an impact measurement different research methods can be used. Besides determining the research approach it is also important to determine the research method used. Three different types of research methods can be defined; quantitative, qualitative and participatory research methods. It is important to keep in mind that quantitative and qualitative methods do not equal quantitative and qualitative data. Each method can use quantitative (numbers) and qualitative (text, photo's, film etc) data. The difference between the two methods is that quantitative methods are often perceived scientifically rigorous and credible while qualitative data is difficult to compare without quantification (ISEAL Alliance, 2009). The three methods all have a different scientific origin explained subsequently.

3.2.1 Quantitative research methods

Quantitative methods have a positivistic foundation. In their perspective good research is conducted through quantifiably measuring the world along defined rules of inquiry. They believe that the 'truth' is out there and that it can be measured as long as objectivity is attained (O'Leary, 2004). Quantitative methods are derived from experimental and statistical methods in natural science. Quantitative research is about determining whether a hypothesis is true or false and whether it should be modified in order to measure the theoretical constructs of interest correctly. Hypotheses comprise of statements which are derived from theories. They comprise of statements about associations among constructs and about associations between constructs and observable indicators, also known as variables. A variable could be any attribute

that changes the values across things or people under study. In order to determine whether a hypothesis is true or false determining the validity and reliability of the research is of utmost importance.

A hypothesis consists of two constructs; a causal construct and an effect. The causal construct is the independent variable while the affected construct is the dependent variable. The degree to which both variables accurately measure the constructs of interest is the construct validity. With high construct validity all variables on which data is collected capture the constructs in the hypothesis in a successful manner. Construct validity is important to inform the theory and to prevent misleading results. The main problem with maximizing construct validity is that variables never solely measure the construct of interest. Other irrelevant characteristics are measured as well, called constructs of disinterest. Furthermore, a variable is also likely to contain random errors of measurement. One of the most important things to do to realize construct validity is to measure each construct in more than one way. This allows comparison between the different results. If the results are similar and we believe that the only thing these two variables measure in common is the construct of interest and this is evidence for their construct validity (Bryman, 2001).

Besides construct validity internal and external validity also play an important role in determining the quality of the research. Internal validity concerns the extent to which the method is measuring what it aimed to be measuring and whether there are no other causes for the measured results. It is thus about the causal relation between one variable on the other. Internal validity determines whether the manipulated variable is the cause for the results and that there are not any other influences. Consequently, a threat to internal validity is a variable that is 'alien' to the research but has a relation with the dependent and independent variable and consequently causes the results. The only way to get around these threats is to conduct a randomized experiment in which the variables are randomly determined. Generally, a variable that is randomly determined is unrelated to all other variables (Bryman, 2001).

External validity concerns the extent to which the results of the research can be generalized to the populations and settings of interest. When maximizing external validity it is important to specify before the research is conducted the limits of desired generalization. The group of people and the settings for which we believe the hypothesis will hold need to be precisely defined. To be able to realize generalization from a sample to the population of interest a random sample need to be drawn. When it is too difficult to draw a random sample generalization must be done on a theoretical basis. Speculation must be done about the possible differences between the sample and the objects that were left out of the sample. Replicating the research in other settings with other samples is of importance for realizing internal and external validity (Bryman, 2001).

Another important aspect in doing proper quantitative research is assuring reliability. Reliability is a prerequisite for validity; therefore it must be addressed before validity. "The reliability of a measure is defined as the extent to which it is free from random error" (Bryman, 2001: 83). Random error could be mental mistakes, slips of the pen, lack of concentration and so on (Bryman, 2001).

Random error is likely to vary from one occasion to the other. There are different methods to measure this random error. The correlation between scores measured with the same measure at a different point in time for example provides an estimate of the reliability. This correlation is called the test-retest correlation. However, it is often difficult or costly to do a retest. Therefore, there are also other methods such as internal consistency reliability can be used in which the correlation between different items is tested (Bryman, 2001).

In quantitative research the main tools being used are large-scale surveys which allow statistical analysis. In most cases random or stratified random samples are taken, which decreases the likelihood of bias. Furthermore, objectivity of measurement and analysis is in most cases to a large extent accomplished. The 'hard' data collected will be highly convincing, also towards policy makers (Mayoux, 2006).

Drawbacks in quantitative methods

Within quantitative measurements everything is focused on measuring the process or phenomenon you want to measure and nothing else. To do so the focus is very clear and narrowed down in order to avoid biases. However, this a priori focus might overlook some of the relevant issues brought up during the research. There is little room for addressing new issues brought up during the research. Furthermore, in realizing valid and reliable measurements many difficulties can arise. Practice shows that many samples taken are, in fact, not random and minorities can be under-represented in aggregate conclusions (Mayoux, 2006). In the case of SSCG-systems random selection is almost impossible since voluntary certification is a self-selection process. Furthermore, knowledge of the important variables is presumed when selecting a control group (ISEAL Alliance 2009; Mayoux, 2006).

The reliability of the information collected is not easily achieved. The type of questions asked is very influential as well as characteristics of the person conducting the research. The motivation of the respondents can also influence the results. Moreover, no matter all precautions taken it remains difficult to prove causality of the results. Besides, the type of data gathered is in most cases not specifically adjusted to be able to realize direct benefits for the respondents from these data (Mayoux, 2006).

The methods considered important within quantitative research cannot be applied in every situation. There are many contexts in which it is impossible to draw large samples, to assure random

sampling and to conduct all the different tests necessary to assure full internal and external validity and reliability (ISEAL Alliance, 2009). This is especially the case for research in a development context. Data bases are often lacking, people are illiterate and the researcher depends much more on the information that is available in contrast to making decisions on what information it is going to obtain. Therefore, the researcher is often forced to make use of more qualitative research techniques. Finally, the large scale character of quantitative methods sets high resource demands in terms of costs.

3.2.2 Qualitative research methods

The focus of qualitative methods is not on precise measurement of predetermined hypotheses but at realizing a holistic understanding of complex realities and processes (Mayoux, 2006). Qualitative research methods focus on the complexity of reality. Their approach is more aimed at discovering what has caused a certain process or phenomenon to occur rather than taking on an a priori focus and trying to prove this. Instead of testing theory qualitative research is an inductive process where theory is generated from research. Data generation and creation of theory go hand in hand. The focus is on creating an understanding and interpretation of its participants of the social world instead of adopting a natural scientific model. They thus take on an epistemological position. (Bryman, 2004; Robinson, 1998).

The open and flexible attitude offers opportunities to capture underlying meanings, local perceptions and sensitive issues and can be considered more in tune with social complexity. The in-depth analysis is a good tool for uncovering processes and causality. Furthermore, the large details often provided in the research offer the opportunity to emphasize a contextual understanding of social behaviour. Moreover, the focus is more on individuals rather than aggregates. This creates more room for individual values. Reality can be conceived different by different groups and individuals. In qualitative research the sampling methods are less randomized which offers the opportunity of focusing on specific cases and issues of interest. The data gathered is not necessarily fixed in its meaning. It is often multi dimensional and unstructured (Mayoux, 2006; Robinson, 1998).

Within qualitative research attention can be paid to assuring direct advantages for local communities. Opposite to quantitative research, qualitative research is not concerned about numbers but about stories. The research method leaves room for showing empathy and understanding. Furthermore, it gives people the opportunity to tell their story, an opportunity they might have never had before (Mayoux, 2006).

Within qualitative research realizing internal and external validity and reliability cannot be attained in the same manner as within quantitative research. Thus different methods have been found that try to attain reliability and validity in similar ways as in quantitative research (Bryman, 2004).

Attaining external reliability is difficult since it is impossible to freeze a social situation in order to make it replicable. This problem can be partly accounted for by either exactly describing the research situation and the position of the researcher or by realizing that more than one observer agree about what they see and hear. By presenting the exact situation in which the research took place and how it is conducted a future researcher might be capable of recreating the situation or analyzing and accounting for the influence of the difference in context (Bryman, 2004). Qualitative research offers the opportunity of capturing subjectivity. Furthermore, the threat of falsification can be minimized in in-depth longitudinal investigations (Mayoux, 2006). Realizing internal validity is one of the strengths of qualitative research, because the relatively long period of participation within the research group can ensure high correspondence between the theory and observations. Finally, external validity is difficult to realize since qualitative research is mostly focused on small scale samples (Bryman, 2004).

Qualitative research has a background with several different traditions. Therefore, it is difficult to describe a clear outline on what qualitative measurements ought to be and how it should be conducted. Different types of research techniques within qualitative research can be used that differ considerably from one another including ethnography/ participant observation, discourse and conversation analysis, case studies, focus groups discussions, qualitative interviewing and the collection of qualitative documents and texts (ISEAL Alliance, 2009; Bryman, 2004; Robinson, 1998).

Drawbacks in qualitative methods

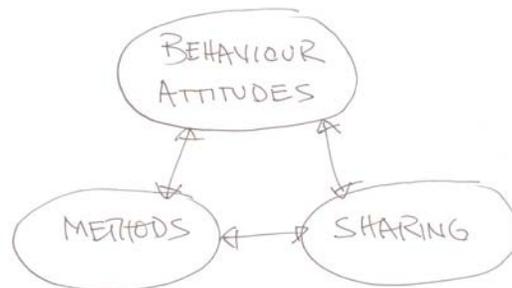
A drawback of qualitative measurements is the difficulty to make generalizations out of small scale non-randomized samples. The small scale level also makes the method very open to bias. Besides, objectivity is not assured which makes over-influence caused by the biases of the researcher a very likely situation to occur. These both influence the reliability of the data as well as the credibility of the analysis of the data. Credibility might furthermore be influenced through presence of the investigator. Besides quantitative measures are often very time consuming. Localizing respondents, accessing them and conducting the interviews is all very time intensive. Finally, because of its open attitude towards multiple realities qualitative research might lack focus (Mayoux, 2006).

3.2.3 Participatory appraisal approaches

Participatory appraisal methods have their origins in development activism: non-governmental organizations and social movements. Since the '70s the development of these methodologies has accelerated. Rapid Rural Appraisal (RRA), Participatory Rural Appraisal (PRA) and Participatory Learning and Action (PLA) are an important part of this development. Rapid Rural Appraisal was introduced to be faster and better than more quantitative methods such as large scale surveys as it offers the opportunity to rapidly gather a large amount of data. In the late 1980's and early 1990's PRA evolved out of RRA which in 1995 was followed up by PLA. All three methods have their own distinctness and are still being used. Whereas in RRA the information is mostly extracted and elicited by outsiders, in PRA this process is better shared with others. PRA is one of the most commonly used participatory appraisal methods (Mayoux, 2006; Chambers, 1994).

Obtaining knowledge is not the purpose of the participatory methods per se, but social change and empowerment are. It is meant to give voice to the most vulnerable and marginalized groups in society. The process will provide an increased understanding of development processes and provide more reliable and representative information to researchers (Chambers, 2007). This information is based on local perceptions and priorities, captures diversity and brings voice to the most vulnerable (Mayoux, 2006). Through empiricism good performance is hoped to be achieved (Chambers, 1994). The representation of the local voice is hoped to bring empowerment (Mayoux, 2006). Besides it gives insights into the thoughts of local people about their own situation and about wealth and well-being. This is shown to deliver useful insights. Especially since there are believed to be multiple realities which are location dependent. Primacy of the personal is an important aspect in this process. This aspects point to the importance of using your own best judgement at all times (Chambers, 1994).

Fig. 2 Chambers three connected circles



Source: Chambers, 2007: 8

PRA is based on three principal components expressed in connected circles. These three components are behaviour and attitudes, methods and sharing. From the beginning behaviour and attitudes are considered more important as the method used. The focus is not solely on the behaviour of the participants but just as much to that of the facilitators. The facilitators should assure to only fulfil their role as facilitator and not as leader. In order to do so important precepts are "Handing over the stick", "Don't rush", "Sit down, listen and learn" and "Use your own best judgement at all times" (Chambers, 2007: 8). Sharing refers to sharing resources between citizens such as food, knowledge, ideas, training and also sharing between governments and NGOs of materials, methods and insights. Local people are perceived and approached as experts. Relationships also have become part of the circle. Methods, the third circle, are a very distinctive part of PRA. In the appraisal different groups can be involved ranging from individual observations and small focus group to larger workshops (Chambers, 2007).

The methods used are visual and tangible and contain different tools such as diagram tools, oral and visual tools. This method makes it easier for non-literate citizens to engage in analysis and discussion and it can cross-border language groups (Mayoux, 2006; Chambers 2007; Chambers 1994). As mentioned before participatory research is seen as an alternative for large scale surveys. Participatory research provides more depth, richness and realism of information and analysis. The different methods open to use make direct and indirect triangulation and crosschecking possible which increases reliability of the information and the credibility of the analysis. Through the different methods used the facilitator gains personal experience which makes interpretation of change and causality possible. Besides, it can make analysis of complex issues possible (Mayoux, 2006; Chambers, 1994). Common tools for conducting participatory research are visualizations, narrative case studies, direct observations and group interviews (Mayoux, 2006).

Drawbacks of Participatory appraisal approaches

One of the drawbacks of participatory appraisal approaches is that they are often techniques for data collection to determine the viability of livelihoods but not so for research. The data collection methods are

often not considered proper for scientific research (Chambers, 1994). Moreover, the data gathered is in many instances difficult to analyze and compare over time or between different places. This causes interventions to be based on impressions rather than scientific analysis. Moreover, people might not express their true feelings publicly which leads to unreliable information. When expressing their true feelings they might be made more vulnerable.

Another problem is that of sample selection. A tendency is that only vulnerable groups are asked to participate since they are the ones that need empowerment the most. However, without participation of powerful groups it is very unlikely that practical policies will be formed based on the process. Besides, for research to be truly participatory it is important that all groups are represented. Moreover, empowerment may be more assumed rather than actually prevailing. Furthermore, some argue that there is a natural tendency for empowering the better of rather the worse, men rather than women and adults rather than children and young people.

The participatory approach's rising expectations towards the communities is not only positive. High demands and co-occurring pressure might cause some form of 'burnout' within communities. Finally, the facilitator is very important for guiding the process in an appropriate manner. Its role in assuring an atmosphere in which everybody feels free to speak up and all participants are being heard is a very decisive factor in the success of the method (Mayoux, 2006; Beazley and Ennew, 2006).

3.3 Finding a balance

Thinking back to the objective of this research the goal was to realize a valid, reliable and manageable measurement system in order to measure 2nd and 3rd order effects of SSCG-systems. The concepts of validity and reliability have their roots in quantitative research methods and formed since their existence the basis for all scientific research. In general a research project is expected to be viable and reliable in order to be able to make any claims about the research results.

When analyzing the theory behind participatory appraisal approaches and qualitative research methods it becomes clear that in some manner this reliability and validity is sought as well. However, it is also clear that in certain contexts realizing reliability and validity as described in quantitative methods is very difficult or almost impossible. This is especially the case in developing world contexts where lack of information is a daily issue. There are no databases to obtain information from and to draw baselines upon. People are often illiterate and hard to reach. In this context surveying becomes impossible.

Some of these issues are partly accounted for in the participatory appraisal approaches but large concessions are being done on validity and reliability. Therefore, there is a continuous struggle between assuring validity and reliability and making sure the research remains manageable. Manageability in this research is determined by the applicability of the measurement system in a development world context, the costs and the inputs (time, technical staff) necessary to conduct the measurement. This aspect of manageability is even more important because the measurement system is hoped to be used worldwide and on a frequent basis. The costs for certification are already high enough and should be maintained as low as possible. Thus the question is how to be able to find a common ground between the different aspects considered important. How can a manageable measurement system be designed while not giving in too much on validity and reliability?

O'Leary (2004) and Bryman (2004) both have attempted to take the indicators of good research as laid down in quantitative methods and transform them into more workable indicators. According to O'Leary (2004) the essential part of doing research is assuring that it is credible. "... it must have the 'power to elicit belief'" (O'Leary 2004, p. 56). Validity, reliability and objectivity are indicators to assure this credibility. To assure that these concepts are better manageable (in a development world context) these indicators are transformed into new concepts.

In quantitative research methods internal validity is to assure that the reality is correctly represented. It is to assure the trustworthiness of the conclusions drawn in a research. Conclusions based on the research results should accurately reflect what was studied (Bryman, 2004). However, some authors might claim that it can be difficult to capture this reality since different realities might exist. In order to assure that the results are credible and trustworthy the principle of authenticity can be used instead. Authenticity will determine the acceptance of the results by others (Bryman, 2004; O'Leary, 2004). Authenticity can be achieved by reflexive and rigour practice. Besides, the research should be carried out according to 'the rules' of good research and the findings should be submitted to the people under study. Through respondent or member validation they are able to reflect on the results obtained in the research. This process can either be done in a group or individually. Member validation is to assure that the results are in congruence with the views of those that are under study (Bryman, 2004).

External validity regards whether it is possible to draw generalizations from the research results to larger populations. In quantitative research this is assured by drawing random samples. In more small scale local research the focus is more to the uniqueness of the situation and the local context (Bryman, 2004).

Furthermore, the samples are very unlikely to obtain the size needed to draw generalizations. Transferability can be used instead which shows the opportunities for the applicability of the results on other populations (O'Leary 2004). Realizing transferability can be done through describing a local situation into great detail, also called thick description (Bryman, 2004). This description should also obtain a detailed description of the methods chosen and the specific research context (O'Leary, 2004). Thick description provides details to others to judge about whether the results can be transferred to larger or other populations (Bryman, 2004). However, how to exactly bring this concept of thick description into practice is not described by Bryman. Additionally, based on theoretical discussion assumptions can be made regarding how the research would have been different when the objects excluded would be included (Bryman, 2001).

Reliability concerns whether the data and results gathered will be the same under repeated trials and thus whether the methods used are consistent. Internal consistency test or test-retest measures are only applicable on large samples. In order to account for these criteria in small scale more qualitative research dependability can be used. Dependability does not assure consistency since it is very difficult to assure consistency in human responses. Responses from people can change every day and be influenced by the smallest things. Therefore, the goal of dependability is not to assure that the same results will be obtained in repeated trials but to show that the research is designed and developed in a consistent, systematic, logical and well documented manner. By following this strategy the researcher can account for research subjectivities (O'Leary, 2004). However, once more the exact strategy for bringing this concept in practise is not provided by O'Leary. A method of assuring dependability is through auditing of the research process by peers who function as auditors. The researcher is supposed to keep notes of all the different steps taken in the research process, from problem formulation to fieldwork notes. These notes will be analyzed by the auditors who then decide whether there are proper procedures are being followed. This method is very unpopular however since it is very time consuming (Bryman, 2004).

Next to reliability and validity, objectivity and reproducibility are also important indicators of good research. Objectivity is to assure that the results are not influenced by personal subjectivities and is to prevent bias. The relationship between the researcher and the research subject should be mediated by protocol, theory and method. However, it is very difficult for a researcher to have a totally objective position towards the research subject. Confirmability addresses the issue of objectivity in a less strict manner. It stresses the importance for the researcher to engage in reflective practices in order to acknowledge their own subjective positioning (Bryman, 2004; O'Leary, 2004). In the case of a researcher having a clear agenda managing subjectivity is not easy. In this case the researcher should assure subjectivity with transparency which means managing subjectivities through making them transparent. By doing so the researcher assures that the biases do not influence data analysis (O'Leary, 2004).

Reproducibility concerns if a methodology when used in different studies with a similar context will yield the same results and conclusions. However, reproducibility can only be achieved when the influence of context, the subjectivity of the researcher and variability in or of the research are accounted for (O'Leary, 2004).

Based on theoretical foundation this is the balance in achieving validity, reliability and manageability that can be sought, as summarized in table 1. To further gain insight in how this can be transformed to a measurement system some examples from practice were studied and analysed. These lessons are presented subsequently. In chapter 5 the lessons learned in this chapter as well as the subsequent chapter are combined into a valid, reliable and manageable measurement system.

Table 1. Balance in achieving validity and reliability

Internal Validity	Authenticity: <ul style="list-style-type: none">• Reflexive and rigour practice• Respondent validation• Rules of good research
External Validity	Transferability: <ul style="list-style-type: none">• Thick description:<ul style="list-style-type: none">- methods- research context• Theoretical discussion
Reliability	Dependability: <ul style="list-style-type: none">• Research design:<ul style="list-style-type: none">- consistent- well-documented- systematic- logical• Peer audit of research process• Note of the whole research process

4. Practical foundation: reviewing practices of measuring impacts in developing countries

From chapter 2 several key issues in measuring effects in a developing context can be discerned. It would be most interesting to analyze the solutions brought up by previous attempts to measure effects on all these different key issues. However, there are several themes which make it difficult to follow this strategy.

First of all it proved difficult to find many previous attempts to measure effects of SSCG-systems. It is a fairly new subject and searching the internet and academic databases did not provide many results. Especially finding academic research on this specific topic, where the focus is on measuring effects on farmers and their environment showed very difficult.

I did encounter a fairly large amount of literature focused at monitoring environmental aspects in developing countries. One of the important authors in this field who has been working on this topic for several years is Finn Danielsen. He and his colleagues share a large amount of information on this subject via their website¹. Via this website a large international network of researchers and practitioners share their knowledge on innovative approaches to monitoring of natural resources, livelihoods and governance.

Although this is very interesting, most of the information encountered here was not useful for this research. First of all, because the focus in their work is on monitoring and not on measuring effects. Monitoring a situation necessitates different data as measuring effects. When measuring effects one is looking for causal relations and explanations while the focus in monitoring is on recording the situation in different points in time. Secondly, the focus in all the monitoring schemes encountered at their website and in their articles is mainly on monitoring environmental aspects and how they can be best managed by the local communities. This mainly includes technical measurements to determine the well-being of different aspects of the environment such as animal populations and the use of resources. Although this is very useful information their research has no relation to supply chains and the certification of products. They do not focus on measuring the change being caused by compliance rules as put down within certification schemes, but on effectiveness in natural resource management. Therefore, their focus and approach lies in different aspects. Because of all above mentioned reasons this share of information was not of use for this research.

Most information found on measuring effects of SSCG-systems was practice based, developed by institutions working in the field such as Organic Exchange and ISEAL. In order to not only make use of these practice based experiences I also included some academic studies where the focus was not on measuring the effects of SSCG-systems specifically but on other types of policy interventions. In these cases lessons can be learned from their approach taken to measure effects. Since there are also some general issues it is interesting to study how these issues are tackled by different researchers.

Since not many measurement systems exist yet for SSCG-system some of the advice provided by different authors and organisations working in the certification domain was also used when creating the measurement system. Many NGOs are currently developing their own measurement systems and their obtained insights in this process can be of use.

Analyzing all the different key issues in the articles encountered can be a bit difficult because they do not all explain their motivation behind their research approach. The main focus of the articles is on explaining the research goals and results but not so much their research strategy. Thus it might be that their strategy of tackling some of the key issues can be discerned while others not. Off course each author could be approached in order to ask them about their motives but this is a very time consuming task. Since I only had a limited amount of time to conduct the research I choose to not invest this time in learning from previous experiences but rather invest this time in testing my system in practice and learning from this.

In the following different methods of measuring effects are in short explained. Each method is analyzed for the key-issues as discerned above. The complete analysis can be found in annex 1. Some of the key-issues relate to realizing validity, reliability and manageability. Consequently, in the table in the annex the issues are ordered according to these three categories. Only the issues from which specific lessons could be learned are further explained after the general description of each method.

4.1 RISE

The Swiss College of Agriculture has developed a computer-based tool to assess the sustainability of agricultural production called RISE: Response Inducing Sustainability Evaluation. This tool covers

¹ www.monitoringmatters.org

ecological, economic as well as social dimensions. The goal of the tool is to realize a holistic improvement of the sustainability in agricultural production by uncovering the strengths and weaknesses of farm and bringing up intervention points. In order to be able to conduct an impact measurement a second analysis is needed at a later point in time. Data is gathered at farm level using a comprehensive questionnaire. The results on the indicators are retrieved through computer-aided calculations which are followed by a feedback discussion of the results with the farmer. In this manner measures to improve the situation can be yielded (Studer, 2009).

The tool consists of 12 indicators each having their own parameters outlining the state of the system and others outlining the pressure on or driving force within the system. State parameters (S) vary between 0 and 100, 0 being the worst case and 100 the best case. Driving forces (D) also range between 0 and 100. However, driving forces have a negative pressure on the system and therefore 0 is the best case and 100 the worst case. The Degree of Sustainability (DS) is calculated as $DS = S - D$.

The output is presented in a sustainability polygon to assure easy and clear representation for all stakeholders. However, if desired the scores on each parameter can also be traced back to be capable of highlighting the necessary intervention points (Ibid).

The RISE system appears to be a credible system with a large share of attention to statistical reliability. However, because of searching this high level of statistical reliability and high academic rigor it gives in on manageability. The amount of time needed for an interview is large and there are high demands towards the knowledge of the researcher. Data availability also proved an issue in their system. In a western context this was no problem but in the developing world context it made it sometimes more difficult to conduct the analysis. In western countries it might not be difficult to find the trained persons but in developing countries this situation is the opposite, which makes the system not so manageable. Moreover the high time and knowledge demands make it a costly undertaking.

Another drawback in the system is that it does not measure impacts but only monitors a current situation. To be capable of measuring impacts two measurements at a different point in time are needed. Obviously this raises costs and it also demands perseverance in waiting several years before getting any results.

Although it has some drawbacks there are also some strong points in the system that could be a lesson to other systems. The use of driving forces provides an interesting insight as it tries to capture some of the dynamics which change the system over time. The clear representation makes it very comparable between different regions as well as sectors. Moreover, their trials showed that the representation in the sustainability polygon was very much appreciated.

4.2 The COSA Project

The Committee On Sustainability Assessment (COSA) is coordinated and led by the International Institute for Sustainable Development (IISD). The committee includes several research centres located in Nicaragua, Costa Rica, France and Switzerland and was created with initial support from the United States Agency for International Development (USAID), the Norwegian Agency for Development Cooperation (NORAD) and the Belgian Development Cooperation Agency (BTC). The Committee is in the process of developing a rigorous assessment tool to assess the costs and benefits of undertaking sustainability practices. With this assessment tool the Committee hopes to provide farmers and policy makers with information on which they can base rational choices about sustainability (Sustainable Coffee Partnership, 2007).

The Committee has created an integrative tool focusing on social, economic and environmental subjects. It has a quantitative basis to be capable of comparing and evaluating the effect of sustainability interventions. This data can be used for different types of statistical analysis to assure validity and credibility. However, the COSA also recognised the importance and integrity of the indicators themselves. Therefore, they also included multi-criteria analysis. In this form of analysis distinct variables are measured simultaneously without attempting to translate variables into a single common unit. By doing so the integrity of the variables is maintained. Local stakeholders are capable of assigning a weight to indicators, specific for that task and context. Moreover, it offers the opportunity of comparing the performance of different certification systems or with conventional practices on a farm-by- farm level. Finally it also allows for correlation analysis between different parameters of a variable (Giovannucci and Potts, 2008).

In order to test their method, COSA has applied their questionnaire with four coffee sustainability initiatives across five countries. Different lessons were learned during this testing period.

It showed that in general farms do not draw up reports, which again addresses the difficulty of data availability. In very few cases some data on farm-level was found but this was almost never found to be useful information in order to determine sustainability. This was especially the case for the economic data. Gathering this data thus proved difficult and time consuming. To overcome this problem COSA suggests developing a simple farmer "economic log" which farmers can use on an ongoing basis. The information collected in this manner is valuable for COSA as well as the farmers themselves (Ibid.).

A different issue was the difficulty of locating farms that recently adopted a sustainability initiative and that were only involved in one sustainability initiative. Besides, not every farm located was willing to participate in the research because they feared release of commercial information (Ibid.).

Also some problems were encountered with the format of the questionnaire. The meaning of different measurement units differed considerably between countries but also between different regions within countries. Thus elements of the questionnaire should be adapted to the specific local context. Besides, the questionnaire was too long. This was caused because of the difficulty in gathering economic data as well as confusion about who to ask which questions. Streamlining this process will shorten the interview time needed and the costs as well (Ibid.).

In respect to causality and the attribution problem it is difficult to determine whether the taken approach is appropriate or not because they do not exactly describe how the measurements are conducted. It did become clear from the documents that they do not pay attention to the local context. This makes it hard to understand the results in this local context.

Just as with RISE the COSA program has to conduct annual measurements for a period of at least three years in order to be capable of putting claims concerning changes caused by the certification. Again this puts pressure on time and resource limits.

4.3 Impact evaluation of the NFPP on rural household livelihoods

In their study, Mullan et al. (2009) tried to estimate the impact of the Natural Forest Protection Program (NFPP) on two facets of household livelihoods: income and off-farm labour supply. The NFPP aims to protect watersheds and conserve natural forests. To conduct the measurement they made use of the micro-econometric policy-evaluation techniques difference-in-differences and propensity score matching.

Because they were unable to measure the impact on income and off-farm labour supply in both the presence and absence of the NFPP they treated the program as a natural experiment. They measured the change in income and labour supply through time in areas where the program was in place, the target group, and areas where the program was not in place, the control group. They made use of multiple parametric and semi-parametric policy-evaluation techniques to assure more robustness of the results (Mullan et al., 2009).

The difference-in-differences and Propensity Score Matching (PSM) techniques both deal with the identification problem which is often encountered when trying to measure policy impacts. The impact of the program is measured by Y_1 , the outcome when an individual participates in a program, minus Y_0 , the outcome when the same individual does not participate in the program. Thus the impact is measured through the following equation:

$$\text{Impact} = Y_1 - Y_0$$

However, we can never know the value of one individual in both situations because one cannot measure the outcome of participating of non-participants and the other way around. Thus techniques are needed to solve this missing data problem and to create a counterfactual. Non-participant outcomes can be the counterfactual but only when program participation is random. If program participation is not random different outcomes can be expected for those participating in the program. This is either because they choose to participate based on expected returns or because the policy only targets specific groups of society. In the case of the NFPP the first issue was not a problem because the program is compulsory. Another problem occurs when macro (economic) trends have different impacts on the control group and the target group. There are different methods to overcome this problem such as including a vector of observable characteristics that can potentially influence the dynamic of the outcome. A different method is selecting individuals in the target group and control group based on the similarity of their observed characteristics (Ibid.).

Several lessons can be learned from the study in China. The first is that it is important to realize that the results only represent an average impact on households. The exact impact on each household can be higher or lower as displayed in the results. Second, it is important to realize that the results only pertain to a specific region. Different institutional, socio-economic and environmental conditions might lead to a different impact on households of the NFPP.

Problem of PSM is that the sample should be large and it is difficult to find a match. This is especially the case in a development context where little information is available on respondent characteristics. Moreover, very specific cases drop-out because it is impossible to match them. Thus one might wonder whether you can assure good representation.

A third lesson to keep in mind is the manner the researchers try to correct for external influences. As described earlier determining who the agent of change is and attributing this change can be very difficult. Several contextual factors are at play. The solution these researchers offer is not so much contextualizing

the results but they try to keep the external factors as similar as possible by making use of propensity score matching. In this manner they try to minimize the impact of contextual factors on the results.

4.4 Livelihood Change in a Changing Natural Resource Management Context

Mahdi et al. (2009) have conducted a research to measure the effects of a changing natural resource management context on livelihoods and their sustainability in the uplands of Lembang sub-watershed in West Sumatra, Indonesia. They have made use of Rapid Rural Appraisal and qualitative methods. They conducted household interviews, interviewed key persons and organized focus groups. They measured changes at two points in time, 1996 and 2006.

The change in livelihood was measured by quantifying the indicators of access to capital assets. These indicators are based on the expected change in the capitals caused by the changing natural resource management context. By using these quantitative methods they claim to be more capable of accurately determining the change in access to capital assets. Two statistical tests were conducted on the quantitative data to determine the significance of differences in access to capital assets in the two points in time and amongst and between the different income groups (Ibid.).

In their research they also address the importance of using qualitative analysis because the information gathered here goes beyond the scope of quantitative measurement. Qualitative methods support and strengthen quantitative methods (Ibid.).

A drawback of quantifying the capital assets is that not all building variables of a capital asset can be included, which can lead to bias. They claim that this can be mitigated by selecting those variables that most strongly represent the effects of a changing natural resource context (Ibid.). However, in many cases it will be difficult to fully represent the meaning of an asset by using only a few building variables.

Another drawback of their method which they also recognize themselves is that they cannot determine who the agent of change is. On basis of their research they can only claim that the measured change in access to assets is caused collaboratively by all changes in context. Besides, measuring two separate points in time only offers the opportunity to make statements about these points in time and not about the livelihoods dynamics between them (Ibid.).

4.5 Integrative Framework for assessing the impact of fair trade coffee

Karla Utting (2009) has created an impact assessment framework to assess the positive and negative local-level impacts from responsible trade interventions. The framework has three main objectives: ..."first, to explore changes in livelihood systems, standards of living and sustainable development; second, to identify conflict of interest and trade-offs between different local stakeholders; and third, to identify any potential for innovation and practical improvement of responsible trade initiatives (Utting, 2009, p. 131)".

For developing her framework Utting has especially made use of the Department For International Development's Sustainable Livelihoods Framework and their Stakeholders Analysis methodology. Her framework consists of four stages (Utting, 2009):

1. *Vulnerability context*: In this stage the context in which people live is studied and the stakeholders involved are identified.
2. *Impact assessments*: As the name says, this stage analyses the impacts of responsible trade interventions on a multidimensional level: the impacts on livelihoods of primary stakeholders, on the socio-economic condition of community stakeholders, on organizations and the environment are assessed and finally policy and institutional impacts are assessed.
3. *Conflicting interest and trade offs*: The third stage analyses the different conflicts and trade offs amongst local stakeholders. The question is how these conflicts and trade-offs influence the possible impacts on local stakeholders.
4. *Discussion of local-level changes*: The fourth and final stage consists of discussing the gathered data and coming up with possible improvements to better achieve the stated goals as laid down in the responsible trade interventions.

The framework was tested in northern Nicaragua on the fair trade coffee industry. A combination of various quantitative and qualitative research methods was used. These methods included; secondary sources,

published summaries of organizations, close-ended surveys, open-ended surveys, structured, semi-structured and unstructured interviews, rapid rural appraisal, focus group analysis and observations (Ibid.).

The assessment framework was found to be useful for assessing the impacts on local livelihoods. The assessment included different social, economic and environmental assets. This offered the opportunity to present in detail the effects on local livelihoods. This shows that development is not just a matter of economic improvement but also of empowerment of local stakeholders (Ibid.).

The method of analyzing the impacts through four stages assures more knowledge on the context of the impacts. One is more capable of understanding the local context and external influencing factors which might have caused a change in livelihoods. This is reinforced by all the different interviews conducted with local institutions and community members that are not direct stakeholders. By asking them about changes in all aspects of social and economic life that occurred in the last years the researcher is better able to appropriately attribute the measured changes in livelihoods.

A drawback of this research-method is that it is very qualitative. Most of the interviews are semi-open interviews which produce qualitative data. The format for the questions and answer categories for the different respondent groups are very diverse. The interview-formats themselves within each respondent category was also much diversified, which complicates result analysis. It is time consuming and far more difficult as when quantitative data is being used.

4.6 Organic Exchange

Organic Exchange is the only organisation I know of that has developed a monitoring system in order to monitor the impacts of organic cotton certification. Organic Exchange has developed a self-assessment methodology to measure the progress in organic cotton production (Organic Exchange, 2010).

The self-assessment methodology is a tool to monitor progress, problems and interventions required by using Key Performance Indicators (KPIs). The focus is on capturing progress and realizing improvement. It does not only assist in acquiring information on social, economic and environmental performance but also in increasing transparency and farmer visibility. Finally it can also offer a form of justification to the investments made by different stakeholders in organic cotton production (Ibid.).

The system is very inclusive, addressing a wide range of social, economic and environmental aspects. The method used can be considered a PRA since the assessment is conducted in semi-formal discussion groups with farmers and other parties. In these group meetings the current social, economic and environmental situation of the farmers under consideration and the status of organic production are scored according to predetermined categories (Ibid.).

For all the different aspects a description is given about what this aspect comprises. Based on this definition five different situations are presented in which a farmer can place itself. These situations range from being underdeveloped to being far-developed on the aspect. During the group-meetings each aspect and the current level of development on that aspect of the group of farmers is discussed. Based on this the discussion facilitator together with the farmers will determine in which state of development the farmers are at that point. Subsequently these results can be presented in charts and compared to other regions. The group meetings are held every year in order to be able to detect changes throughout the years (Ibid.).

The monitoring system is useful to get a picture of how the current situation in an area is in respect to different sustainability issues. However, there are several problems, often encountered in PRA-methods, which are also an issue in the self-assessment method. It is difficult to assure good sample representation. Besides, dominant participants might lead the evaluation in a certain direction which can cause bias in the results. Another important problem is proving causality. Not much attention is paid to the attribution problem. However, because of their years of experience they realize these problems exist and are trying to overcome these problems.

Assuring good sample representation is especially very difficult in their approach because they conduct the group-meeting the day of bringing in the yield. Because of doing so their samples depend on the persons that will bring in the yield to the buying company. An attempt to assure good representation could be conducted by inviting men, women and children to the assessment on beforehand². Moreover, farm group leaders can help choose the farmers to include³. However, this does not rule out the probability that only very motivated and good scoring farmers are willing to participate.

To assure that dominant participants or discussion facilitators do not dominate the discussion and influence the results the potential blocks to discussion can be analyzed before starting the meeting. Whenever these blocks seem present one should try to change these dynamics to create a relaxed atmosphere in which each person feels free to speak and share their thoughts. The atmosphere should be relaxed and non-hierarchical. Creating a relaxed atmosphere can also partly be attained by conducting the

² Interview with Simon Ferrigno, Organic Exchange, England on: 12-01-2010

³ Interview with Prabha Nagarajan, Organic Exchange, India on: 12-01-2010

assessment when the pressure of farming is low, i.e. post harvesting. Also during the meeting the group should be as relaxed as possible. Starting the meeting with an open discussion can contribute to realizing this. Besides, addressing over and over again that the meeting is meant to help the farmers and not penalize them can also assist in creating a relaxed atmosphere and obtaining honest answers that reflect reality (Organic Exchange, 2010). Finally, the facilitators should guide but not direct the discussion. The scores on the different indicators should be a group agreement rather than the facilitators' interpretation⁴.

Another problem encountered in their system is proving causality. At the moment the focus of their system is at monitoring the year to year change in organic cotton production. However, in the future they want their system to function also as an impact measurement system. In order to be capable of functioning in such a context more attention should be given to the attribution problem. Their conducted monitoring can be used as baseline data however it still remains difficult to determine who the agent of change is. Some of the solutions they offer for this problem are calibration with different stakeholders at different levels^{5 6}. Besides, feedback with farmers can also be a useful method. The notes taken by the discussion leaders and its assistants can also provide good insight into the progress of the assessment and the possible external influences at play. But one has to realize as well that it is almost impossible to measure a pure cause-effect relationship⁷.

4.7 ISEAL Alliance

ISEAL alliance is a global association for environmental and social standard systems. ISEAL guides and works with existing and emerging standard systems to strengthen their effectiveness and expand their impacts. Part of their current work is the facilitation of an international consultation process in order to define a Code of Good Practice for Measuring the Impacts of Social and Environmental Standards (ISEAL Alliance, 2010).

Their Code of Good Practice is based on the theory of change. The theory of change tries to explain impacts by analyzing the process as a whole, from inputs to outputs, from output to outcome and finally from outcome to impact (Ibid.).

The ISEAL Alliance has developed a draft of their Code of Good Practice together with the input from several sustainability initiatives around the world. Although they are still in the process of further developing and finalizing the Code I will explain their approach so far.

The Alliance does not provide a finished measurement system which interested parties can use without making any adjustments. They provide a guide with steps that should be taken in order to develop a sound and credible impact measurement system. Subsequently this line of reasoning will be shortly described because they show some interesting lessons and point of view.

First of all determining the scope of the evaluation programme is important as well as ensuring stakeholder engagement. Stakeholders often have specific expertise or knowledge and besides it can help in building trust and legitimacy. This can facilitate access to participants. Stakeholders should be asked about their perspective on importance indicators to be assessed, data collection and unintended consequences of the standard system (ISEAL Alliance, 2010).

The second part consists of describing the intended change the standard systems aims to bring about. This will help in determining where impact is likely to occur and how and thus provides insights on how these impacts can be measured. In this second part it is very important to describe the expected short, medium and long-term results. These three types of results are consistent with outputs, outcome and impact. Gaining insight on this causal chain will provide knowledge on how change is achieved. Besides, external factors that can influence the short, medium and long-term results should also be mapped (Ibid.).

The third part consists of monitoring. To be capable of conducting the monitoring for each short, medium and long-term result the indicators for measuring these results should be determined. To simplify this process one should first determine the area of impact, second the desired impact, third the strategies the standard system employs to achieve the desired impact, fourth the expected outcomes and fifth the indicators needed to determine whether the expected outcomes are occurring. For each indicator selected a protocol should be written on how the data should be collected (Ibid.).

Whenever possible the manager has to establish as baseline for each indicator or compile external data such as control groups against which the data can be compared. To be able to contextualize the information it is advised to gather general context information about a region. Beside, using a mix of

⁴ Interview with Liesl Truscott, Organic Exchange, England on : 12-01-2010

⁵ Interview with Alfonso Lizzaraga, Organic Exchange, Peru on: 12-01-2010

⁶ See footnote 4.

⁷ See footnote 4.

methodologies is advocated in order to cross-check data. However, this might not be possible in each case because of higher costs involved in using different methods (Ibid.).

When creating the monitoring system the data management should also be kept in mind. When using quantitative data, results can be formatted in tables or graphs and when using qualitative data this should be organized systematically so that comparisons and conclusions can be easily drawn. Before analyzing the data gathered evaluators should review the completeness of the data, errors and uncertainties, sampling errors and data validation methods. Moreover it is important to know the recipients of the evaluation as this will influence the type of report (Ibid.).

Finally transparency throughout the whole process is advocated. All decisions and steps taken should be described and explained. Realizing transparency about the process and conclusions drawn will strengthen the credibility of the results (Ibid.).

The ISEAL Alliance has provided a well thought through method to create a monitoring system. Different important issues are discussed and tackled. However, it does not provide an actual monitoring system as such. There still remains a great amount of work for standard systems before they can start measuring the impacts of their system. This fortifies the need for creating a measurement system in order to measure 2nd and 3rd order effects. While the indicators measured by the system created in this research might differ from those considered important by the standard systems it does offer an example of how one can actually measure these impacts. On the other hand it is important to realize that assigning the impacts to interventions might be difficult when the total causal chain is not mapped. Therefore, it is important to realize that creating a measurement system in itself is not totally sufficient in order to provide causal explanations; it should be considered part of a bigger analysis.

4.8 Summary

The lessons learned from analyzing the different systems can be summarized in the subsequent manner.

When choosing a research method for the measurement system the researcher can make use of a mix of methods. This can positively influence the completeness of the data; it is a manner of cross-checking the information. It is also important to pick a method that keeps the balance between straightforwardness and the complex reality.

Also attention should be given to the attribution problem. The method should be capable of measuring causal relations. In many of the systems analyzed addressing the attribution problem proved very difficult. One of the solutions was making use of a control group and drawing a baseline can also be a viable option. Qualitative methods, such as key-person interviews, are also very useful in gaining more insights in the actual cause-effect relationships.

The analysis also drew attention to the importance of contextualization. The results gathered at micro level will have to be put into context. Understanding more of the local context will assist in understanding the results better and interpreting them in the right manner. But also regional, national and international contexts can influence the results. Although some measurements can be done at the local level it is important to always have some awareness of the international context along the supply chain in order to correctly interpret the results.

Data availability proved a big problem. Often the necessary data is not available which influences the manageability of the system to a great extent. Other practical issues such as locating farms and the time of interviews are also important considerations when developing a measurement system. These practical issues should be kept in mind while developing the system.

Concerning the specific indicators used it is important to assure that the target group will understand the meaning of indicators as used. If an indicator of concept is misunderstood this could influence the results. To assure that the concept is measured to its full significance multi-criteria analysis can be used. It increases the chance of measuring the indicator in all its aspects and hence assuring good representation of reality.

During the whole measurement process transparency should be accomplished. This includes transparency concerning the methods used and steps taken as well as transparency in the results. Clear and good representation of the results is part of that. First however the completeness of the data should be reviewed. Keeping the recipients of the results in mind is important, but the results should also be comprehensible for the producers. An appreciated method for results representation is by using graphs or a sustainability polygon in which the score on varying sustainability dimensions is displayed.

Analyzing these previous attempts of measuring effects has provided many lessons and insights. These lessons are together with the insights learned from analyzing theory on quantitative, qualitative and participatory research methods used to create the measurement system as presented in the subsequent chapter.

5. The measurement system design: balanced impact measurement for SSCG-systems

Different approaches and experiences for conducting research have been discussed. All methods have their own positive features and drawbacks. None of them is able to assure the perfect representation of reality. Simply choosing one method for designing the measurement method would be a waste of all the opportunities the other methods offer and the knowledge obtained via analyzing previous studies. Moreover, it has become clear that the use of one method does not exclude the use of other methods. All methods stress the importance of triangulation and cross checking and nowhere is stated that the different methods cannot be used in a cooperating manner. Therefore, my measurement system presented in this study is not based on a previously determined research method, but is a balanced system in which reliability and validity are assured while maintaining the system manageable. Thus, the system is based on a mix of quantitative, qualitative or participatory appraisal research approaches to achieve the best possible measurement of 2nd and 3rd order effects in a development world context. This theoretical background of research methods is combined with the practical lessons learned in order to create a viable, reliable and manageable measurement system.

Subsequently the measurement system is explained. Manageability, validity and reliability will be addressed in this order. Afterwards the practical appliance of the measurement in six steps is presented. Furthermore, these will be elaborated upon within the context of the fieldwork I conducted on organic cotton certification in Paraguay.

5.1 Addressing the measurement challenges

Manageability

When choosing the research approach addressing manageability is very important. The content of the interviews is aimed at producing answers and results that are easy to analyze and use. However, it should not rule out the possibility to obtain information about social issues. Besides it is also important to be able to ask different questions about one topic, but on the other hand the interview cannot take up too much time, in order to maintain the respondents' interest and concentration. The often used method in western countries of paper surveys is not a viable option in a development context. Thus the question was how to formulate questions to which answers are easy to analyze, that do not take up too much time but that does not rule out the value of a concept either. The solution is to make use of a Likert-scale system.

A Likert scale offers the opportunity to quantify the indicators while not having to translate them in concepts that do not truly represent the indicators. A Likert-scale question is a closed question in which the respondent can choose between different answer categories. The categories in general range from very bad to very good. The precise categories and the amount of categories differ per research. To each category a value is attributed. In this case there are five categories with values ranging from 1 (worst case) to 5 (best case). A Likert-scale enables the researcher to measure more socially based indicators. Moreover, since a Likert-scale has predetermined answer categories answering the questions goes quickly and many different questions can be asked in a reasonable amount of time.

The results from Likert-scale questions are easy to analyze since they can be displayed in graphs. In this manner the results for the target group as the control group can be easily compared and analyzed. The results are also easily compared amongst groups from different regions or nations which adds value to them.

Although putting the results into graphs is fairly easy, the great amount of questions will also lead to a great amount of graphs. To overcome this problem the results of all questions concerning one indicator will be summed and divided to come up with one mean result for each indicator. This means the results can be presented in sustainability polygons to give a clear and quick insight into the results. If someone is interested in the build up of this mean result of an indicator, for example to see where improvements can be made, it is possible to look at the results on specific questions. By using this method the researcher is able to obtain a clear and quick oversight of the results while also being capable to go into detail when wished to do so.

Conducting an interview based on Likert-scale questions is relatively easy and the researcher does not have to be very experienced or knowledgeable. No experts will need to be hired which keeps expenses low. Costs are also kept minimal because of the relatively short time span of the interviews. Moreover the research does not fully depend on the availability of external documented data. They are only needed to be able to crosscheck the results and put them in context but without them results can also be obtained.

Table 2. Manageability characteristics

	Characteristics
Manageability	<ul style="list-style-type: none">• Likert scale questions• Research schedule: six steps

Validity

Construct validity should be realized in order to assure that the concept being measured truly reflects what the concept is about. Construct validity is to a large extent realized through an extensive analysis about what the concept of sustainable development comprises, which indicators should be included in the system and how these indicators are defined and founded in theory. Secondly, each indicator will be measured in several manners. A multi-criteria analysis will be conducted through asking multiple questions about each indicator.

Next to addressing construct validity internal validity should also be realized. Internal validity is realized when we are actually measuring what we claim to measure and there are no influences from external variables. Internal validity is partly realized by minimizing the attribution problem and all the measures that are taken to assure that. Realizing insight in the cause for the change as measured creates more insight in the internal validity. It enables the researcher to determine whether the independent variable is the cause for the change as presented in the results or whether an 'external' variable has influenced the results. How to address the attribution problem is discussed further on.

Internal validity can also be strengthened by the researcher through conducting a reflexive and rigour practice. The research should be carried out under rules of good research and member validation has to be executed. Member validation will be done by organizing focus group meetings after the individual interviews. Two meetings will be organized; one with the target group and one with the control group. In these group meetings the preliminary results will be discussed and there will be an opportunity to provide feedback to these results.

The attribution problem is overcome by conducting the measurement with a target group as well as a control group. Initially the idea was to analyze a community in which organic cotton is produced and analyze a different community in which conventional cotton is produced. In this manner the results of both communities could be compared with one another. The advantage of comparing two communities is that the impacts on the community can also be measured. Thus people living in the community where organic cotton is being produced can be questioned about the effects on their situation. However, when testing the measurement system in Paraguay I discovered that they are not separated by community. The different types of producers are living together in each community. Therefore, the approach changed to interviewing people from the target group as well as the control group in the same community. Questioning the community about the changes they experienced proved to be ineffective because it is almost impossible to attribute the changes they have undergone to the certification of some farmers. Therefore, only farmers and key-persons will be interviewed.

By interviewing the target group as well as the control group the attribution is partly addressed. However, it still remains very difficult to attribute the change in the social, economic and environmental situation to the intervention: organic certification. To be able to measure change a baseline is favoured. However, in many instances this is lacking. No baseline is drawn up before starting the intervention. To overcome this problem the respondents should be questioned about their current situation but also about their situation in the past. This is preferably about the situation of the respondents before entering the certification scheme. However, when the period between these two moments in time is very large it might be that the respondents cannot or only vaguely remember the situation before entering the certification. In this occasion a different approach should be sought. In general however, this is not expected to form a problem since most SSCG-systems emerged fairly recently. The questions have to be posed to both the organic cotton producers as well as the conventional cotton producers. In this manner the change in the situation of the two types of producers over the years can be compared and the difference can be attributed to the intervention.

However, there are still many different other external influences that can cause a certain change. By measuring the target group and the control group some of these external influences can be ruled out. For example local policies are likely to influence them both in the same manner and will not cause the difference in their situation. But also more family specific situations, apart from being part of the certification scheme or not, can influence the results.

More certainty about the cause of the measured change can be obtained through triangulation. Triangulation can be realized by contextualizing the information. This sheds light on external factors influencing the research results. Therefore, this context will be analyzed through documents, interviews with key persons and if available by analyzing previous research. The focus in this analysis will be on external

factors that might have caused the change as measured or that can explain the results, such as new policies and interview method is used.

External validity determines to which extent results can be generalized. Often this is achieved through randomizing the experiment. This might be very difficult to realize in a development context where the researcher is very likely to be dependent on assistance of local organizations. In this context the researcher often has not much choice but to interview those people selected for him. Whenever possible a list of the total research population should be used to randomly select the participants. If this is not a viable option thick description of the methods used and of the research context can assure some form of external validity. By using this method people reading the results can decide whether and to what extent they can generalize the results to other situations and contexts. This thick description should consist of very clearly describing all the steps taken in the research. The six steps research schedule presented subsequently can assist in this. Besides, an as detailed as possible contextual description should be presented to assure that all influential factors are addressed. Finally it is important that the researcher critically analyzes the results and mentions factors that might have influenced the outcome.

Table 3. Validity characteristics

Validity	Characteristics
Construct	<ul style="list-style-type: none"> • Theoretical Foundation • Multi-criteria analysis
Internal	<ul style="list-style-type: none"> • Reflexive and rigour practice • Respondent validation • Triangulation: <ul style="list-style-type: none"> - Key person interviews - Documented data - Causal explanation questions • Causality: <ul style="list-style-type: none"> - Target & Control group - Two communities - Before and after
External	<ul style="list-style-type: none"> • Random sample • Thick description • Critically reflect upon results

Reliability

Reliability should also be addressed in the measurement system. The research is reliable when it is free from random error. This is a very difficult aspect to realize in the context the measurement system will have to function in. People are social beings and showing consistency in their answers can be very difficult. There is not enough time or money to do a retest. Therefore there will always be some random error in the system. However, to overcome this problem as much as possible the research has to be designed in a consistent, systematic, logical and well documented manner. A high level of transparency towards the method used is demanded.

Table 4. Reliability characteristics

	Characteristics
Reliability	Research design: <ul style="list-style-type: none"> - Consistent - Systematic - Logical - Well documented - Cross check

5.2 Conducting the measurement

When measuring the effects of SSCG-systems not only the tool for conducting the measurement is important but also the manner of conducting the research should be considered. As discussed previously conducting research in a developing context necessitates a different approach. There are some practical difficulties, besides theoretical difficulties, in conducting research in this context which should be considered. As mentioned before these are determining the research population, the sample size, sampling, data availability and the costs of conducting the measurement. The practical elaboration of overcoming these issues will be explained subsequently.

The different steps that should be taken in order to be able to conduct the research are described in their order in time, starting with the first action and finishing with the last. While these steps are described, the different difficulties will also be discussed and accounted for.

Table 5 Six steps research schedule

Conducting the measurement: Research Steps						
	1	2	3	4	5	6
Source	Databases Local organisations Key persons	Local org. Databases Local governments	Respondents Local organisations	Respondents Key-persons	Respondents Key-persons	Results
Action	Approach contact organisation Determine key-persons Feedback questionnaire	Depict region Determine sample size Draw random sample	Test the questionnaire Adjust the questionnaire	Questionnaire Interviews with key persons	Group meetings Interviews with key-persons Dispersion of results	Entering results in SPSS Analysis Dispersion of results
Location	'Home' Research nation	Research nation/ area	Research area	Research nation/ area	Research area	'Home' Research nation

Research Steps
1. Gathering contextual information and approaching local contacts
2. Sample selection
3. Test & Adjust
4. Conduct
5. Reflect
6. Analysis & Dispersion

Step 1: Gathering contextual information and approaching local contacts

When the sustainability initiative, product and nation of choice for the research are determined general information should be gathered on this sustainability initiative and the product in its local and international context. The researcher should obtain insight in the function of the initiative, the latest trends and the important local stakeholders.

Based on this first analysis the most important stakeholders for keeping the sustainability initiative running should be approached. Preferably cooperation with one or a few local organisations should be sought. The local organisation can be very helpful in determining the precise research area, sample selection and approaching the respondents. They can assist in gaining access to different respondent groups and key-persons in the subject. Approaching key-persons can be initiated from the beginning of the project. However, it is advisable not to conduct all the key-person interviews from the beginning. During the research the knowledge of the researcher on the subject will increase and the interviews with the key-persons can become more focused. Issues can be addressed that the researcher might not have been knowledgeable about beforehand.

Moreover the expert knowledge of the local stakeholders on the subject can provide further insight in the situation and context of the sustainability initiative. Based on this knowledge the final questionnaire can be constructed and discussed with the local organisation. They can provide all sorts of feedback on the specific local situation of the sustainability initiative as well as on practical issues such as the level of understanding of the respondents. Concepts that are likely to be incorrect or misunderstood by the respondents can be adjusted. Consequently their input will most likely affect the quality of the research and its appropriateness for the local context.

In all the choices that have to be made during the research the input of the local organisation can prove very valuable. Whenever considered necessary their input can be sought.

Step 2: Sample selection

When the country of choice is known it is important to determine the region for conducting the research. There are different factors that can influence this choice. Two of the important factors are the time and money available to conduct the measurement. When there is a great amount of money at hand it would be best to conduct the measurement in different regions in the country in order to obtain a much diversified measurement. However, in many instances the product is produced in one region or there is not enough money available to conduct the research in different areas. In these instances a choice has to be made. This choice can be based on different aspects.

When there are different regions in a country where the certified product is produced the existing information on the both regions should be quickly analyzed. The reason for doing so is to assure that the research area is not a very exceptional situation which will influence the results and thus influence the image of the situation in the whole country. However this situation is better overcome by not making generalizations out of the research results to country-level but to only apply them only to the local situation.

Another very important aspect in choosing the research area is the accessibility of the area. Accessibility should be realized in two ways. First, the research area should be literally accessible in terms of infrastructure. Visiting the respondents should not take up a great amount of time and along with that large cost. Second, access to the research population is very important in terms of social relations. Contacts with local people or organisations can be very helpful in approaching the research population and making appointments to conduct the measurement, as also partly explained in step 1.

Approaching respondents via local contacts can provide trust and more interest in partaking in the research. However, it can also cause some form of bias. The contact might only introduce the researcher to either very positive or negative cases which obviously will bias the results. Besides, when the contact person is present during the interviews he might influence the answers given by the respondents. These problems can be overcome by following some simple procedures/ measures.

After choosing the research area the respondents should be sampled. The best method is to conduct a random sample. In this manner bias, for example through the influence of the contact person, can be avoided. However, the chance is relatively high that the contact person or organisation does not have an exact list of the whole research population. While he might have an overview of the target group population the chance that such a document exists for the control group population is small. Thus for the two groups a different approach should be taken.

For sampling the target group the contact person or organisation can be asked for a list with the population where the sample can be drawn from. When there is not such a list the researcher can create one in cooperation with the contact person or organisation. Often people working in the area on this specific subject have knowledge about which people are participating in the initiative. In this manner I was able in Paraguay to draw up a list with all organic cotton producers. From this list an a-select sample can be drawn. When drawing the sample for the control group, the conventional farmers, a document with this population can be sought for at local government bodies or organisations. From this list a random sample can be drawn.

When there are no lists a transactional walk or ride, depending on the size of the research area, can be conducted. The area of sampling should be determined and within this area every fifth household should be interviewed. When the area is big this could also be every 7th or 9th household. At each household the owners should be questioned first whether they produce conventional cotton. If so, the interview can proceed. If not the neighbour of the household can be questioned to participate. By doing so, a form of random sampling can be realized. However, this form of sampling has one drawback. The researcher should have a great mobility to be capable of conducting this sampling method. Not every situation offers this mobility, especially when the researcher depends for transport on others.

Independent on the actual approach taken, the method of sampling should be clearly described and the greatest manner of achieving randomized sampling in the specific local context should be sought after.

The size of the sample should also be determined. This largely depends on the time and money available. However, the smaller the sample, the less representative it is. The size of the target group as well as the control group has to be determined. The size of each group can be proportional to the population

form which it is drawn. For example, if the population from which the control group will be drawn is larger as the target group a relationship of 3 to 2 can be used.

Step 3: Test & Adjust

When the sample is drawn and the questionnaire finalized with help of the local organisation it can be tested. The focus of the test is on whether the questionnaire is clear, the concepts and questions are well understood by the respondents, the time-span is not too long and whether it provides sensible answers and thus results. This test demands some sensibility of the researcher. During the test interviews the researcher should not only be focused on gathering data but also pay attention to situational factors. Does the respondent seem to understand everything, do the responses on questions of the same topic correspond and is the respondent able to maintain its attention during the whole interview. Whenever obscurities prevail the questionnaire will have to be adjusted.

Step 4: Conduct

The fourth step is conducting the questionnaires. Depending on the centeredness of the respondents and consequently time spent on travelling three to six questionnaires can be conducted a day. For the key-persons this amount is a bit less since in general their residence is expected to be more scattered.

Before conducting the interview it is important to create a relaxed atmosphere between the researcher and the interviewee. Bringing someone of the contact institute can largely assist in this. Probably, the producer will know and trust the people working with the institute which immediately creates form of trust towards the researcher. Before starting the interview it is important to stress that there are no right or wrong answers. Additionally, it is important to mention that whenever they make critical comments towards their personal situation or the sustainability initiative that they will not be penalized. Rather their comments will be used to improve the situation.

During the interview the researcher has to be very alert. Despite the tests with the questionnaire it can still contain ambiguities which could cause biased answers. By paying attention inconsistencies in the answers most misunderstandings can be filtered out and biases prevented.

During the test in Paraguay it showed that it is very important to continually ask and reassure that the questions are understood in the right manner. This can be done by re-asking the question or re-phrasing the answer. A good way of discerning misunderstandings is by checking whether the answers to the Likert-scale questions are consistent with the answers to the open questions. If not, the researcher should verify whether the questions are correctly understood and the answers are accurate.

Step 5: Reflect

After and during the interviews are taking place the researcher should analyze the results. Most likely there will be insufficient time to put all the data in SPSS already. Therefore, the researcher should obtain a general impression of the results. These results should be presented to the target and control group.

The feedback of the results to the two respondent groups should take place during two focus group meetings. During this meeting the researcher will present the results he has gathered. The respondent group is asked to reflect on the results as presented.

During the focus group meeting it is of utmost importance to assure that the respondents feel free to openly reflect on the results. Therefore, the results should not be presented as the truth but rather as an impression of reality obtained by the researcher.

Preferably the two respondent groups are gathered in separate meetings. This is to assure that the different respondent groups do not get offended, jealous or influenced in their responses by one another.

Besides the value of calibration of the results through the group meetings it is also a manner of dispersing the results to the respondents. Accessibility and illiteracy are issues which complicate the dispersion of the results to the respondents, especially the producers. The group meetings are a good alternative manner of providing feedback on the results to these groups.

Step 6: Analysis & dispersion

When all the data is gathered the answers on the questionnaires can be filled in SPSS. After doing so the results can be analyzed. The researcher should analyze the means and make cross tabs to gain more insight in the scores on the individual Likert scale items for certain questions or topics.

To all the answers on the Likert-scale questions a value is assigned. This value ranges from 1 the worst situation to 5 the best situation. In most cases 'no/none' is the worst situation and 'very much' the best. However, in the case of pesticides used this is the other way around. This difference is accounted for by assigning a value to the best or worst situation in stead of an answer category. It is important to note this difference in assigning the values because after regrouping in means the scores cannot refer to the individual Likert scale items anymore.

In the SPSS form as constructed for this research this is already accounted for. When however, one decides to use different questions and create a new SPSS format this should be kept in mind.

An overall mean score on each sustainability aspect can be obtained by transforming the separate variables that all belong to one aspect into one new variable; the mean score. This is realized by summing up all the different scores on the questions belonging to that specific aspect and subsequently dividing them by the number of questions. This mean score is calculated for 2006 and 2009 for organic and conventional producers. In this manner the change in the score on the sustainability aspect is presented for organic and conventional producers. Again 1 is the lowest score and 5 the highest. This calculation offers a quick insight in the overall change and can be used to put in a sustainability polygon that displays the score on all the different sustainability aspects.

A bit more detailed insight in the overall score was obtained by regrouping the mean scores. This regrouping was conducted with help of SPSS as follows:

1 - 1,5 = 1
1,5 - 2,5 = 2
2,5 - 3,5 = 3
3,5 - 4,5 = 4
4,5 - 5 = 5

By regrouping the means insight is obtained in the score of the two producer groups in the two years on each Likert-scale item. For example how many producers think have obtained 'little' income in 2006 and how many in 2009. Although the mean scores provide some general insight in the change the scores on the Likert-scale items provide more precise insight in the actual change. The results on the Likert-scale items can be calculated through crosstabs in SPSS and subsequently exported to Excel in which the data can be transformed into graphs to assure clear presentation.

To further specify the results both calculations as presented so far can be conducted on the individual questions or a grouping of questions related to a specific sub-topic. Besides the questions concerning the level of change the respondents perceive they have undergone can also provide more insight in the precise situation.

The questions in the questionnaire concerning the cause of change are used to gain more insight in the causal relations. Crosstabs can be used to analyze how often the different causes occur. Again excel can be used to realize graphical presentation.

There are also some questions that do not consist of Likert-scales neither are aimed at obtaining a causal explanation. These questions are those with 'yes/ no' answers and numerical answers. For the 'yes/ no' answers the percentage of prevalence can be calculated. The numerical answers either speak for themselves or can be used to calculate mean change.

When using quantitative data analysis such as SPSS one of the important issues is whether the data obtained is statistically significant. Significance tells something about the chance that the results are caused by coincidence. It tells something about the strength of the causal relation. Therefore, it is important to also test the statistical significance of the results.

When all the data is analyzed and documented the results should be dispersed to all stakeholders involved.

6. Sustainability aspects framework

When creating a measurement system not only the methods used are of importance. Determining what to measure is just as important. The measurement system is meant to function for different SSCG-systems with a focus on different product groups. Each SSCG-system is likely to have its own specific goals. However, the goal of the measurement system is to be comparable amongst different systems and regions. Consequently, the aspects used for the measurement system are not based on the goals of one specific SSCG-system but rather on an overarching concept. This overarching concept is sustainable development. Although SSCG-systems have their own specific goals, in general one could state that the reason for creating or being part of a SSCG-system is to be able to contribute to sustainable development. Therefore, from this concept the aspects used in the measurement system are derived.

So what does sustainable development actually mean? The concept sustainable development is widely used. Since the concept emanated in the 1980s in the Brundtland Commission report the term has been used more and more and the concept has evolved. The concept is being interpreted in different ways, as different actors are trying to give their own interpretation to the concept. This ongoing development of the concept is part of an ongoing dialogue concerning the concept. To be able to derive aspects from the concept it is important to use a clear definition of what sustainable development should comprise. This also assists in assuring that not anyone can adjust it to fit its own purposes (Kates et al., 2005). The definition of sustainable development that is used, is that as first stated by the Brundtland Commission. They define sustainable development as:

“Humanity has the ability to make development sustainable to ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs” (UN, 2009).

In this definition economy is not perceived as the only catalyser of development. Ecology and economy are an interwoven net of causes and effects which cannot be approached separately if we want to live within natural boundaries. The realisation came that international economic relationships are causing a large share of the problems concerning environmental management in developing countries (Nelissen et al., 1997). Therefore, an important aspect of the concept concerns the goal of not only assuring economic development for poor nations but also realizing an equal division of resources (Kates et al., 2005; Nelissen et al., 1997).

The definition of sustainable development as posed by the Brundtland Commission provides us with a concept that is understood intuitively. However, it remains difficult to express in concrete operational terms (Kates et al., 2005; Labuschagne et al., 2005). Since the 2002 World Summit in Johannesburg more direction was added to the concept by dividing it into the three pillars. An agreement was formed that there was “a collective responsibility to advance and strengthen the interdependent and mutually reinforcing pillars of sustainable development – economic development, social development and environmental protection- at local, national, regional and global levels”(UN, 2002). However, there is no universal agreement on the boundaries of these three pillars and their meaning. This is especially true for the social pillar which is characterized in many different manners (Kates et al., 2005). Clarifying the boundaries between the three pillars is important in creating a clear aspects framework (Hutchins and Sutherland, 2008).

Since there is no agreement yet on these boundaries, Bertil Abbing and I collaborated to create our own definition of the three pillars of sustainable development. The environmental pillar comprises all direct impacts on the environment on all levels (local, regional, national and global). The economic pillar consists of two parts; the economic sustainability and socio-economic sustainability. The latter comprises all aspects which directly impact the economic growth of the community under concern. For the former Labuschagne’s (2005: 377) definition will be used; “Internal operational initiatives that directly contribute to the overall profitability of a company”. The social pillar comprises all aspects which influence the development of society towards a good quality of life. Whereas the environmental pillar comprises aspects on all scale levels, the socio- economic and social pillar only focus on those areas in which the certification scheme is active. The reach of this area is determined by the place of settlement of the workers or farmers taking part in the certification scheme. In this manner seasonal workers are also taken into account.

The theoretical basis for these three definitions as well as transformation of these definitions into aspects will be explained in the subsequent text.

However, before continuing it is important to explain that the focus in this research is on creating an aspect framework, and not an indicator framework. Aspects are a broader description of often several indicators. For example the aspect ‘good labour conditions’ can exist from the indicators ‘fair incomes’, ‘good working conditions’, ‘safety measures’, ‘employment compensation’ and so on. In some cases indicators can be perceived as provisions necessary to realize a sustainability aspect. Part of realizing good working

conditions is assuring fair income. The choice for focusing on aspects is because it gives more flexibility to apply it to different supply chains and situations. In each supply chain and situation the indicators that together determine the impact on an aspect could differ. A large difference is for example whether the framework is used for farming activities or factory labour. Consequently the list of indicators could be endless. Since it was not the main goal of this research to create an inexhaustible indicator framework, and it was the goal was to create a framework that can be used in different sectors and areas, an aspect framework was created. This framework was subsequently adjusted to the cotton sector. Finally, several key-persons working in this sector have chosen the most important aspects from the adjusted list. The aspects that were left are operationalized into measurable indicators.

In the subsequent text the word 'indicator' is used when talking about indicator frameworks that are studied and discussed. The reason being that the authors of these frameworks talk about 'indicators', rather than 'aspects'. However, when discussing our aspect framework that was created the concept aspect is used.

6.1 Theoretical foundation

Many different indicator frameworks have been developed in an attempt to display a comprehensive set of indicators. The indicators used in these frameworks are often based on other already existing frameworks. In almost all cases there is no theoretical foundation for the choice of indicators used (Briassoulis, 2001) or the choice remains unclear.

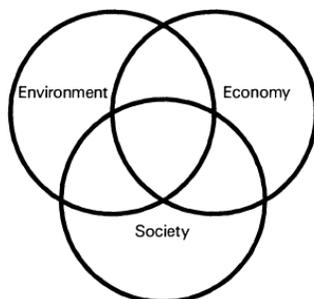
In this research the aspects do have a theoretical foundation. However, experience from practice is not neglected. Therefore, besides creating a theoretical foundation a few already existing indicator frameworks are analyzed to assure impact from practice as well.

The theoretical foundation lies in sustainable development theory as well as development theory. The reason for using development theory is threefold. On the one hand the aspect framework will have to function in a developing context. This context brings along context specific goals to be attained and difficulties in data availability for which the framework has to be adjusted (Hutchins and Sutherland, 2008). Secondly, making use of development theory can provide more insight into the objective of achieving sustainable development for the special development context. Keeping the context in which the aspects will have to function in mind can strengthen their utility. Finally, sustainable development and developing theory have become largely intertwined since the 1980s. This is visible in definition of a sustainable livelihood as defined Chambers. A summary of this definition is presented by Carney (1998):

A livelihood comprises "...the capabilities, assets (including both material and social resources) and activities required for a means of living. A livelihood is considered to be sustainable when it can cope with and recover from stresses and shocks and maintain or enhance its capabilities and assets both now and in the future, while not undermining the natural resource base."

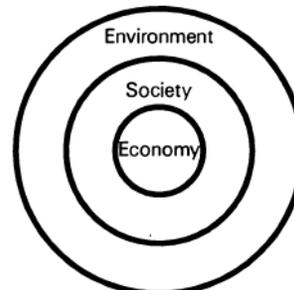
This definition does not only address the overlap between sustainable development and development theory but in the same time addresses the interrelatedness between the three pillars of sustainable development.

Fig. 3 Three ring Circus



Source: Levett, 1998 p. 295

Fig 4. Russian Dolls



Source: Levett, 1998 p. 296

Traditionally the interrelatedness between these three pillars is displayed in the 'three ring circus model' as also discussed by Custance and Hillier (1998). However, there are claims that the interrelatedness in this model is not enough. The interdependency of the three pillars is much larger (Levett, 1998).

The environment can be considered a prerequisite for economic and social life. Besides, the economy is a social construct which only functions because human kind has created the assumptions, behaviours and institutions that make it work. The reason for creating this social construct is that we believe that the economy can assist us in achieving our needs. Based on this Levett states that: "sustainability is about ensuring that human society lives within the environment's limits – and that the economy meets society's needs" (Levett, 1998: 295). This assumption does not fit within the 'three ring circus model' but is presented in the "Russian Dolls" model. The most important questions based on this assumption are whether we are living within environmental limits and if we are achieving good quality of life (Levett, 1998).

6.1.1 Social pillar and Economic pillar

The second part of the question as posed by Levett (1998) – achieving good quality of life- forms the basis of the definition of the social pillar. This goal is also in agreement with the goal of sustainable development as originally posed by the Brundtland Commission (Nelissen et al., 1997). However, it only provides a direction in which the different aspects should point but not the aspects themselves.

Creating an aspect framework to measure good quality of life has proved a difficult task. Many different types of indicator systems have been developed in an attempt to measure quality of life. All sorts of aspects are deemed important in attaining good quality of life. However, many indicators systems have encountered difficulties in providing an all encompassing indicator system. Most systems are only addressing parts of quality of life, such as poverty, human development and human security, and are leaving out others. Furthermore, there is an enormous overlap of themes. Examples of these indicator frameworks are the OECD Pressure- State- Response- approach, the Human Development Index, the Millennium Development Goals, the Gender related indexes, the Human Poverty index, the Index of Human Security, the Disability- Adjusted Life Year and the Health- Adjusted Life Expectancy. Furthermore, all these systems are measuring sustainability in a quantitative manner and depend on the availability of documented data and statistics (Hilderink, 2004).

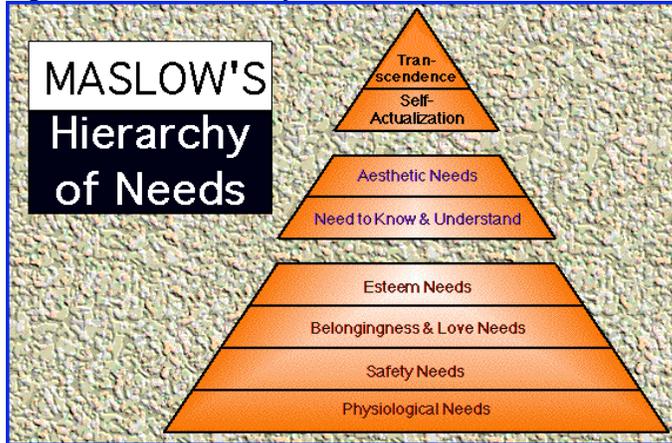
Obtaining these types of data in a (local) developing context is very difficult. If they are available it is most likely to be on national level. Consequently, within the aim of this research in which we focus at the local and regional level these indicator systems in place are not very useful. Thus a more all encompassing aspect system had to be created which has a local focus and which does not solely depend on the availability of statistical data.

In order to create this system the goal of achieving good quality of life had to be split up into several measurable aspects. In order to do so a clear definition of 'good quality of life' had to be put forward. Two classic authors that have done an attempt of determining what 'good quality of life' is are Abraham Maslow and Herman Daly.

Maslow has created a Holistic Dynamic Needs Hierarchy in which human's needs are presented, from their most basic needs such as food and shelter up to life fulfilment needs. He describes in eight sequential levels the needs of individuals (Hilderink, 2004). These needs are divided in two groups; deficiency needs and growth needs. Level one to four falls within these deficiency needs and level five to eight are growth needs. Within the deficiency needs a person needs to meet the lower level before continuing to the next level. All the deficiency levels need to be met before being able to fulfil the growth needs (Huitt, 2004).

Only the first two levels have a focus on aspects outside of the individual, such as hunger, thirst, bodily comforts and safety. From the third level onwards the needs concern aspects such as love, needs to understand, self-actualization and self-fulfilment. Therefore it is difficult to be used for defining sustainable development (Hilderink, 2004).

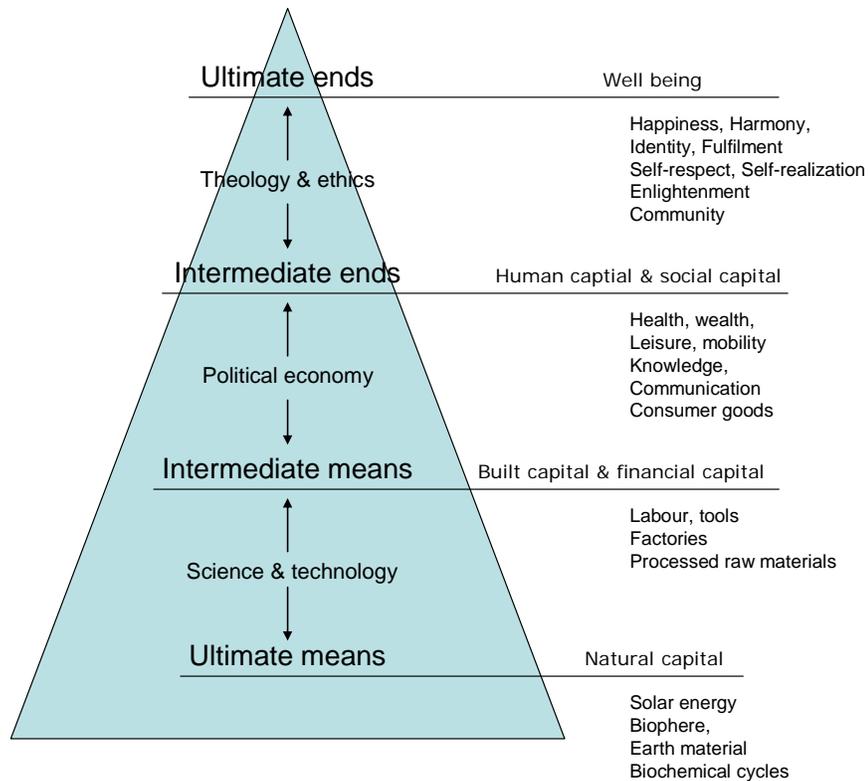
Fig. 5 Maslow's Hierarchy of Needs



Source: Huitt 2004: 1

Daly's triangle has a more practical approach and is more focused on earthly matters, although not excluding personal development needs. He makes a distinction between means and ends. Within these two categories he makes another distinction between intermediate and ultimate means or ends (Daly, 1990). The ultimate means are the natural environment and are in Daly's triangle approached as a precondition for human life. In his triangle politics and economy, the intermediate means and ends, are a vehicle for reaching the ultimate end: well being. Sustainable development in his triangle is about how ultimate ends can be reached for the largest number of people and in which manner the ultimate means can be most efficient and sustainable used for reaching this goal (NSSD, 2003). As such he has a more societal approach.

Fig. 6 Daly's Triangle



Source: Hilderink 2004: 36.

Daly's triangle is both all encompassing in terms of indicators included and also creates the possibility of applying it to a local level. He does so by linking the World Bank capital approach to the lower three levels of his pyramid. This capital approach is a very practice based approach used to determine the sustainability of livelihoods. Linking the capital approach to the different levels offers a more qualitative, local indicator system in which good quality of life, the ultimate end, can be measured through access to the different capitals (Hilderink, 2004).

The first level consists of natural capital which entails natural resources. The second level is physical capital and human capital such as production equipment. The third level consists of human capital and social capital such as health, wealth, knowledge and communication. The final level, 'well being' is not part of the capital approach but describes the ultimate end of the pyramid. Well being is about self- respect, happiness and enlightenment amongst others (Hilderink, 2004). This ultimate end is in agreement with the original aim of sustainable development and the second part of the sustainability question as put forward by Levett (1998); to attain good quality of life. Moreover, Daly believes that the environment is a prerequisite for human life.

The capitals approach as linked to the means and ends in Daly's triangle offers an opportunity to translate the goal of achieving 'good quality of life' into more practical terms. The capitals linked to the triangle can form a fundament for the aspect framework. However, we must note that the capitals only refer to the bottom part of the pyramid as proposed by Maslow and neither includes the ultimate ends as described by Daly. Nonetheless, I believe that the goals of SSCG-system are more focused on achieving these 'deficiency needs' which relate to matters such as bodily comfort, hunger and thirst rather than the growth needs, in Daly's pyramid the ultimate ends. Growth needs seem to be on a more personal development level and are hard to be influenced through value chains. Therefore, the capital approach is a viable option to assess social and economic sustainability. In the following the assets will be described which together form the different capitals. These assets can consequently be converted into aspects. Before explaining how this conversion is conducted some more background information on capital approaches will be presented.

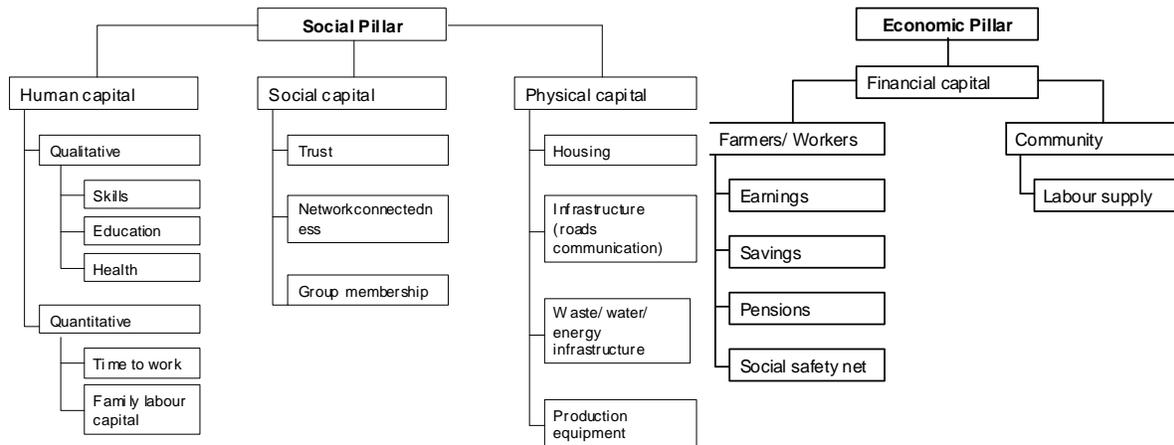
Sustainable Livelihoods Approach

Using capitals to describe the sustainability of livelihoods is not only conducted by the World Bank. The concept has its roots in the work of Chambers. The concept is interpreted and used by different organisations and academics in the development field. An often used capital approach is described in the Sustainable Livelihoods Approach. This theory tries to capture human well-being by analyzing their access to different capitals. At household, community and societal level access to assets together forms a stock of capital. People and social units will need to be able to call upon these stocks of capital to assure their sustainability (Rakodi, 1999).

The five capitals that together determine the sustainability of a household are financial, physical, natural, human and social capital. Natural capital is said to comprise the natural resource stocks from which the household can obtain resources that are useful to the household. This includes land, water and other environmental resources. Often these resources are common pool resources. Physical capital or produced capital concerns basic infrastructure such as communication networks, shelter, water, and energy provision. It also concerns different production equipment and means assisting households to pursue their livelihood. The human capital of a household is constrained by several aspects. These comprise both quantity and quality. The level of education, skills and the health status of household members is an important determinant of the status of human capital. However, also the demands of household maintenance are influential. The time available to household members to engage in income earning activities is determined by the technology for and access to household provisioning activities, the household dependency ratio and the stage in the household- life cycle. Financial capital is not only about earnings but also about remittances, savings, credit and pensions (Rakodi, 1999). Social capital is defined as 'the rules, norms, obligations, reciprocity and trust embedded in social relations, social structures, and society's institutional arrangements, which enable its members to achieve their individual and community objectives' (Narayan, 1997: 50). The importance of social capital was made particularly apparent by Putnam (1993). Through his study of civic traditions, democracy and regional development he showed that social structures, networks and organizations are the decisive factors in regional governmental effectiveness and economic performance (Putnam, 1993). Social capital can be obtained at household, community and societal levels. It is a relational concept and therefore cannot be measured in its own right (Rakodi, 1999).

Based on these descriptions the following aspects can be attributed to the social and economic pillar:

Fig. 7 Indicator framework based on SLA



6.1.2 Environmental pillar

Although the livelihoods approach also pays attention to the environmental pillar in the form of natural capital this attention is very minimal. It is insufficient to answer the second part of the question as posed by Levett (1998); how to live within environmental limits. Therefore, environmental literature has been studied in order to create an aspect framework for the environmental pillar.

Determining environmental limits is very difficult and doing so could be a research subject in itself. Since environmental limits remain debated a general approach is to try to minimize the negative environmental impacts. This is not any different for the supply chain. Therefore, the focus when creating the aspect framework was at minimizing environmental impacts. In the light of this research this translates to determining which aspects are of importance when minimizing negative environmental impacts from the supply chain. Different methodologies are present to determine environmental impacts such as Material Flow Accounting, Environmental Risk Assessment, Input-Output Analysis and Life Cycle Assessment (LCA). Only the latter offers a good methodology to take the different life phases of a product into consideration (production, use and waste phase) and thus can be used to base the framework upon (Hutchins and Sutherland 2008; Bauman and Tillman 2004). Furthermore, LCA is a widely acknowledged and the most used methodology for measuring ecological impacts and therefore it is becoming increasingly effective (Hutchins and Sutherland, 2008; Guinée et al., 2002).

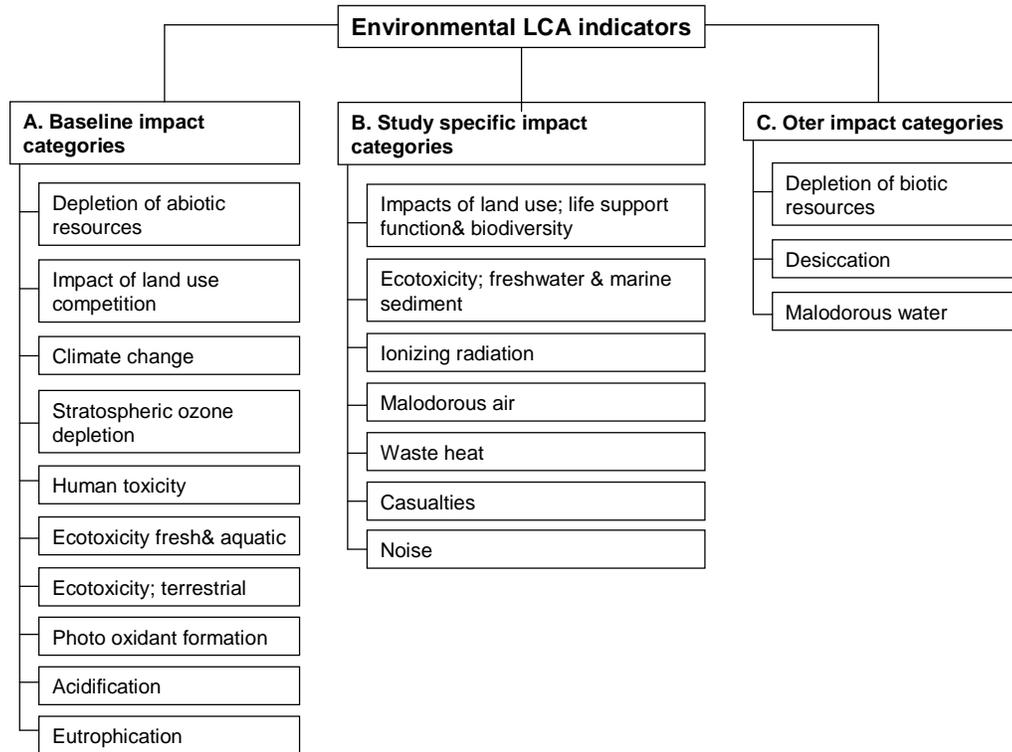
The aspects which represent the environmental pillar are based on those used in the LCA as described by Guinée et al. (2002). All impact categories as required by ISO 14042 (concerning environmental management, life cycle assessment and life cycle impact assessment) standards are met in their framework. Therefore, it is likely to give a good representation of what is strived for in attaining environmental sustainability in SSCG-systems.

The different indicators as mentioned in the LCA are categorised according to the United Nations Commission for Sustainable Development (UNCSD) themes to be able to provide a more clear structure amongst the different aspects included in the framework. These themes are; atmosphere, land, water, biodiversity and natural hazards (UNCSD, 2007). The UNCSD indicator framework is based on and adjusted to years of experience in developing and testing their framework. It provides a clear, all encompassing and straightforward categorisation which brings a clarifying order in the different sustainability aspects.

The fifth category as used by the UNCSD, natural hazards, is left out of this research because the supply chain cannot influence natural hazards. In the eye of the definition of environmental problems natural hazards are problems not caused by humans and thus we cannot impact their occurrence.

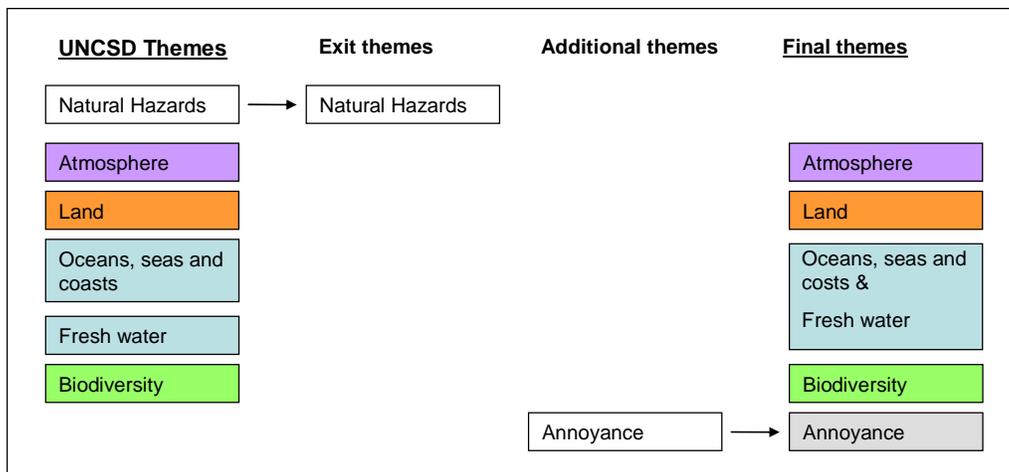
In the following the UNCSD and LCA frameworks are presented.

Fig. 8 Environmental LCA indicators



Source: Guinée et al., 2002: 68

Fig. 9 UNCS D themes & Final themes



When analyzing the aspects as presented by Guinée et al. (2002) it becomes clear that noise and odour are difficult to group under the five themes as withdrawn from the UNCS D framework. Therefore, a new category is formed under the heading 'annoyance'. This theme concerns annoyance for humans amongst which odour and noise can be grouped but also sight hindrance is an important aspect. This last aspect will thus be added to the framework under the header, landscape pollution.

6.2 Comparison

For all three sustainability pillars aspects are selected with a theoretical foundation. This gives more weight to the aspects chosen. However, as explained earlier, there are a lot of different indicator frameworks already existing. Although their foundation is unclear or seems not theoretically grounded it would be a pity not to look them over and analyze whether there are any aspects which should be added to the frameworks as presented so far. Moreover, the aspects chosen for the economic pillar and social pillar currently only have a foundation in development theory. Taking on an environmentalist point of view might provide new insights. Therefore, the indicator frameworks as presented so far are to be compared with previously constructed frameworks.

6.2.1. Comparison economic pillar and social pillar

The aspects proposed so far in the social and economic pillar have their main foundation in development theory. However, there are also different attempts done to create a social and economic indicator framework from an environmentalist perspective. In order to assure a full coverage of aspects originating in development theory as well as sustainable development theory the aspects based on the theoretical foundation as described above are compared with three indicator frameworks. These indicator frameworks are used amongst other reasons because they are three of the most extensive frameworks encountered after a literature review.⁸ The first is the social sustainability indicator framework as described by Labuschagne et al. (2005). This indicator framework is based on the analysis of four indicator frameworks (the Global Reporting Initiative, the United Nations Commission on Sustainable Development Framework 2001, the sustainability Metrics of the Institution of Chemical Engineers and the Wuppertal Sustainability Indicators) complemented by additional Social Impact Assessment guidelines and frameworks as well as literature on corporate social responsibility. The result is a thorough framework which displays a broad array of social and economic indicators (Labuschagne et al., 2005) (see annex 2).

The second indicator framework that is used is the United Nations Commission on Sustainable Development Framework 2007. This latest version of the UNCSD framework is also used to provide the themes for the environmental indicators. It is a well-grounded and extensive framework (UNCSD, 2007)(see annex 3).

The third framework is based on the indicators as presented by Organic Exchange in their Key Performance Indicators for Self-Assessments (Ferrigno et al., 2008 and 2009) (see annex 4). While writing this thesis they were in the process of renewing their indicators. The environmental pillar and the economic pillar consist of new indicators while the social pillar was still in development so is based on some new indicators that were already finished and some old indicators.

By using a framework with an academic background and two with a more practical background, of which one focuses on national situations and the other on more local situations, a full representation of different existing frameworks is presented.

The framework for the social pillar and those concerning the economic pillar are analyzed at the same time. The reason for doing so is that there is a large overlap of aspects between the two pillars. Furthermore, the division chosen between the two pillars as conducted within the analyzed frameworks might differ from the division used in this research. Therefore, the two pillars are analyzed together. Afterwards the different aspects are divided again between the social and economic pillar. The aspects not already included in my own framework and of interest in achieving sustainable development in a SSCG-system context are circled.

Analyzing the frameworks shows that there are different aspects missing in the frameworks for social and economic aspects as constructed so far. However, it is important to note that different frameworks are each constructed with their own goals. These goals might differ from that of creating a sustainable supply chain and thus their impact areas might also differ. Therefore, some of the aspects presented in the frameworks are not deemed relevant for the framework created in this research and consequently are not taken up in the framework. These aspects are found within the UNCSD and Labuschagne's framework and are: demographics, macro economic development, public finance, tourism, trading opportunities and external financing. External financing in the UNCSD framework is mainly about Net Official Development Assistance and foreign direct investment and remittances on a national level. This is not likely to be influenced through the supply chain. Also trade is meant to be on a global level. These impacts and the other aspects

⁸ Articles reviewed; Briassoulis (2001), Guinée et al. (2002), Hilderink (2004), Hutchins and Sutherland (2008), Kates et al. (2005), Labuschagne et al. (2003), Labuschagne et al. (2005), Levett (1998), Organic Exchange (2008), Piorr (2003), Rakodi (1999), Schmidt et al. (2004), UNCSD (2001), UNCSD (2007), Veleva and Ellenbecker (2007)

mentioned might play a role in value chains on the global scale, but it is unlikely that for any of the aspects mentioned local impacts as a consequence of sustainable supply chain management will be noted in a direct manner. However, it is important to mention here that the trading opportunities can be impacted on a local level, as I learned during my research. Therefore, I propose to add this to the aspect framework. Because I only learned this during the research it is not part of the aspect framework which I used in Paraguay.

Furthermore, some of the aspects presented in the frameworks are provisions instead of impacts. Since the aim of this framework is assessing the impacts on social, economic and environmental sustainability and not whether the necessary provisions are adhered to, only aspects which are expected to be impacted will be part of the framework. Moreover some of the aspects are already represented in other aspects. Sanitation is for example a provision for health and good contracts of working conditions and therefore does not need to be added. These aspects that are either provisions or already represented in other aspects are good contracts, sanitation, information and communication technologies, information provision, internal and external stakeholder influence in decision-making, research and development, monitoring, legislation and enforcement.

Finally, before discussing the adding of some of the aspects, it is important to note that the aspects included in the framework will present basic categories. In order to assure that the framework will not become too extended and to provide room for adjustment of the categories to each specific situation these categories will not be divided in specific indicators yet. Thus aspects mentioned in the frameworks which are only a specification of a general category will not be included as well but might be used in the operationalization of the aspects into indicators.

Two of the aspects mentioned in both the Organic Exchange framework as well as Labuschagne's framework are information provision and stakeholder influence in decision making. Although these two aspects are provisions instead of aspects that are impacted, they do point towards important impacts to be included in the framework: fairness, equity and awareness. Moreover, they are not represented in other included aspects yet. Creating awareness of the sustainable system one is acting in is of utmost importance to realize the sustainability goals. In the Organic Exchange framework this aspect is an often returning subject in the form of creating understanding. The importance of this aspect will be explained further when discussing the environmental aspects. Furthermore, fairness and equity is not only important in the light of social relations as mentioned in the next paragraph but also in economic relations. Equity in influencing decision making is part of that. This particular aspect of equity would be better grouped under the economic pillar since it is influencing the economic situation of the participants. However, it is not part of financial capital. Therefore there is a new category added within the economic pillar which covers fairness and equity as well as ownership which will be explained later. The new category will hold the name 'economic relations'.

The importance of equity is also acknowledged by Labuschagne et al. and Organic Exchange. Equity in Labuschagne's framework encompasses racial and gender equity, while Organic Exchange also addresses equity in profit distribution. The UNCSO does not mention equity as such but does stress the importance of abating government corruption. While government corruption will not be an issue in this research the aspect of corruption can be conceived part of the equity concept since a corrupted system is unfair. The concept of equity is an important concept since the first report by the Brundtland Commission on sustainable development. Equity of use and benefits of resources particularly received attention. Therefore this aspect should definitely be included in the framework.

Finally there are some aspects only mentioned in one of the frameworks. Labuschagne et al. (2005) discuss the importance of security and sensory stimuli. Security can concern product security (safety of the product), internal security (safety during work) and external security (potential danger of production process for surrounding communities). Sensory stimuli are about aesthetics, odour and noise. These aspects were not mentioned in the social and economic pillar but form part of the environmental pillar. Therefore, they are not included to the social or economic framework.

The UNCSO displays different indicators concerning material consumption patterns as indicators for economic sustainability. The three indicators mentioned in this aspect, energy use, waste generation and management and transportation are in my perspective part of the environmental pillar and therefore discussed later in the text.

Organic Exchange also displays different aspects not included in the framework yet. For the social pillar these are investments in profit distribution, child labour, labour conditions and food security. The latter aspect is often not an issue in a western context but in a developing context where severe poverty is a daily issue including food security is important. The same is the case for child labour which is still an often prevailing issue in developing countries. Finally the aspect 'investments' concerns how profits are invested. This might assist in addressing if profits are invested sustainably and in an equal manner along all the actors in the supply chain. This aspect is also displayed in the economic pillar as presented by Organic Exchange in a more extensive manner. Referring back to the definitions attached to the three pillars, the aspect of social and community development can be grouped under the social pillar. The investments influence the development of a community towards quality of life while investments in diversification and processing influence the economic position of the farmer and thus can be grouped under the economic

pillar. Ownership is another aspect which can be added to the economic pillar. Ownership can play an important role in whether benefits are obtained by the farmer or other actors in the supply chain. It thus also directs to the importance of fairness and equity as mentioned previously.

Labour conditions address the circumstances people will have to work in. It might display the provision 'good contracts' which is mentioned in different frameworks. Partly the indicators proposed under labour conditions by Organic Exchange are already addressed by aspects such as equity between men and women, racial equity, safety measures etc. However, some indicators are not specifically addressed which are important such as working hours and facilities present at the work place and can thus be addressed under the aspect of labour conditions. However, they will not be added as an aspect on their own.

A final remark after analyzing the different frameworks is that the Organic Exchange framework has a much more practical character than the other two frameworks. The indicators provided by Organic Exchange are often already divided in sub-indicators which are easier to operationalize. Therefore, for operationalizing the final into indicators these indicators presented by Organic Exchange were used.

Comparing the current framework with the UNCSD, Organic Exchange and Labuschagne's framework showed that the capitals as described in the Sustainable Livelihoods Approach do not cover all aspects considered important when measuring impacts on sustainability. This leaves the question whether the SLA is an appropriate approach to determine the important aspects of sustainability. According to me it can still be perceived as such. It is interesting to note that all lacking aspects in the capital approach do not concern assets which can be obtained by an individual but aspects imposed from outside. Good labour conditions or influence in decision making are not assets which one has access to and can use to their benefits on their own. The aspects can be perceived as assets which have to be obtained and maintained through (economic) relations with others and are especially relevant from the perspective of SSCG-systems.

6.2.2. Comparison environmental pillar

The aspects chosen so far in the environmental pillar are mainly based in environmental theory since this proved much more extensive than presented in the developmentalist SLA. Herein the different academic indicator frameworks are studied. Subsequently, the aspects derived here from are compared to those as presented by Organic Exchange in their KPI- system as well.

When comparing the framework as presented by Organic Exchange with the proposed framework so far it once more becomes obvious that the Organic Exchange framework uses more operational terms. However, most aspects are covered in the proposed framework only in more abstract terms. Some aspects though are not covered yet (see annex 5).

Water is taken into the framework but no attention is paid to the access to water and its depletion level. Both aspects are of great importance to sustainability as also shown in the natural capital in the SLA and thus should be added to the framework. Furthermore, there is no attention paid to energy use as such. This theme is also discussed in the UNCSD indicator framework but under social impacts. In the same part waste generation and management and transportation are discussed. Transportation is already mentioned under the aspect of infrastructure. However, energy use should be included in the framework. In this research energy use is interpreted to consist of non-renewable energy and mineral energy. Together with the aspect water this can form an additional theme in the framework; depletion. Waste generation is a somewhat special aspect. Waste generation is amongst others one of the aspects causing biodiversity loss. It is an indicator which can be used for measuring the aspect of biodiversity loss, but it is not an environmental aspect in itself that can be impacted.

Dependency on external inputs (such as pesticides and fertilizers) might seem a subject not mentioned previously. These external inputs concern inputs necessary to be able to cultivate cotton. According to me this dependency actually refers back to the biodiversity and soil fertility since fertile land with good biodiversity does not need any or will be impacted by additional inputs. Therefore, this is not added as an additional aspect.

Finally, an important aspect which came to the front analyzing the self assessment systems as created by Organic Exchange is 'understanding'. Understanding the process of organic cotton production, understanding the different impacts it might have and understanding how the possible solutions function. Without this understanding realizing sustainable cotton production is far more difficult. Perceiving understanding as an aspect might be difficult since it can also be partly perceived as a prerequisite. Therefore, it can be perceived as part of the knowledge and skills of a producer.

Based on the theoretical background and the comparison with existing indicator frameworks the following aspects are selected (Fig, 10, 11 and 12). These aspects can be used as a starting point for future research on sustainability. Based on the specific product and or context certain aspects can be left out and they can be converted into case specific indicators.

In the subsequent chapter the indicators are further tailored to the specific context of cotton production so they can be used for testing the developed measurement system in Paraguay. However, before presenting this adjusted indicator system some background information on the cotton chain will be presented.

Fig. 10 Social aspect framework

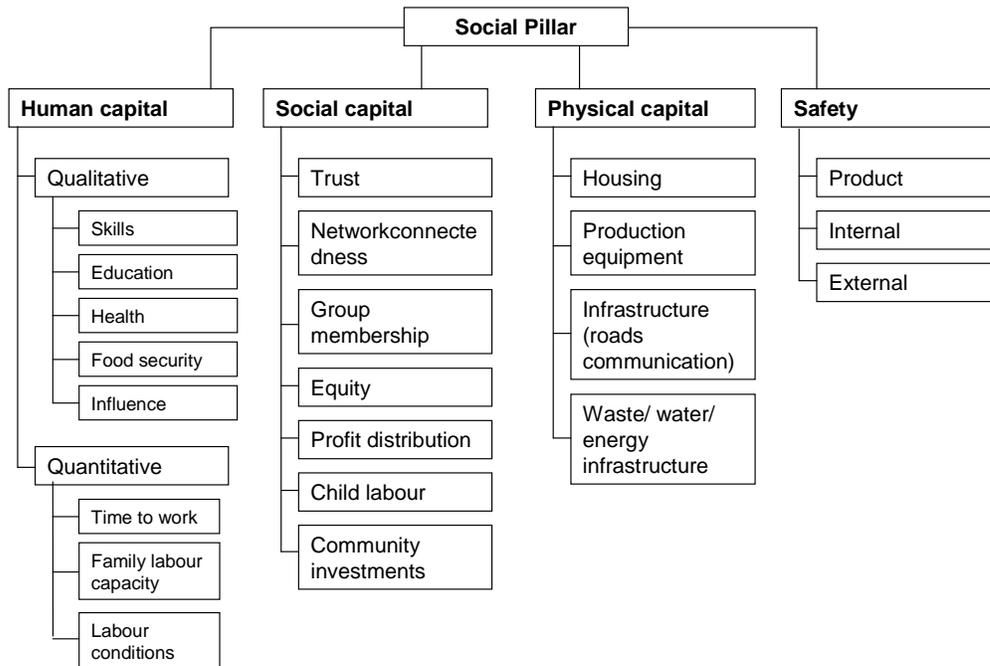


Fig. 11 Economic aspect framework

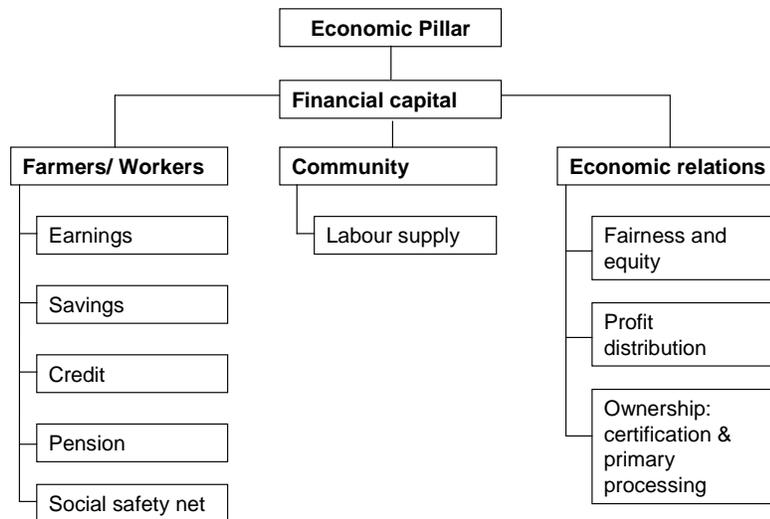
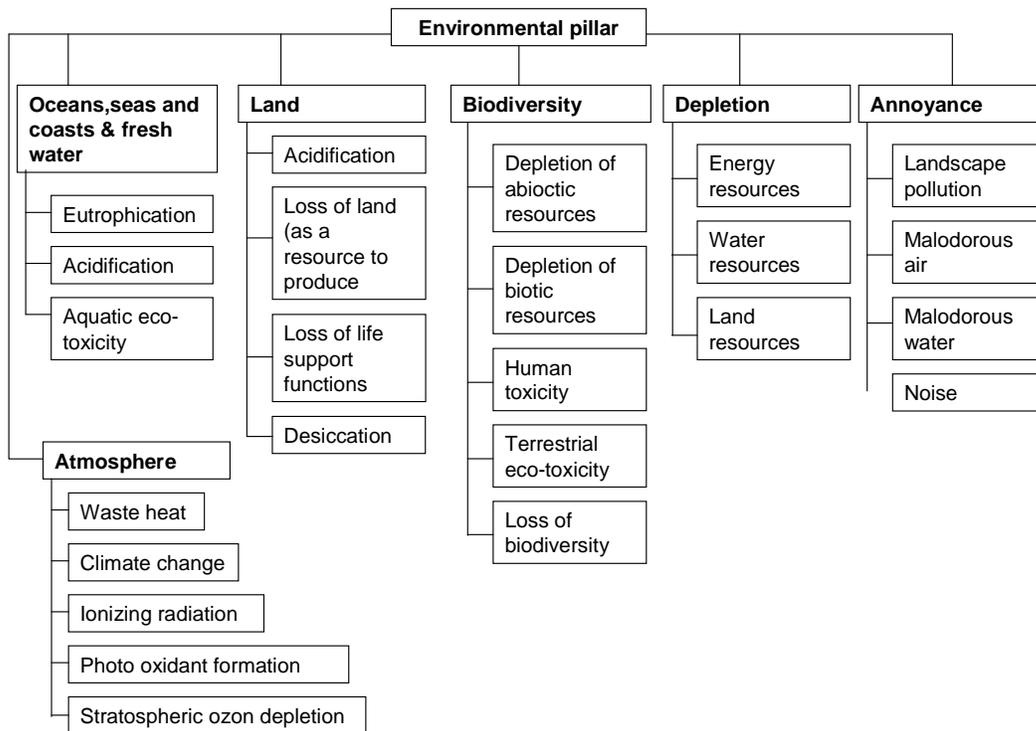


Fig. 12 Environmental aspect framework



7. Organic cotton certification

After creating the measurement system based on a theoretical background and lessons from practice the empirical part of the research started. The measurement system was tested to gain insight in its strengths and weaknesses.

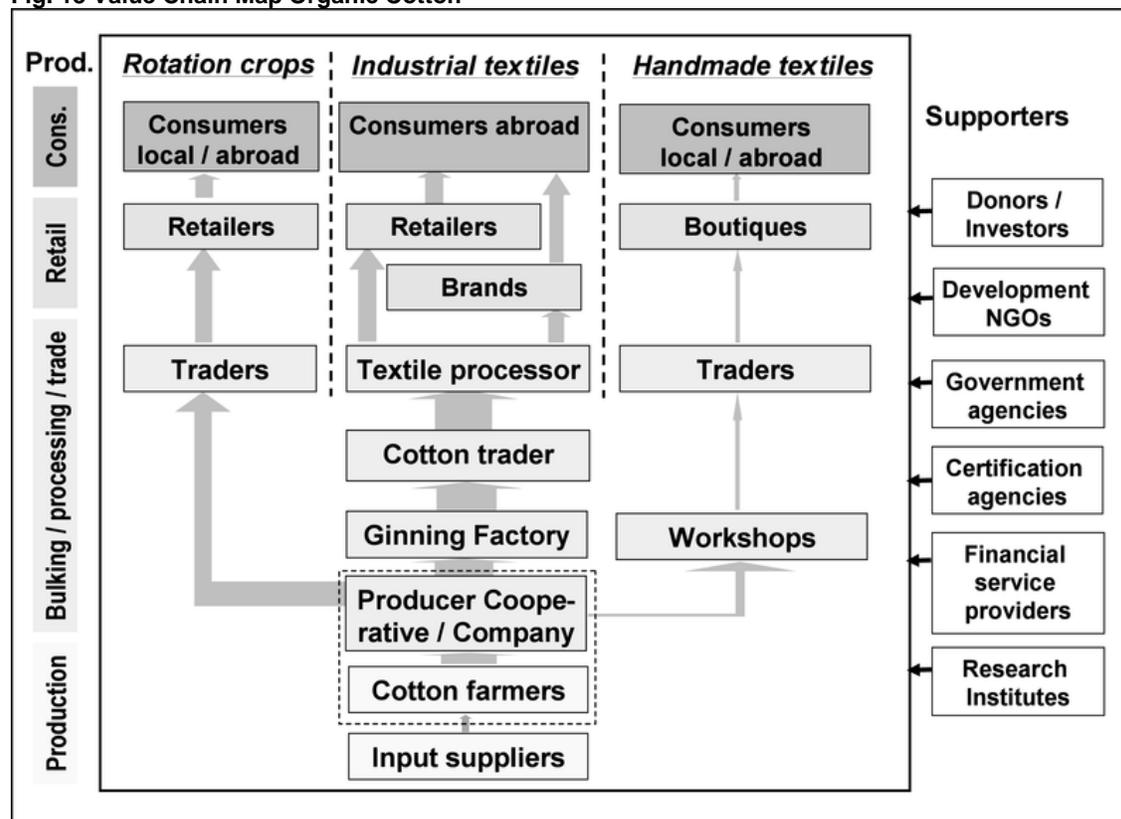
The system was tested on organic cotton certification in Paraguay. Before being able to conduct the test the aspect framework as presented needed to be tailored to the specific local cotton situation. It is important to know that the cotton producers in Paraguay are small-scale farmers that have a very low level of mechanization in their cotton production process.

Before presenting the aspect framework as used for the test-case first some general background information on the cotton chain is presented.

7.1. The cotton chain

The Cotton chain is long and complex. There are different end products in which the cotton can be transformed. Dependent on this end goal the cotton will follow one of the following roads through the value chain.

Fig. 13 Value Chain Map Organic Cotton



Source: Creative Commons 2010.

The chain presented here only displays the steps at which value is added to the product. Middlemen and exporters and importers are not explicitly mentioned in the chain presented but are mainly part of the retailers or traders.

In 2007 cotton was produced in 90 countries. It is produced in North and South America, Africa, Asia, Europe and the Middle East. The cotton sectors in the developed and developing countries differ notably. In the developed countries it is very unusual to encounter full-scale vertical integration, whereas developing countries still exhibit vertical integration from growing the cotton to marketing the end products.

Besides, the size of cotton farms in developed countries is larger and the level of mechanization throughout the whole production process is higher (UNCTAD, 2010).

Production trends

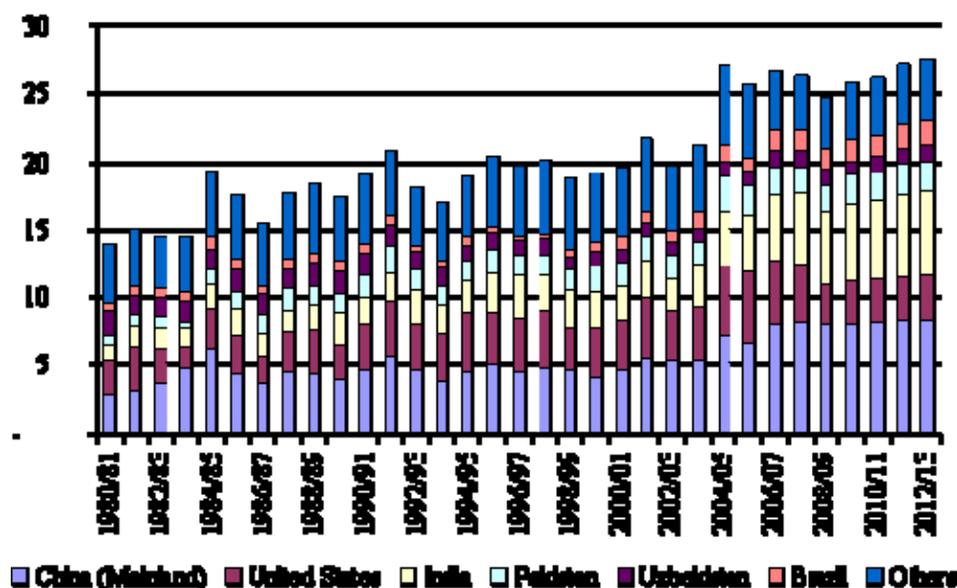
Cotton's share in textiles fibres declined significantly since the 1970s. In 1960 cotton's share was 68,7% against 21,8% for chemical textiles. In 2002 these proportions have changed to 39,7% cotton and 57,7% chemical textiles. Nevertheless, the role of cotton in natural fibres remains non-debated (UNCTAD, 2010).

The largest expansion in production area occurred after the Second World War. Between 1945/46 and 2006/07 the land used for cotton cultivation increased with 35%. However most of the cotton growth production in this period is due to improved yields. The output per hectare quadrupled from 0,2 tons per hectare to 0,8 tons per hectare (UNCTAD, 2010).

China has caused the largest share of growth in cotton production. This is in large the consequence of production stimulating policies implemented by the government. They made use of price incentives and above-quota premiums in cotton production to stimulate production. This led to an increase in production from 4,6 million tonnes in 1983/84 to 6,3 million tonnes in 1984/85 (Ibid.).

The financial crisis of the recent years has also had its impact in the cotton market. The structure of the market is changing which brought along particular financial stress in 2009. Extreme volatility in cotton futures in 2008 caused financial insecurity. Banks decreased lending and the demand for cotton declined. Consequently many merchants experienced extensive losses which bankruptcy or merger as a consequence for some. The crisis is also having its impact on the volumes traded, mostly visible with the large companies (ICAC, 2009).

Fig. 14 World cotton production (million tonnes), by main countries, 1980/81 - 2012/13



Source: UNCTAD, 2010

Exports

In the period of 2002-2006 the United States was the largest cotton exporter with almost 3 million tons of cotton and 3,7 billion dollars exported, around 40% of the total world exports in that period. Uzbekistan became the second largest cotton exporter after the collapse of the former Soviet Union, with a share of 10% in the period 2002-2006. Western countries together do also account for a large share of cotton exports. In the period 2004/05- 2007/08 they accounted for 16% of world exports (UNCTAD, 2010).

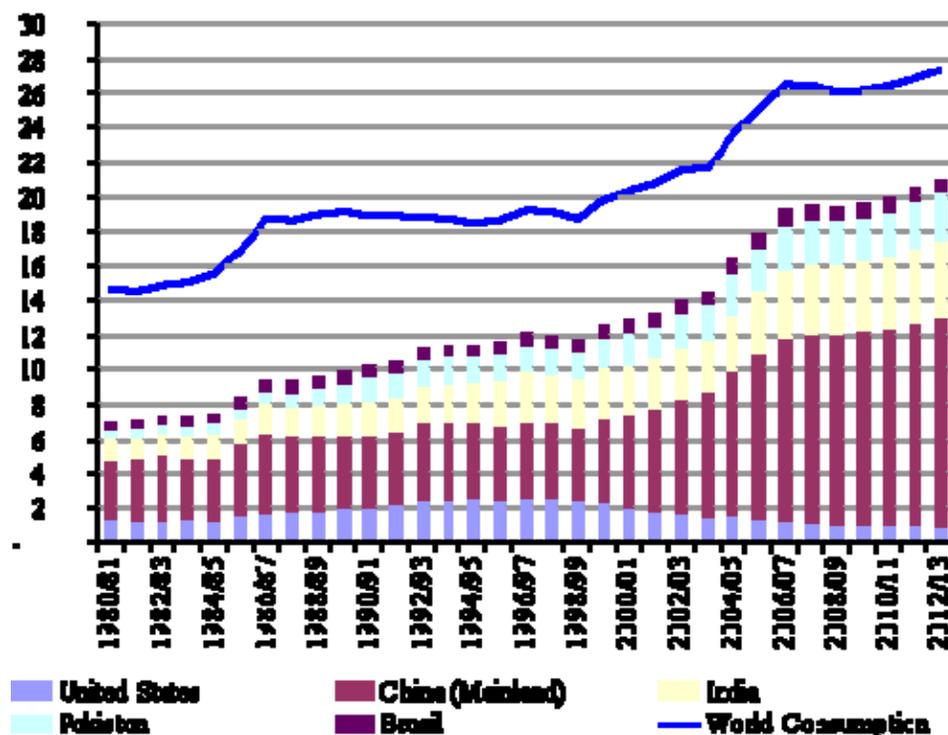
Consumption

Between 1980 and 2006 the number of cotton importing countries almost doubled from 85 to 150. The share of traditional cotton importers has decreased as we note for the EU, the former USSR countries and

East Asia. Contrary the share of China has risen from 52 000 tons in 2000/01 to 2,5 million tons in 2007/08. In the crop year of 2012/13 it is expected to import 46% of world imports. The share that developing countries have in imports is increasing as well. Between 1981 and 1999 they accounted for approximately 78% of global cotton consumption. The term consumption is in this case used instead of import because it also concerns processing of cotton produced within the country. Since 2000 this increased above 80% and in 2010 they are expected to consume 94% of global cotton production (UNCTAD, 2010).

As a cause for this shift increasing wages in the developed countries are mentioned. The textile sector is very labour intensive. About 1/6 of the production costs are from labour expenses, which makes such a shift to low-wage countries very viable. However, after some year of specialisation of the developed countries they were able to regain some of their comparative advantage through differences in quality, new technologies and infrastructure (Ibid).

Fig. 15 Cotton consumption (million tonnes), by main countries, 1980/81-2012/13



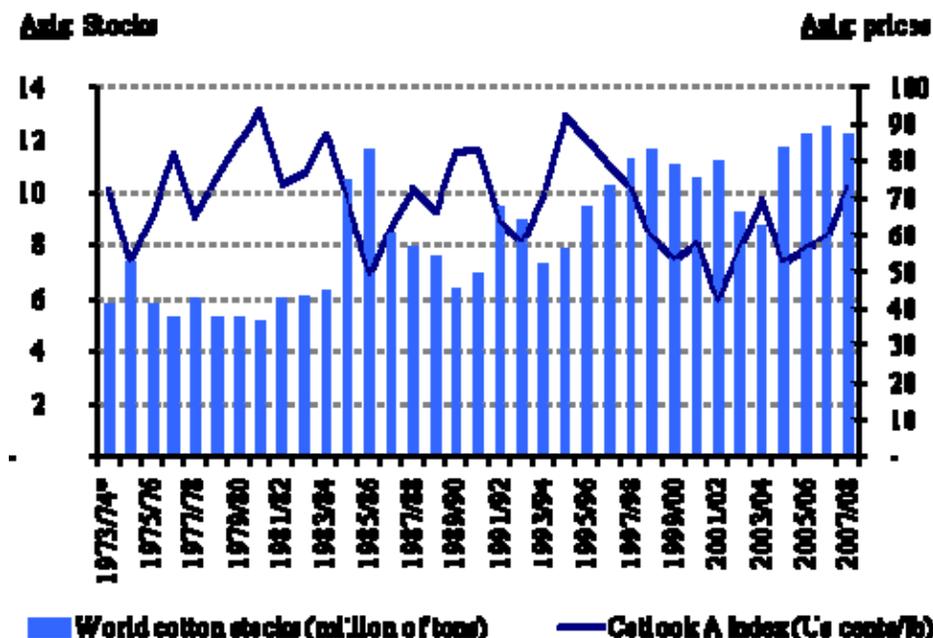
Source: UNCTAD, 2010

Prices

Cotton prices fluctuate depending on world markets, quality of the product and the variety grown. Additionally, government pricing policies influence international cotton prices. In the United States and Europe the prices paid to domestic cotton farmers were 90% and 154% above world cotton market prices in the season of 2001/02. But also in China direct support to producers is provided which is largely affecting the global cotton prices because of China's large share in global production (25%). These price interventions negatively influence the efficiency of the global cotton market. They artificially inflate production and consequently take down world market prices. Especially, developing countries which economically depend on their cotton exports suffer the negative consequences of these pricing policies. Based on data of ICAC estimations the aggregate level of subsidies provided by governments was US\$2,7 billion in 2007/08, US\$5,6 billion in 2006/07 and US\$ 7,7 billion in 2005/06 (UNCTAD, 2010).

Prices are generally based on the cash price paid for cotton in short-term contracts (2-4 months) for forward delivery or in transactions. Cotlook Limited, a UK consultancy, monitors the world prices and transfers these to their price indexes; Cotlook A and B. Cotlook A is the average of the cheapest five quotations of the main upland cottons traded and Cotlook B is an average of the cheapest three quotations for "Course Count" cotton which is transported to Europe. The latter is cotton used for spinning coarse count yarn (Ibid).

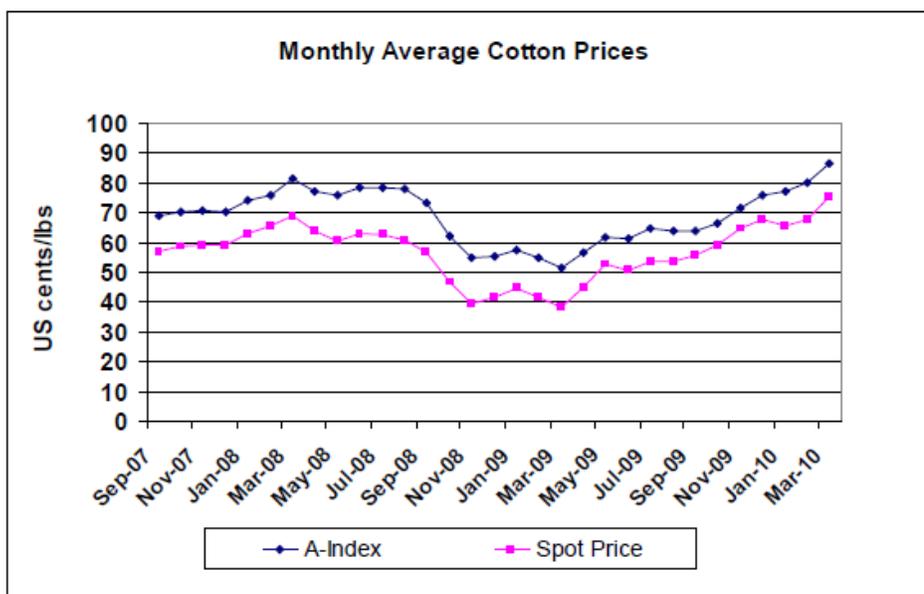
Fig. 16 Long-term price developments for cotton (Cotlook A-Index, 1973/74-2007/08) and world cotton stocks



Source: UNCTAD, 2010

A steady increase of cotton stocks in the middle of the 1980s is visible which continued in the late 1990s and early 2000s. The rise in stocks is caused by the excess supply as a consequence of the subsidies in China and the US. The effect of these policies is visible in the downward slope of the Cotlook A Index during this period. In 2007/08 prices recovered to a relatively high level (UNCTAD, 2010).

Fig. 17 Monthly Average Cotton Prices (Sep '07-Mar '10)



Source: Meyer et al., 2008.

A-Index: Cotlook A-index

Although the global recession has affected cotton demand since the beginning of 2009 prices started to increase again. Demand has also begun to recover but the production is still lagging behind. Not only low prices are the cause but also low yields due to bad climatic conditions in several producing countries. The

prognostics for 2010/11 show an increase of 10% in world production as a response to rising prices. Also cotton demand is likely to grow moderately making global stocks recovery difficult to achieve (Meyer et al., 2008).

Organic Cotton

Cotton production takes place in 22 countries covering the same regions as conventional cotton based on 2008 statistics. In the same year 145,872 metric tonnes of fibre were produced which was 0,55% of global cotton production. This was 152% higher than in 2006/07. This increase can be partly attributed to better data gathering. However, 60% can be attributed to increased production levels within existing projects. Around 217.000 farmers are involved in organic cotton production of which 18% are women (Ferrigno and Lizzaraga, 2008).

In the Middle East and East- Asia the largest amounts of organic cotton are produced. In the season 2007/08 India even produced 50% of the global organic cotton product. In Latin America only 1,09 percent of global organic cotton production is produced. Within Latin America organic cotton is produced in four countries; Nicaragua, Brazil, Paraguay and Peru. Peru is producing the largest share of organic cotton in Latin America. They produce 67% of all the hectares produced. Paraguay comes second with 19% (Ferrigno et al., 2008).

In Peru the production is almost exclusively for export while in Brazil and Paraguay a large share is also produced for domestic and regional markets. There is a great potential for expansion of organic cotton production in Latin America especially in Brazil due to its large farm lands and the existence of a domestic textiles sector (Ferrigno et al., 2008).

Cotton developments in Paraguay

Historically cotton has been an important cash crop for the economy of Paraguay. In the production year 1990/91 they cultivated 560.000 hectares of cotton. However, in the year 2008/09 this has come back to only 52.000 hectares, an enormous decline. This decline is due to several causes; low productivity, low seed quality, degraded soils, low accessibility of credits, lack of interest of producers, bad technical assistance and above all low international cotton prices (Dietze, 2009).

Table 6. Surface, production volume and yields

Cotton	Fibers		
	In thousands hectares	In thousands tonnes	Yield in kg/ha
2000/1	308	98	318
2001/2	170	48	282
2002/3	240	60	249
2003/4	295	110	373
2004/5	250	66	262
2005/6	200	57	283
2006/7	130	35	268
2007/8	65	20	312
2008/9	52	12	221

Source: Dietze 2009: 3

Between 2006/07 and 2007/08 there was a decline of 40,9 % in hectares used. In the district Misiones more specifically the decline between 2005/06 and 2006/07 was from 7000 to 3100 hectare of land cultivated (IICA, 2010). The cause for this huge decline in land cultivated with cotton was a subsidy policy from the government that was misunderstood by the farmers. Before 2005/06 the government would always subsidize half of the seeds. In the year 2005/06 they started to subsidize all the seed expenses. However, the producers misunderstood this subsidy and thought the pesticides were also subsidized. After the harvest none of the producers was able to pay of their loans for pesticides because they did not account on it. The same was true in the subsequent year, 2006/07. In this year the higher links in the chain started to claim their money with the lower links in the chain. However, these links did not have any money and went

bankrupt. This distorted the whole cotton chain in Paraguay and shook up production levels. Besides, it negatively influenced the competitive position of Paraguay on the international market. In the years of 100% subsidy many producers said they would cultivate cotton because of the advantages it offered but in fact they did not. Since cotton is normally sold on forehand the Paraguayan exporters were unable to deliver the amount of cotton they had already sold. Because of that they lost their reliability and since then can only sell the cotton that is currently in stock.⁹ This development is displayed in its degrading share in export value. Between 2006 and 2008 the share of cotton production in total exports declined from 1,8% to 0,4% (IICA 2009).

The production yields in Paraguay have been relatively low over the years due to lack of soil recovery, misuse of investments, problems with the density of sowing, seed quality and more issues alike (Dietze, 2009). Also droughts and other climatic issues influenced the production yield. Between 2006/07 and 2007/08 the yields however improved from 955 kg per hectare to 988 kilos per hectare (IICA, 2010). The highest production yields were obtained in 1992/93. In the same production year a new seed was introduced (Delta Pine) and moreover the climatic situation was favourable (Dietze, 2009).

Table 7. Price relation of cotton between de international market and domestic market

Cotton	Liverpool ¹⁰ cent per pound	Farm Gs ¹¹ /Kg
2000/1	49,93	750
2001/2	37,13	875
2002/3	53,81	2100
2003/4	64,52	2150
2004/5	49,88	1620
2005/6	53,14	1940
2006/7	52,58	1555
2007/8	71,39	2350
2008/9	45,71	1050

Source: Dietze 2009: 16

Cotton prices in Paraguay depend on the international market. Between 2005/06 and 2006/07 the prices most likely decreased because of the international excess supply and pricing policies of the US and China and the national problems due to the failed seeds subsidy. Between 2006/07 and 2007/08 the global prices recovered somewhat which is also visible in the prices for cotton in Paraguay. However, the impact of the financial crisis on the prices has been large. The price decreased between 2007/08 and 2008/09 with 55,3%. This decrease is likely to have its negative consequences in the amount of hectares produced this year.

The negative curb in cotton production does not only influence the directly involved producers it has many external effects as well. Before the downturn cotton productions influenced higher investments and consumption levels and also labour demand, due to its labour intensity. The decrease in production has caused migration to the cities of many rural inhabitants with the hope to find alternative sources of income (Dietze, 2009).

7. 2 Indicator framework for the cotton chain

In this research the indicator framework was used to measure the effects of organic cotton certification. Therefore, the indicator framework as presented above is tailored to the specific cotton context based on often occurring environmental, social and economic impacts within the cotton supply chain. First some more background on these cotton supply chain specific impacts.

⁹ Interview with Carlos Céspedes Zarza, Aratex, 18-10-09.

¹⁰ Indicator for the international cotton price on the market.

¹¹ 1000 Paraguayan Guaraní is 0,22 US\$.

7.2.1. Environmental impacts

Cotton is a crop produced all over the world; the United States of America, Australia, China, Africa, India, Central Asia and South America are all cotton producing regions. It is being produced on large fields in highly technological settings as well as on small fields with almost no technological input. Nonetheless, it is considered to be a difficult crop to grow. The cotton crop is sensitive to drought, low temperatures and various insects. It can be grown on different types of soils although medium and heavy textured, deep soils with good water holding capacity are favoured. Often cotton is rotated with other crops to ensure maintenance of soil fertility and manage fungi and nematodes. However, cotton crops bring in more money than other crops which make reduction or removal of the rotation crops tempting. In response more pesticides are needed to control insects and soilborne plant pathogens and to maintain soil fertility (Kooistra and Termorshuizen, 2006).

During the years pest populations have become resistant to pesticides and pest predators are being eliminated. Therefore, ever higher doses of pesticides are required to control the pest populations, referred to as the 'pesticide treadmill' (Eyhorn et al., 2005; Ferrigno, Lizarraga, 2008; Banuri, 1998).

Cotton production consumes 11% of the world's pesticides while it is only grown on 2,4% of the arable land. Moreover, rough estimates state that in developing countries 50% of all pesticides being used are applied to cotton cultivation (Kooistra and Termorshuizen, 2006). Forty-six different insecticides and acaricides are used which comprise 90% of all pesticides used on cotton. More than half of these insecticides and acaricides are moderately hazardous or worse (Soth, 2006).

Pesticides can be applied in different manners. Aerial spraying accounts for 13% of all sprayings, tractor spraying for 35%, hand-pumps for another 35%, and power sprayers for 17% of all sprayings. In subsistence farming pesticides are often not used because it is unaffordable to the producers. In more developed countries pesticide use is often regulated by governments, however illegal use of pesticides still occurs. The pesticides sprayed do not only attack target populations but also affect non-target organisms. These non-target organisms can sometimes be beneficial to the cotton cultivation which makes the attacks on them an even less wanted consequence (PAN Germany, 2007; Kooistra and Termorshuizen, 2006; Eyhorn et al., 2005; Banuri, 1998). Consequently biodiversity is negatively affected.

These effects are aggravated through run-off, aerial drift, evaporation, leakage into groundwater, and seeping of the pesticides. Some of the compounds end up in rivers which is extremely dangerous to fish populations (Kooistra and Termorshuizen, 2006). But also domestic animals, wildlife and biological diversity are negatively affected. Besides, soil quality and fertility are severely affected (Ferrigno and Lizarraga, 2008; Banuri, 1998). Furthermore, the use of pesticides also has large negative impacts on human health, as described later on in the text. Finally, defoliation, the removing of leaves before the plant is automatically stripped has the same negative impacts as pesticides on the natural and human environment (Kooistra and Termorshuizen, 2006).

Besides the use of pesticides the use of (synthetic) fertilizers also negatively affects the environment. The use of fertilizer causes nitrogen and phosphorus leaking into groundwater. Phosphorus furthermore binds to sediment and contributes to eutrophication if soil is eroded. The leaking of the fertilizers into groundwater can have devastating effects on aquatic life such as algae blooming, depletion of oxygen and a decline in aquatic plants and animals. It is important to note here that the use of organic fertilizers such as animal manure and compost also accounts for nitrogen loss. However, the nitrogen loss caused can be balanced against the positive effects of carbon additions to the soil which increases the level of organic matter in soil (Ibid.).

In addition to fertilizers negative impact on aquatic life, the production of nitrate and ammonium used in fertilizers is using large amounts of energy which contribute to global warming. However, the impact of a single farm is considered to be negligible. Nonetheless, it remains an aspect to keep in mind since the impact of total cotton production is not negligible (Ibid.).

Finally, the use of synthetic fertilizers is an important determinant in the occurrence of acidification. Effects of acidification are found in affected soil communities in vegetation damage and shifts in aquatic communities. The effects depend on the sensitivity of the soil and water and also on the magnitude of deposition (Ibid.).

Another huge problem in cotton production is water use. In many, especially large scale, cotton farms irrigation techniques are used. On a global level these irrigation systems are very unsustainable. The average efficiency is calculated to be 40% which means that 60% of the water used does not reach the plant. It must be noted that many organic cotton farms have rain fed production techniques which can be considered sustainable. However, when groundwater is used it is only sustainable when no more water is used as is the recovery rate. The realisation of irrigation schemes often needs the construction of dams and rivers are being diverted. This causes in almost each case water shortages downstream (Ibid.).

The use of large amounts of water not only negatively influences water stocks but it also negatively impacts land quality. Salinisation and erosion are not uncommon effects of long term use of irrigation practices. Poor water management can cause some areas to receive too much water, with saturation and salinisation as a consequence. Salinisation causes loss of agricultural land due to mineral salts accumulation. Salinisation and water logging together form the first two steps towards desertification

(International Crisis Group, 2005). In the twelve leading cotton producing nations 12% to 36% of the cultivated land is affected by salinisation. Of the total irrigated land one third is or is expected to be affected by salinisation (Kooistra and Termorshuizen, 2006).

The occurrence of erosion depends on the type of irrigation system being used. The sensitivity to erosion is for the biggest part dependent on the type of cultivated land. The angle of the field, the rainfall pattern, the soil type and the percentage of land covered with crop all influence the sensitivity to erosion (Ibid.).

While the soil quality is affected in different manners the use of land is also negatively affected by cotton cultivation. Hundred million acres of land are abandoned as unproductive land due to former cultivation. Cotton production has been especially responsible for that. Besides land clearance might occur in order to accommodate for cotton cultivation expansion. Although cotton cultivation is often not the first reason for land clearance it does occur. The clearance contributes to biodiversity loss, soil erosion and the greenhouse effect. Therefore, it can be interesting to pay attention to the origin of the land being used (Ibid.).

7.2.2. Social and economic impacts:

When analyzing the different social and economic impacts encountered in the cotton cultivation process it is important to note that impacts differ per region, nation and local area. Local characteristics such as cultural differences scale of production and local legislation might influence the impact of cotton cultivation on the social and economic situation of producers.

Since the measurement system is meant to be used in different contexts a broad representation of the different impacts will be presented in the system. In the following these different impacts will be described. It is important to keep in mind that the different impacts do not occur in all contexts.

Obviously the production of cotton brings along some positive economic impacts. People working in the sector are able to access a form of income which might deprive them from poverty. However, not in every instance working in cotton cultivation brings along positive economic impacts.

In cotton cultivation pesticides are used in large quantities. This quantities are rising in a fast pace due to development of pesticide resistance, referred to as the pesticide treadmill (Glin et al., 2006; Banuri, 1998). Consequently the costs for pesticide inputs have been increasing in a fast pace. According to PAN UK the pesticides costs per hectare for farmers has risen with 80% on average. This causes the pesticide costs to account for 60% of the production costs (Glin et al., 2006). The pesticides are expensive to the farmers and often need to be bought on credit, not knowing whether the yields will be sufficient to repay the loans (Ferrigno et al., 2005). This can lead to debt and despair, in particular for small scale cotton farmers (Glin et al., 2006).

In addition, it can negatively influence the food security of a family. Although it does not necessarily influence their food production, lack of credit will decrease the amount of supplementary food bought (Ferrigno et al., 2005). The indebted situation can bring along large portions of stress to the farmers. In India the media even describes cotton producing regions where producers live in very distressful conditions as suicide hot-spots (ICAC, 2008).

The large quantities of pesticides that are being used not only negatively influence the economic position of farmers but also negatively influence their health. Indirectly this once more negatively influences their financial situation since they spend part of their income on medical treatment costs. In Ghana these costs were around US\$ 43 a year. Furthermore, they might also loose working days due to illness caused by pesticide poisonings (Ferrigno et al., 2005).

In reality the medical costs spend due to pesticide poisonings are only a small part of the problem. There are no strict implementation- and safety-measures which causes a large occurrence of poisonings. The poisonings caused are often severe and even lead to fatalities. The World Health Organisation estimates that each year 40,000 lives are lost due to pesticide application (Kooistra and Termorshuizen, 2006). Sprayings can be completed using small airplanes, machinery or with manual spraying tools. The manual application of pesticides is an intensive job hence sprayings are heavily inhaled. The workers spray in front of them forcing them to walk through the just sprayed area which increases the risks of poisoning. Often workers doing the job have protein poor diets which makes them more vulnerable to pesticide poisonings. The effects of pesticide poisonings are dizziness, headaches, vomiting, diarrhoea, anxiety, neurological and vision disorders, miscarriages, stomach cramps heavy sweating and even fatalities (Glin et al., 2006; Banuri, 1998).

The health effects not only occur with the workers themselves. All family members suffer from health problems in areas with regular or high pesticide usages. Children might be playing near the fields, women could be weeding while spraying is taking place and get in contact with the pesticides while washing working clothes (Glin et al., 2006). Storage and disposal problems also cause contact with pesticides from family members not necessarily working on the land. Pesticides are valuable and therefore often stored in the house when other safe storing options are absent. In home storage makes access for children easy.

Furthermore, containers are often used for other (food) products when emptied. Next to the impacts of not being cleaned properly it also causes confusion concerning the type of product that is stored in the container.

Finally, misuse of the pesticides causes additional poisonings. Moreover, pesticides are used on other (food) crops in some instances since the specific use of the pesticides is often misunderstood. Besides, they are also used for domicile problems such as bedbugs and headlice, which can cause tragic cases of poisoning. Illiteracy aggravates the problems (Glin et al., 2006).

Other negative social impacts occur in issues concerning labour rights and working conditions. The ICAC reviewed 168 articles on the social impacts of cotton production. Their main focus was on labour, employment and working conditions. One of the predominant themes mentioned in the articles analyzed was child labour, addressed in 44% of the articles reviewed. Many countries have laws on child labour and ratify the ILO conventions on child labour, but they lack the vital resources to adequately enforce these laws. Besides, they might also lack motivation when they do not perceive child labour as an important issue (ICAC, 2008).

Central Asia is one of the areas where child labour is a largely prevailing problem. Children are often forced to work on the land, often without receiving any payments. When they refuse to work then can be expelled and held back or suspended from school. They are brought to the cotton farms for several months where they live in very poor conditions without any sanitary provisions, lack of clean water and inadequate food (International Crisis Group, 2005).

Child labour is not only a problem in that it violates the children's rights, but also in that it is a perpetuating factor of rural poverty and underdevelopment. Children who are forced to work and consequently do not receive adequate education are bound to grow into illiterate adults with not much perspectives on prosperity.

Finally, children are more vulnerable to pesticide exposure because they absorb the pesticides more rapidly. Besides, detoxifying is harder for them and they are at greater risk to develop chronic and long-term developmental adverse effects (ICAC, 2008).

Women are also a disadvantaged group in cotton production. Just like children women are more vulnerable to pesticide exposure. Often they are the ones mixing chemicals and refilling spray tanks. In the poorest regions of the world women play a vital role in smallholder cultivation. However, their important role is seldom acknowledged and fully rewarded. They face a lack of decision making power, lack land entitlements, lack access to credit, lack representation and participation in collective organisation and a lack representation in national labour rights. Women often work long hours, receive little income and almost no respect from men. Moreover, they have almost no influence on how the money that is earned is spent (ICAC, 2008; International Crisis Group, 2005).

More general issues concerning labour rights and conditions is the prevalence of forced work. As mentioned before child labour is still occurring. But not only children are victims of forced work. In Central Asia also students, teachers, businessman, medical personnel and employees of local administration are forced to work in the cotton industry (International Crisis Group, 2005). Among those that work formal work arrangements are not always existent. Work arrangements are not in all cases recognized as formal employment relationships. Often the relation contains dependency and subordination. Therefore, many waged agricultural workers lack rights of social protection. This is especially true for non-permanent workers, since they often have more informal work relationships. Permanent workers on the other hand more often have contracts, certain social security and employment benefits and are often protected by law. Finally, freedom of association is an important issue. The problem concerning this issue is that many small scale producers lack the capacity and resources to form these organisations (ICAC, 2008).

7.3. Final indicator framework for the cotton chain

Based on the analysis made on the environmental, social and economic impacts of cotton cultivation the list of aspects based on the different sustainability frameworks was adjusted to the cotton case. It shows that some of the environmental aspects are not influenced through cotton production. These aspects are stratospheric ozone depletion, waste heat, ionising radiation and photo oxidant formation and will consequently be left out of the measurement system. Desiccation is an aspect not mentioned as a direct effect of cotton production. However, it can be an impact of salinisation which is one of the often occurring negative environmental impacts in cotton production. Therefore, the aspect desiccation will be replaced by salinisation.

Concerning the social and economic aspects it shows that almost all the important aspect are already mentioned in the framework. Health-, pesticides- and fertilizer-expenditures do not have to form new aspects in the framework. Rather, they can be used for determining earnings of farmers and labourers. It shows the importance of making a distinction between gross earnings and net earnings after all production and health expenditures are subtracted. The aspect storage and disposal of pesticides is also

not mentioned as an aspect and neither should be included as an aspect on its own. However, it can contribute to determining the expected human toxicity levels. Once more the importance of having understanding and awareness of all the inputs and outputs in the production process is shown to be important.

Fig. 18 Environmental aspects for the cotton chain

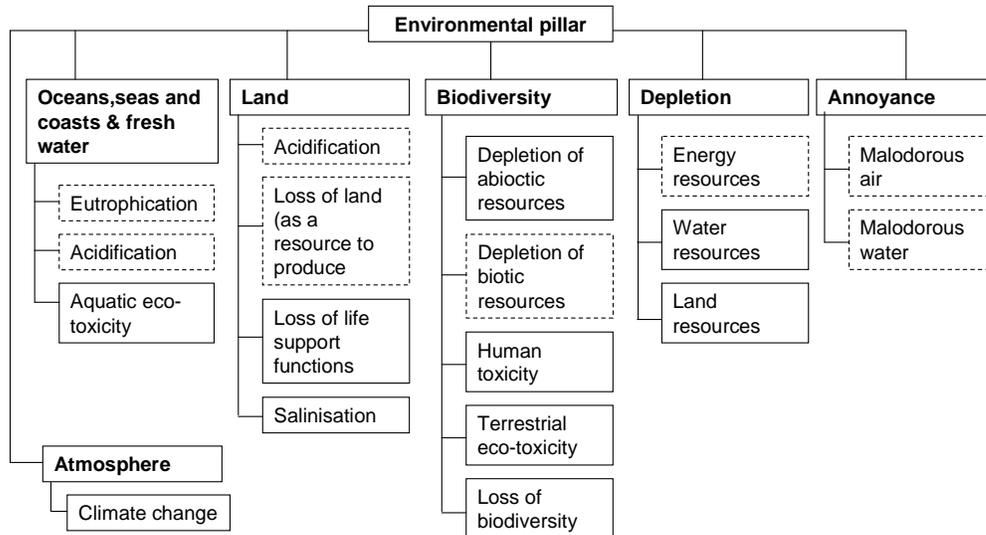


Fig. 19 Social aspects for the cotton chain

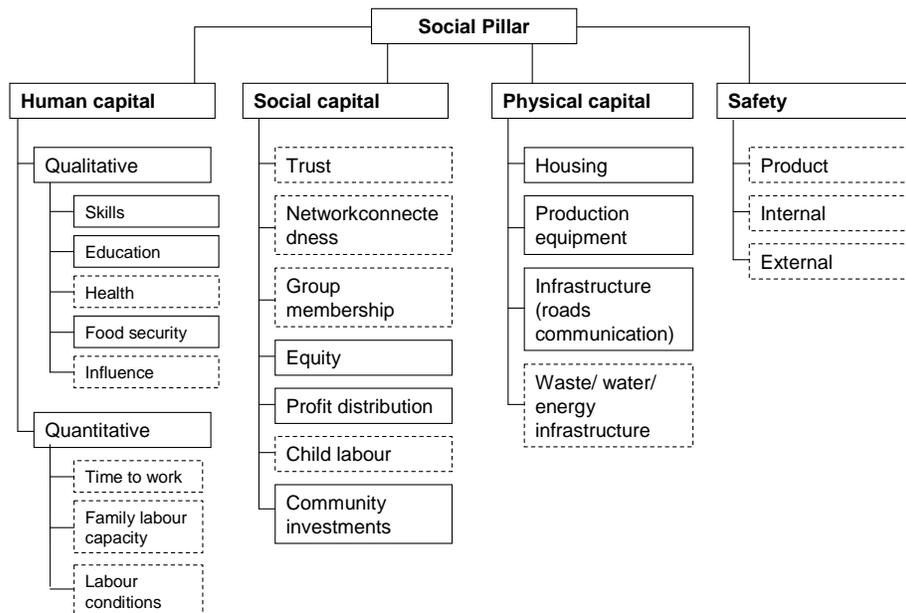
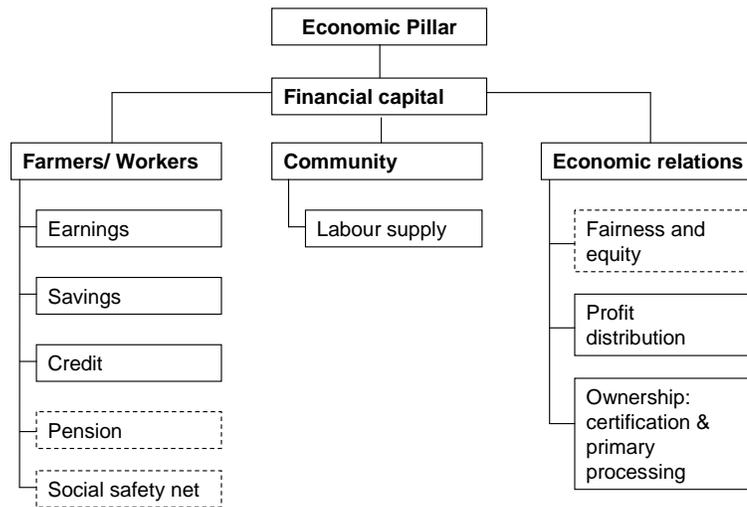


Fig. 20 Economic aspects for the cotton chain



7.4 The aspects used in the test-case

Not all aspects as presented above can be measured in the test-case. In general it is far too time consuming to measure all the aspects that can be influenced by SSCG-systems. Therefore, the indicator frameworks were dispersed to several experts¹² on organic cotton production and certification in Paraguay as well as on a global level. Based on what they consider the most important social, economic and environmental aspects to measure the final selection of the aspects measured in the test-case were chosen. The aspects that are not used in Paraguay are those with the dotted text boxes.

A short operational definition of the all the aspects included in the framework is presented in annex 6. The definitions as presented might not represent all the dimensions that can be attributed to the aspects but those that seemed most appropriate and feasible for this measurement system. Therefore, the definitions as presented could be debated. However, the most important for the reliability of the results is to provide transparency on what is claimed to be measured, and thus the operational definitions. How the operational definitions are translated into questions is explained in more detail in the results chapter.

A final remark concerning the operational definitions is that I lacked the necessary knowledge to make good operational definitions of the environmental aspects. Translating these aspects in measurable units is difficult for someone with no background in this subject. Besides, there was not enough time available in this research process to discuss this issue in more detail with experts in the ecological field. However, I do recommend making use of and expert on ecology to operationalize the environmental aspects when the measurement system is used in the future. The descriptions provided in the annex are mainly focused at which features affect the aspect and should thus be measured.

¹²

Victor Benítez Insfrán, Alter Vida, Executive director
 Hebe Gonzalez, Alter Vida: Director of the Agro-ecological program
 Mario Paredes, Alter Vida: Coordinator of production of the Agro-ecological program
 Simon Ferrigno, Organic Exchange, Director of the farm development program
 Jeroen de Vries, ICCO Kerk in Actie, Verantwoordelijke voor het katoenprogramma in Paraguay.

8. Test results: effects from certification on organic cotton producers

In the previous chapters the line of reasoning behind the creation of the measurement system is presented as well as the system itself. The goal was to create a balanced measurement system in which validity and reliability are addressed without giving in on manageability. Besides, the goal was specifically aimed at measuring 2nd and 3rd order effects of SSCG-systems. Up till now in previous research many difficulties came to the front when trying to measure these orders of effects, especially in development countries. With the strong foundation of the system in theory and practical lessons of existing measurement systems it is hoped to overcome these difficulties and assure that the goal of creating a valid, reliable and manageable measurement system can be realized. However, practice can always prove different. Therefore, the measurement system was tested.

The goal of the test was to study whether these above mentioned goals are actually achieved. Practice can provide great lessons on the strengths and weaknesses of the system as presented and can indicate the areas where improvement is needed. The test was conducted on organic cotton certification in Paraguay. In this chapter the effects from organic certification on the producers are presented. In chapter 9 the methodological lessons from testing the measurement system are put forward.

8.1 Organic cotton certification in Paraguay: Aratex

There is one production company in Paraguay that works in partnerships with producer groups and cooperatives as well as NGOs to produce organic cotton: Aratex. Aratex does not only produce organic cotton but also soy, herbs, seeds and grains. Besides producing they operate ginning, packing, processing, manufacturing and exporting of their products (Ferrigno et al., 2006).

Aratex owns group certifications which means that it owns the certification of the producers and that they have to organize, train and control all the producers. In total about 500 small farmers are contracted which grow organic cotton on 310 hectares. Aratex also has its own farm where it predominantly produces soy (Ibid.).

Aratex is currently certified by the European Union 2092/91, the USDA- National Organic Program for the USA and the JAS for Japan. All the controls are conducted by the Swiss certification body Institute for Market Ecology (IMO) (Ibid.).

Aratex is working together with the national NGO Alter Vida. In some regions the certification and training process is conducted in cooperation with this NGO. One of these regions is the area where the test with the measurement system was conducted. Therefore, the specific certification process of that area is described subsequently.

Becoming certified

Farmers can only apply for a certificate after non-use of herbicides and pesticides for a period of three years. The initial contact is established between the farmers and Alter Vida. First, the will approach and inform producers about the certification and the rules. Alter Vida will visit each farm that want to be part of the certification and controls whether the producers are complying with the rules of organic production. They will fill in a technical compliance form and send it to Aratex's Internal Control System (ICS). When a farmer applies for certification experts of the ICS in cooperation with experts from Alter Vida will conduct technical internal controls. This control is conducted with all the producers. The controls are about checking the environmental state of the farm and compliance to the rules. Based on this control the farm will be accepted in the certification scheme, be neglected or be considered 'in transformation' (that it is in its period of non-pesticide use before it can enter the certification scheme). A farm does not totally have to be pesticide and herbicide free for three years. The producer can also pick one parcel that is poison free for the total period. However, at the moment of entering the certification no pesticides or herbicides at all can be used anymore on the entire farm.

When the producers are accepted by the ICS a list of new producers applying for certification is send to IMO. IMO will control all the documents and will visit around 20% of the producers to control their compliance. Again, this process is assisted by Alter Vida or Aratex. When all the documents are approved the certification is provided to Aratex. The producers that are accepted to the certification will receive notice via Alter Vida. The producers will continue production with help of Alter Vida. De products are harvested and commercialized in cooperation with Aratex.

Trainings

When a farm is accepted in the certification scheme Alter Vida will take care of the further enrolment of the producer and will deliver several trainings. First, trainings will be provided concerning the rules as put down in the organic production certificate. Second, the producer will receive several technical trainings concerning organic production. Third, Alter Vida will monitor the progress at farm-level and will provide follow-up assistance through farm visits.

The technical trainings provided to the producers will always take place at committee level. All the committee members are allowed to attend these trainings. Whether the producer is producing in a conventional or organic manner does not make a difference. In this way the other farmers that are part of the committee but not member of the organic certification scheme also benefit from the trainings provided by Alter Vida.

Auditing and compliance of the farmers

The follow-up visits conducted by Alter Vida immediately also function as a check-up of compliance to the rules as determined in the organic cotton certification. Once a year Alter Vida together with the farmer fills in a technical compliance form in which they describe the environmental situation of the farm. The first time a farmer does not comply with the rules he receives a warning. The second time the farmer will receive a sanction, such as being excluded of sales of his cotton to Aratex for the period of one year. The third time the farmer does cross the rules he will be thrown out of the certification scheme.

IMO conducts a yearly control on a sample of the farms to check whether they continue their rules compliance. A quick overview of these compliance rules is provided in annex 7.

Advantages to the farmers

The producers that are part of the organic cotton certification will in general receive a higher price for their products. However, Aratex does not promote this price difference as an advantage of the certification scheme. The price difference in general is little and Aratex cannot promise that the price difference is the same every year. It will also depend on their financial situation in that harvest year. According to Aratex the main advantage to the farmers lies in the production trainings they receive. Producing in an organic manner will improve their yields and can also assist in improving their health situation.

8.2 Research method for conducting the test in Paraguay

The test in Paraguay was conducted according to the six steps as presented in chapter 5. Some of the steps are taken in a somewhat different manner as proposed in the six steps because some phases of conducting the test were done at the same time as creating the measurement system itself. Moreover, some of the advice on how to conduct the measurement I based on my experience with conducting the test in Paraguay. Consequently I have not exactly followed all these steps myself. The manner of conducting the six steps in the test will be described.

Table 8 Six steps research schedule

Conducting the measurement: Research Steps						
	1	2	3	4	5	6
Source	Databases Local organisations Key persons	Local org. Databases Local governments	Respondents Local organisations	Respondents Key-persons	Respondents Key-persons	Results
Action	Approach contact organisation Determine key-persons Feedback questionnaire	Depict region Determine sample size Draw random sample	Test the questionnaire Adjust the questionnaire	Questionnaire Interviews with key persons	Group meetings Interviews with key-persons Dispersion of results	Entering results in SPSS Analysis Dispersion of results
Location	'Home' Research nation	Research nation/ area	Research area	Research nation/ area	Research area	'Home' Research nation

Research Steps
1. Gathering contextual information and approaching local contacts
2. Sample selection
3. Test & Adjust
4. Conduct
5. Reflect
6. Analysis & Dispersion

The first step was to gather contextual information and approach local contacts. I started with this phase around three months before leaving to Paraguay. My approach was a bit different from how it will normally be since I did not have to focus on a specific product group. In the process of developing my research design I came in contact with ICCO/ Kerk in Actie, a Dutch NGO and Organic Exchange an originally American NGO. While talking to them the plan arose to conduct the test on organic cotton certification in Paraguay. Via them I contacted Alter Vida, a Paraguayan NGO which is working together with Aratex on the development of organic cotton certification. Alter Vida was interested in my research and was willing to help me with conducting the research in Paraguay. Throughout the whole research their assistance proved very helpful. Without their assistance it would have been far more difficult and time-consuming to conduct the research.

While being in Paraguay I discussed the research approach with Alter Vida. They did read over my questionnaire but I did not specifically ask them about their perception of certain subjects mentioned and terms used.

Alter Vida also assisted in choosing the key-persons and approaching some of them. This process was not only in the first few weeks but during the whole research period. Besides, some of the key-persons interviewed came up with new ideas about which persons could be interesting to interview. This evolved as a natural process.

The second step was to depict the research area and select a random sample. The choice for the research area was also made in cooperation with Alter Vida. Alter Vida is assisting Aratex with the certification process in a few villages in Misiones. Therefore, they were able to assist me with my research in this specific area. They offered me assistance in housing, transport and also during the interviews. Misiones did not seem to have any specific particularities which made the area a-typical. Therefore, I decided to conduct the test in Misiones.

Drawing a random sample proved difficult, especially for the conventional producers, since no lists with the total population of the two producers groups existed. For the organic producers I was able to draw up a list with all the producers in the area together with Wilfrido Ortiz, one of the local employees of Alter Vida in Misiones. From this list a random sample could be drawn. The producers were spread over three villages in the area: San Ignacio, Santa Maria and Santa Rosa. For the conventional producers drawing up a list with the total population was impossible. There are many small scale producers in the area and there is no knowledge on which producer is producing which crops, neither at local governments. Therefore, I had to make use of the transactional walk method to select the conventional producers. I randomly passed by

farms in the three villages San Ignacio, Santa Rosa and Santa Maria and asked the farmers whether they were producing cotton and if so whether I could interview them. I tried to select the sample as random as possible by making use of the transactional walk method but sometimes some flexibility was required in order to conduct sufficient interviews.

The sample size I tried to keep proportional to the total population. I interviewed 20 organic producers and 26 conventional producers. Furthermore, I tried to keep the amount of producers of each producer group within each village proportional to the total research population in the village as well¹³.

The third step was to test the questionnaire and adjust it when some aspects were misunderstood or showed any particularities. I tested the questionnaire with two producers. Based on this test I made some adjustments in the questionnaire and filtered out some obscurities. It showed afterwards however that I should have conducted the test more accurate and perhaps with more respondents. Also in the adjusted questionnaire there continued to be some misunderstanding and different interpretations of some of the concepts used. This should be overcome by discussing the questionnaire in more detail with the local contact organisation, in this case Alter Vida, and by conducting the test with more precision. When discussing the questionnaire with the local contact organisation their interpretation of the concepts as used in the questionnaire should be explicitly discussed to assure that the concepts are understood as meant by the researcher.

The fourth step is conducting the research. The final questionnaire can be found in annex 13. The testing and conducting of the research took place in a time span of five weeks. However, only 3,5 to 4 days a week were effectively used. The amount of interviews conducted each day depended in large on the availability of the employees from Alter Vida. I depended on their transport and knowledge of the local area to bring me to the respondents. No detailed maps exist of the area and people do not have addresses as we know them. Therefore, I was unable to visit the producers by myself. Because of my dependency on the local assistance some of the days I could only conduct one or two interviews while other days I was able to conduct up to six interviews. If the researcher does not depend on external assistance, or has this assistance within reach for the whole time he can conduct 5 interviews a day in a fairly relaxed manner. Based on a five days work week the researcher can thus conduct 25 interviews a week. I must note however that this could be a bit less when the respondents live very dispersed.

The assistance of Alter Vida during the interviews made it much easier to obtain the trust of the producers to partake in the research. This was also the case for conventional producers. The fact that a local person guided and assisted me in conducting the research created a lot of trust and made approaching the respondents much easier.

The key-person interviews were conducted throughout my whole stay in Paraguay. Some of the key-persons were proposed by Alter Vida, others were chosen by me to triangulate the information obtained in the questionnaire and others came to the front during the research process. Based on the key-persons interviews and several documents the contextual information is provided in the results for each sustainability aspect. A list of the interviewed key-persons is presented in annex 8.

The key-persons interviews were open interviews. The key-persons were asked about their perception towards the impact of certification on the different sustainability aspects as mentioned in the producer questionnaires. In this way the information from the producer interviews could be triangulated. Besides, they were questioned about their specific area of expertise in order to obtain contextual information about the sustainability aspects.

The fifth step was to reflect on the results in group-meetings and by doing so to disperse the preliminary results to the respondents as well. Some situation specific and language problems made it impossible for me to conduct this fifth step.

On forehand I made the assumption that all producers would speak Spanish but reality proved different. Most of the producers did understand some Spanish but were unable to speak it. Instead they spoke a local language; Guarani, which was impossible for me to understand. Luckily most producers did understand basic Spanish, which was sufficient to understand the questions and the answer categories. Whenever there seemed to be a misunderstanding I called one of the employees of Alter Vida, which were assisting me, to assure that everything was understood correctly. However, in a group meeting a higher level of interaction is needed, which was impossible to achieve with their level of Spanish and my knowledge of Guarani. Not understanding what is being said in a group context and consequently depending on interpreters complicates the process up to such a level that it was almost impossible to conduct the group-meetings. Besides, at the point in time that I wanted to organise these meetings it was harvest-season of a different crop and the producers were unwilling to meet.

The sixth and final step is to analyse the data and disperse the results. Since I had to create the SPSS file and totally think through how to best analyze the results this most likely took more time for me as it will in future research. Around eight interviews can be filed in SPSS a day. When fifty interviews are conducted one and a half week should be sufficient to file them all. The information obtained in the key-

¹³ Santa Rosa: Organic 4; Conventional 9
San Ignacio: Organic 1; Conventional 4
Santa Maria: Organic 14; Conventional 11

person interviews does also have to be filed, which will cost another few days, depending on the amount of key-person interviews conducted.

While filing the results into SPSS it showed that some of the producer interviews had too many exceptions or incongruent answers. Therefore, I decided to leave them out of the research. This was the case with three interviews.

Finally, in the last step the statistical significance of the results was tested. In general statistical significance can only be tested on large samples; $n > 30$. In this research the total n for the individual producer groups was smaller as 30. Therefore, the skewness of the answers on all the variables was studied first. When the skewness is between minus one and plus one a statistical test can be used to determine the significance of the results. If not, the significance cannot be determined.

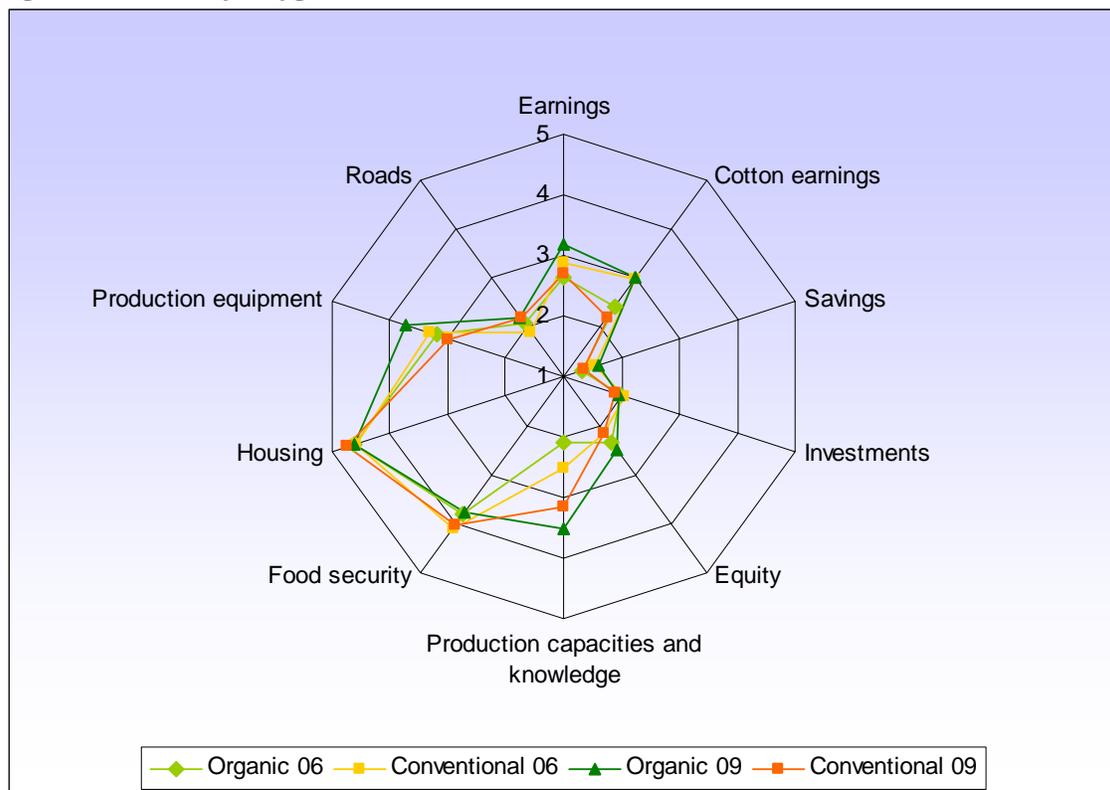
Most results fell within the boundaries of maximum skewness allowed. Consequently a paired T-test was conducted to analyze the means and significance. From the results it showed that almost none of the results were significant. The small n is most likely the cause for this lack of statistical significance.

The results still have to be dispersed to the stakeholders.

8.3 Main conclusions

The biggest impact of organic certification is on 'production capacities and knowledge' and on 'earnings', the latter including income as well as production costs. This impact is also notable in the change in mean score of these two sustainability aspects as presented in the sustainability polygon. 'Production capacities and knowledge' is a special case in which the impact of organic certification is not only felt by organic producers but also by conventional producers. Although the knowledge and capacities increased a lot organic producers still lack some overall knowledge, especially in managing plagues.

Fig. 21 Sustainability Polygon – Social and Economic effects



1 = worst situation 5 = best situation

Table 9 Change in mean scores of social and economic aspects 06-09

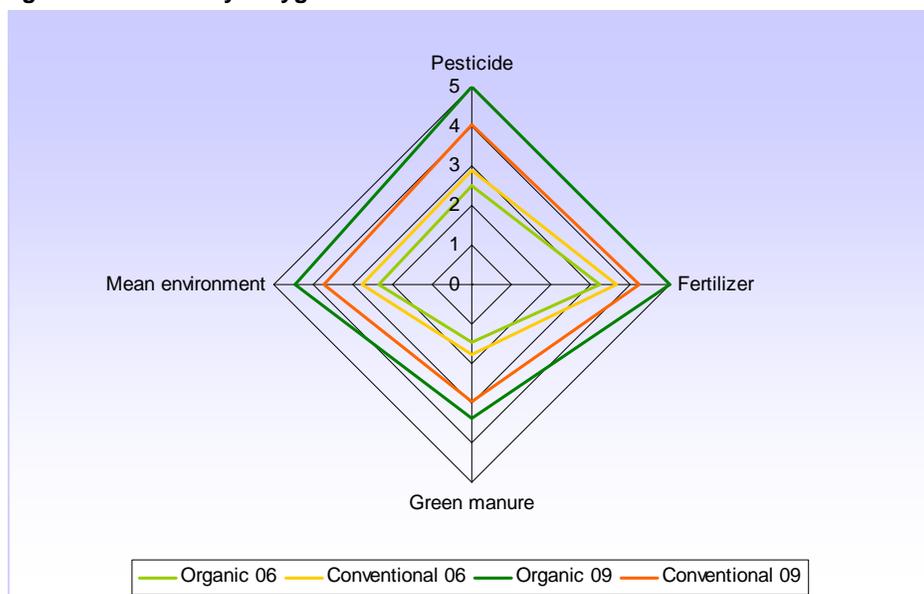
	Organic 06	Organic 09	Conventional 06	Conventional 09
Earnings	2,65	3,18	2,88	2,71
Cotton earnings	2,43	3,03	3	2,21
Savings	1,3	1,6	1,54	1,35
Investments	2,01	1,97	2,06	1,87
Equity	2,36	2,53	2,14	2,15
Capacities	2,1	3,5	2,5	3,15
Food security	3,82	4,08	3,8	4,04
Housing	4,65	4,65	4,62	4,77
Tools	3,2	3,73	3,33	3
Roads	2,1	2,22	1,92	2,19

The impact of certification on food security is also evident although this does not show in all aspects separately. Especially in the sub-aspect 'lack of food' a significant change is visible.

In equity, investments, equipment, savings and labour demand little positive change is being caused by organic certification. The small impact on investments is mostly represented in the amount of investments made within the committee. There is no influence on the investments conducted in the community. The positive impact of organic certification on ownership of equipment is observable in the stable level of ownership of organic producers, contrary to conventional producers who see a decline in their ownership of equipment. In the case of labour demand the positive effect is not so much on the amount of hired labour but rather on the amount of family labour input. The effect on the labour demand in the region is negligible.

Finally, based on these results organic certification does not influence the access to and ownership of credit, the quality and maintenance of roads and the quality of housing.

Fig. 22 Sustainability Polygon- Effects on the environment



1 = worst situation 5 = best situation

The effects of organic certification on the environment are very significant. Organic producers stop using pesticides and synthetic fertilizer and replace this with green manure. This has a positive effect on biodiversity, soil fertility and the health of producers. Moreover, there are positive side-effects visible with conventional producers that also start to produce in an organic manner as a consequence of the trainings provided in the committees.

8.4 Detailed Conclusions

8.4.1 Social and economic aspects

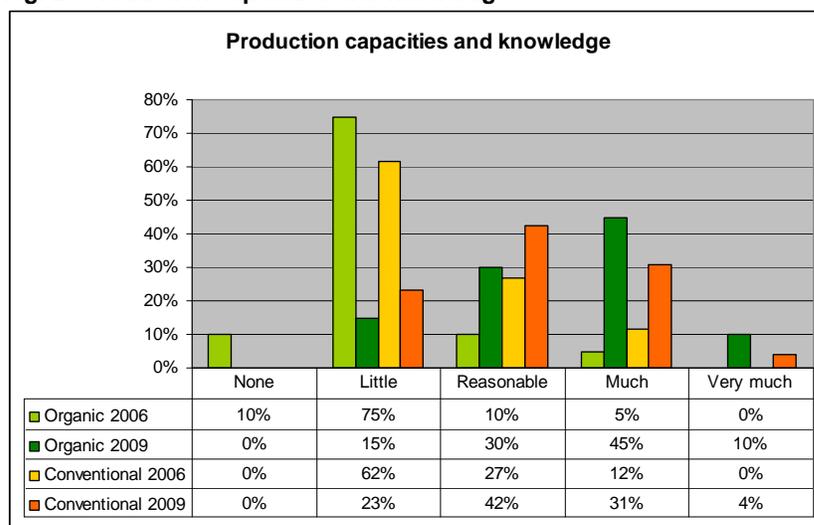
Producing capacities and knowledge

Producing capacities is one of the aspects that are influenced most by organic cotton certification. Aratex has also stressed this to be one of the main advantages for the producers that enter the certification scheme and the results correspond with this opinion.

For measuring the impact on the producing capacities the producers were asked about their level of producing capacities and knowledge in 2006 and 2009. Besides, they were asked whether they learned any additional capacities or knowledge, how many, and who or what the cause was for this change. They were also asked about whether they received trainings and if so from whom. Finally, organic producers were asked whether they miss any knowledge to produce well in an organic manner.

The change in the amount of capacities and knowledge between 2006 and 2009 and is both for organic producers as well as conventional producers evident. The level of knowledge and capacities for organic producers increased from 2,1 to 3,5 while the level of the capacities of conventional producers started at a higher level but they have undergone a smaller increase; from 2,5 to 3,15. Both consider the amount of capacities they have acquired in the last years between reasonable and much. The amount of change the producers perceived they have undergone between 2006 and 2009 was for organic producers 3,2 (reasonable) and for conventional producers 2,47 (between little and reasonable).

Fig. 23 Production capacities and knowledge



N= 46 Organic N=20 Conventional N=26

The amount of organic and conventional producers that consider their own knowledge and capacities little decreases significant and the amount of producers that consider their knowledge and capacities much increases significant as well. The graph also shows that this increase is a bit higher for the organic producers.

Organic producers perceive Alter Vida, their transition to organic production and trainings as the main causes for the change in their level of knowledge and capacities. The conventional producers perceive being part of a committee, their experiences and the assistance of expert, local government and other trainings as the biggest cause for their change.

One of the two most influenced aspects by organic certification is the level of knowledge and capacities of the producer. All the key persons agree that one of the main advantages of entering organic certification is the trainings that the producer receives.

Trainings provided by Alter Vida are not only tailored to cotton production but to production in general. Therefore, the knowledge acquired can be applied to all the products produced on the farm. This provides advantages in many aspects such as higher yields, food security and production costs as explained more later on in this chapter.

The question remains however how the change in the knowledge level of conventional producers can be explained? In part the answer lies in the causes given by the producers themselves as described before. Another part of the answer lies in the fact that the producer committees often consist of organic and conventional producer. The trainings provided to the organic producers are not given on individual level, but at committee level. Therefore, all the conventional producers that are in a committee with an organic producer can reap the benefits of the trainings provided by Alter Vida concerning organic production. This might also explain why many conventional producers name the committee or trainings as a cause for the change in their knowledge level. Thus the trainings do not only benefit organic producers but conventional producers as well. This conclusion will be further supported by some of the results presented later concerning the use of pesticides and synthetic fertilizers.

Organic production capacities

Nonetheless their obtained trainings all organic producers feel that they still lack some capacities to be capable of producing well in an organic manner. Of these producer 50% feels they lack many capacities.

The producers mainly lack capacities in managing plagues in an organic manner. Also a large amount of producers felt they lacked in general some knowledge in all different aspects of producing organically, under which soil preparation received some extra attention.

The key-persons support these results. They also believe that organic producers lack some skills to produce well in an organic manner. Besides, it is sometimes difficult for them to immediately understand the lessons learned and to internalize them. They first need prove before they can fully put their trust in new working methods.

Earnings

Part of the organic certification is a higher price that the producers receive for their product. However, this price difference varies each year and sometimes it is not significant. Therefore, Aratex does not promote it as the main reason why producers should enter the certification scheme. However, the price difference cannot be neglected and moreover the influence on the production costs is large.

For measuring the impacts on the earnings of the producers they were asked about the level of their farm incomes, cotton incomes, total production costs and cotton production costs in 2006 and 2009. This results in different mean scores and graphs of which some are presented below. More graphs on this specific aspect can be found in annex 9.

For the means of change in income between 2006 and 2009 we can see that in almost all results there is an increase in score visible for organic producers. Contrary, for conventional producers there is in all cases a decrease in score visible. The mean earnings and mean cotton earnings are the means in which incomes as well as costs are accounted for. It thus gives a sort of summary of all the other scores displayed in the table.

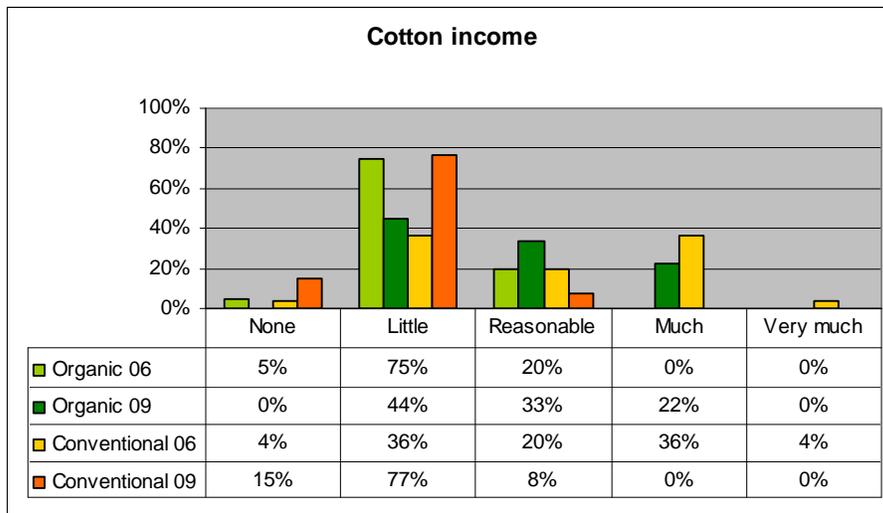
Table 10 Change in earnings

	Orga yes/no	N	Mean
Total income 06	yes	20	2,35
	no	26	2,6538
Total income 09	yes	20	2,95
	no	26	2,4231
Total income cotton 06	yes	20	2,15
	no	25	3
Total income cotton 09	yes	18	2,7778
	no	26	1,9231
Total costs 06	yes	20	2,95
	no	26	3,1154
Total costs 09	yes	20	3,4
	no	26	3
Total costs cotton 06	yes	19	2,7895

	no	25	3
Total costs cotton 09	yes	17	3,3529
	no	24	2,5833
Mean_earnings_2006	yes	20	2,65
	no	26	2,8846
Mean_earnings_2009	yes	20	3,175
	no	26	2,7115
Mean_cottonearnings_2006	yes	20	2,425
	no	25	3
Mean_cottonearnings_2009	yes	18	3,0278
	no	26	2,2115

While these mean results do not show detailed results, from the graphs in which the score on all the Likert scale items are presented some significant differences come to the front.

Fig. 24 Cotton income



N= 46 Organic N=20 Conventional N=26

As we can see the amount of organic producers that believe they have much cotton income increased from 0 to 22 percent, while it declined for conventional producers from 36 to 0 percent. On the other hand the amount of producers that perceive they have little cotton income increased significantly for conventional producers from 36 to 77 percent while it decreased for organic producers from 75 to 44 percent. When analyzing in more detail the change in income in general more or less the same picture is visible but the differences are a bit smaller.

The results regarding production costs show more or less the same trend. The production costs in general are decreasing more for organic producers as conventional producers. When we solely focus on cotton production costs this difference is even bigger. The amount of producers that consider the cotton production costs much amongst conventional producers has increased from 28% to 54%. Contrary this amount has for organic producers decreased from 37% to 12%.

According to the key persons the price paid for organic cotton is a bit higher but not that much. According to them the price difference between conventional and organic cotton is small and it is not considered as one of the big advantages for organic producers. Even Aratex, the organic cotton buying company, does not promote the price difference as one of the advantages of certification. However, there are several changes that occur in the income for organic producers that might explain the results from the questionnaire. First some more information about the price difference between the two types of cotton.

The estimations concerning these prices differ a bit amongst the different key-persons. Based on the gathered knowledge I have made the following price estimation:

Table 11 Price changes - per kg of cotton in Guaranies

	Organic	Conventional	Difference in %
2006/2007	1875	1500	25%
2007/2008	2500	2300	8%
2008/2009	1300	1200	7,70%

According to all the key-persons the price difference decreased over the last three years and some even claim that the last year there was no price difference at all. Although there has been some price difference this has not been substantial, especially in the last two years. The question remains how the results as presented above can be explained?

One of the explanations could be the difference in production yield per hectare. The yield per hectare is much higher for organic producers as for conventional producers. In the harvest season of 2005/2006 the yield for organic producers was 1500 kilos per hectare while for conventional producers this was only 950 kilos per hectare (Paredes and Rojas, 2007). A higher yield and higher cotton prices together can make a substantial positive difference in the level of income.

A different explanation can be found in the difference in production costs for the two types of producers. In accordance with the results presented above the key-persons also state that the production costs of organic producers are much lower as those of conventional producers. Pesticides and chemical foliar are expensive products. Organic manure and pesticides on the contrary can in most of the cases be made with products of the farm, such as plants, egg shell, ash etc. Thus the costs for pest treatment and soil nutrition are very low. Since organic farmers have to make their total farm organic this will not only influence their cotton production costs but also their total production costs. One could reason that the lower production costs, and thus higher net earnings, are perceived by the farmer as higher income, because they do not separate the two. This might be another explanation for why the organic producers perceive their incomes that much higher as conventional producers.

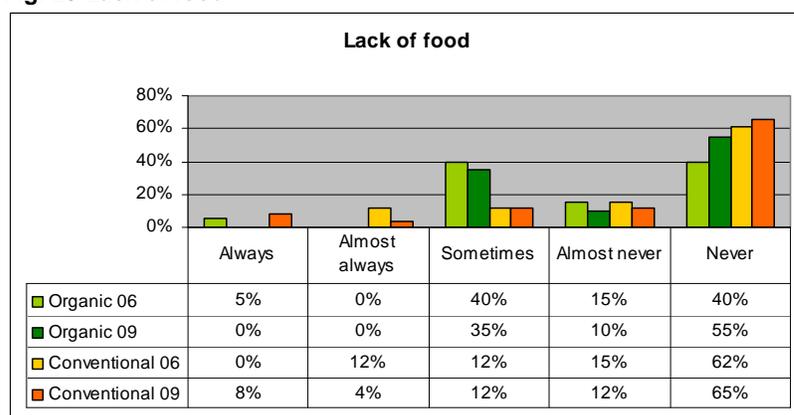
Food security

The direct impact of the certification scheme on food security is maybe not expected to be high. However, there is an indirect impact visible because the producers apply the newly learned production techniques on their food crops as well.

The impact on food security is studied by asking the respondents about the frequency with which they ate all sorts of products, ranging from very basic products to more luxury products in 2006 and 2009. Besides they were asked about their perceived lack of food in these two years. Some questions were asked about their expenditures on food. Finally, questions were asked about whether they felt they had undergone a change in food security and if so, what has caused this change to occur.

If we analyse the means of food security we see that the change between 2006 and 2009 for both organic as well as conventional producers is marginal; from 3,81 to 3,80 for organic producers and from 4,10 to 4,04 for conventional producers. However, in the specific aspect of food security there is a different change visible of the two producer groups.

Fig. 25 Lack of food



N= 46 Organic N=20 Conventional N=26

The mean on food security for organic producers has improved from 3,85 to 4,2 while the situation for conventional producers started at a higher level but slightly decreased towards 2009; from 4,27 to 4,23. Besides, in 2009 none of the organic producers believed they always lacked food while 12% of the conventional producers believed they did. Moreover, the amount of organic producers that believed they never lacked food increased from 40% to 55%. The increase for conventional producers was only 3% but their general score is higher.

Many organic producers believe they have undergone much change in food security. For conventional producers this amount is much lower. However it is also important to note that also ten from the 46 respondents feel they have undergone a negative change.

In the results there is no significant difference in the expenditures on food and luxury products between 2006 and 2009 for both producer groups.

The image as presented by the key-persons corresponds with the results as presented above. According to the key-persons the food security of the organic producers increases because they apply the production techniques learned for organic cotton production on their other crops as well. Consequently their auto consuming capacity increases. Moreover, organic farmers that form part of the organic markets are even better off. They are specialized in food production and have higher incomes because of their sales at the market, which they can use to provide in their own food security.

Labour demand

Labour demand on the farms can be expected to be influenced by organic certification via increased earnings. When the farm becomes economically more viable there is a chance that it will expand its production. As a consequence they will need more labour input and they have more financial means available to be able to afford this. Labour demand is an important aspect in this specific context since there is a visible flow from (young) labourers to neighbouring countries because of the lack of labour demand in the area.

The impact on labour demand is measured through asking the respondents about the number of external labourers working for them in general a month, the total number of days these labourers work on the farm a month and the number of family members and the frequency with which they work on the farm. All these questions were asked for the situation in 2006 and 2009. Finally, they were also asked about the cause of the change in labour demand.

From the total of 20 organic producers 8 producers believe they amount of hired labour on their farm has decreased and one farmer believes it has increased over the past three years. From the 26 conventional producers 12 believe that the amount has decreased and 8 believe it has increased.

The number of labourers hired each month changed for organic producers from 10,75 to 6,3 and for conventional producers from 12,88 to 10,48. The amount of labourers hired by conventional producers was and remains higher and the decline is smaller. This difference can be partly explained by the smaller amount of family members that work on conventional farms. For organic producers this was on average 3,0 in 2006 and 2,3 in 2009. For conventional producers this was 2,58 in 2006 and 1,88 in 2009. Some of the decrease in both labourers and family members working on the farm is accommodated by the higher frequency with which family members work on the farm. The increase is small though. For organic producers the increase was from 4,6 to 4,85, for conventional producers this was even smaller from 4,69 to 4,88.

An explanation for the difference in hired labour between organic and conventional producers can also be sought in the higher amount of cotton cultivation by conventional farmers in both 2006 and 2009. Cotton cultivation is very labour intensive which necessitates a lot of labour input. However, the amount of hectares owned in total by both type of producers presents a bit contradictory results. Organic producers own more land in 2009 as conventional producers (for 2006 no data is known). This would rather direct to higher labour inputs demand by organic producers. Without exact knowledge on how all the hectares on the farm are used it is difficult to make any statements about this.

As a cause for the negative change in hired labour organic farmers mostly mention that they cultivate less products in general and less cotton, thus they need less labour. They also mention that they lack the money. Before they used their credit but currently they do not have that anymore. Finally, one farmer said that he receives more family help at the moment. Conventional producers see the cause mainly in the low cotton price and low price of products in general. They also mention that they stopped or cultivate less cotton or other products. Finally, some of the producers have higher other expenses, such as school fees.

However, there are also some farmers that hire more labour as before although this does not show in the mean tendency. This is the case for 8 conventional farmers and 1 organic farmer. This could explain the higher mean score of conventional farmers. As the main reason for hiring more day labourers they

mention that they produce more products in general. A different cause lies in receiving less family help and consequently depending more on external labour. The one organic producer that started to hire more labour believes that he is producing more because of the training that he received.

The key-persons believed that the influence of organic certification on labour demand is little. Since organic cotton production is labour intensive the demand has increased a little bit. But this is mostly accommodated within the family since in general there is a lack of work. In the last years there is a trend of young people leaving, to in most cases Argentina, because this labour shortage. The labour demand fluctuates a lot which makes the work available very instable. Thus the influence on labour demand within the family is reasonable while outside the family the impact is very little, which is aggravated by the bad economic situation in Paraguay as a whole.

Savings

The savings of the producers are expected to be influenced by the certification scheme via increased earnings. Savings are considered important to the sustainability of the producers because they influence their resilience in cases of economic downturns and their ability to make (big) expenditures when necessary.

The impact of savings is measured through asking the respondents about their amount of savings in 2006 and 2009 (not in numbers but in Likert-scale items).

The savings for organic producers have increased a bit, from 1,3 to 1,6 while they have decreased for conventional producers, from 1,54 to 1,35. Nonetheless the savings of both groups remain very little. Still more as half of the organic and conventional producers had no savings at all in 2009. Of the organic producers in 2009 45% had some savings. For conventional producers only 27% of the producers had some savings in the same year. Based on these results it seems that organic producers have a bit more savings as conventional producers. They might also understand the importance of savings better since the amount of producers with savings has been increasing of the years for organic producers while decreasing for conventional producers.

In my interviews with key persons I did not gather much data about the saving behaviour of producers. It was not a subject that was brought to the table by the interviewees themselves either. However, what I did grasp is that the organic producers might be a bit more aware of the importance of savings. They are very busy with developing themselves and their farm and because of that become more financial aware. Moreover, their incomes increase a bit thus they have more opportunities to start saving as well.

Credit

Credit is expected to be influenced via the certification scheme in a similar manner as savings. Due to being part of the certification the economic viability of a producer could increase which gives them more credibility. Besides, the fact that the producer is a member of a certification scheme in itself could influence their economic credibility.

The impact of the certification scheme on credit is measured through asking the producers whether they had credit, had access to credit, and the difficulty of obtaining credit in 2006 and 2009. There were also asked to describe what the conditions are when someone wants to obtain credit.

It is interesting to see that the percentage of producers that did have or did not have a credit in 2006 was almost the same for organic and conventional producers (70% and 69%). It is also interesting to note that for both groups the amount of producers that possessed a credit has declined (55% and 50%). For the conventional producers this decline is a little bit larger but the difference is not significant.

Although the majority of organic and conventional producers believe that they have access to credit they also perceive it difficult to obtain access to credit. Organic producers perceive it much more difficult to obtain credit as conventional producers. This was the situation in 2006 and even increased in 2009. Why organic producers perceive it so much more difficult is hard to explain. Both producers groups agree about the necessary arrangements to be capable of obtaining credit. Of the producers 78% mentioned that in order to obtain credit one needs a certificate of ownership of any kind. In most cases this concerns ownership of land or an animal. The remaining 22% did not know what was needed or they believed a certificate of salary was needed.

The results as presented above correspond with the information withdrawn from the key-person interviews. According to the key-person having (access to) a credit is not influenced by organic certification. Both producers groups need the same documents in order to obtain credit. As stressed by the producers they

need to show some sort of document of ownership. This can be ownership of land or an animal. If a producer does not have this prove of ownership it becomes very difficult to obtain a credit. Many small scale farmers do not have a proof of ownership which makes it hard for them to obtain credit. The only solution is to put a mortgage on the land. This, however, is very dangerous. In case of not being able to pay back the loan or credit in time the producer will loose its land.

Some more information on options and rules for obtaining credit are presented in annex 10.

Investments

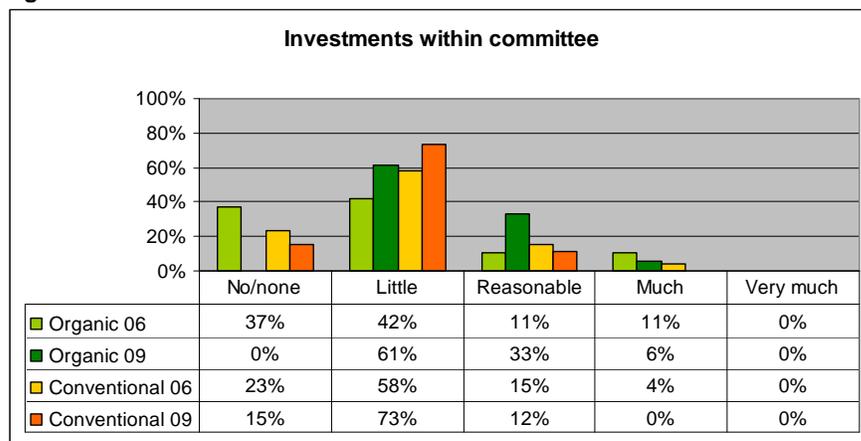
Investments can be influenced in different manners through the certification scheme. Increasing earnings can lead to more investments of the producers in their community. Other actors in the supply chain could also make investments in the local area, as a part of the certification, for example the cotton buying company. These investments focus mostly on the impact of certification on the community and thus tell us something about 3rd order effects of certification.

The impact of certification on investments is measured by asking the respondents about their investments in the community, such as in schools, health care and infrastructure. They were also questioned about the investments within the committee and the investments of the committee in the community. All these questions were asked in relation to the situation in 2006 and 2009. Finally, there was also a question posed about the investments of the cotton buying company in the community.

The mean change in the amount of investments done by both by organic and conventional producers has declined minimally between 2006 and 2009 from 2,01 to 1,97 for organic and from 2,06 to 1,87 for conventional producers. Although the decline for organic producers is a bit smaller as for conventional producers it is difficult to make any claims based on these means. The same change is visible for the amount of money spent on investments by organic and conventional producers; a small decline.

The change in investments conducted by and within committees shows almost the same tendency. From the organic producers 90% is member of a committee or association and 92% of the conventional producers are member. The change between 2006 and 2009 in the amount of investments done outside of the committee by both organic and conventional producers is almost negligible. Both increased slightly. The investments within the committee however have changed for organic producers from 1,95 to 2,44 while it decreased for conventional producers from 2,00 to 1,96. These results are more in detail presented in figure 28.

Fig. 26 Investments within committee



N= 45 Organic N=19 Conventional N=26

Both producers groups believe that there are almost no investments done by the cotton buying company. Only 10 out of 46 respondents believed some investments took were conducted. This small amount of respondents makes it difficult to make any statements about the importance of the investments. Especially since the answers seem to be a bit random. Therefore, no conclusions can be drawn concerning this specific topic.

Most committees are aligned in an association. These associations meet on a monthly basis. Each committee sends some delegates to these meetings which have to bring the monthly fee. The gathered

money is invested in the committees. This can either be financial help, seeds, tools and the costs of running the association. Some associations even were capable of buying a tractor in this manner.

However, in general the amount of investments done within and outside of the committees is very small. People simply lack the money to do investments outside of their family. If they do investments these are in 'kind' rather as in money.

The influence of this small increase of investments buy organic producers does not influence the situation in general in the region according to the mayor of Santa Maria and the other interviewed key persons. The amount of investments and producers is too small to have any impact on the region so far.

Equity¹⁴

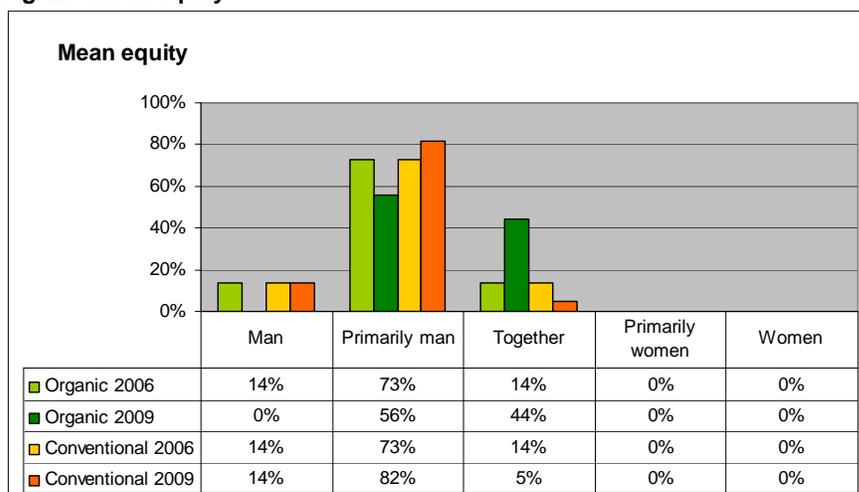
Equity between husband and wife can be influenced by the certification scheme as a result of participation in the production trainings. Women and children are stimulated to participate in the trainings. Because they develop skills they might start helping on the farm which could consequently change their position within the family and might lead to more equity between husband and wife.

Equity is measured through asking the respondents about ownership of the farm, the income generating person, the person who makes decisions concerning finances, household and family issues and farming issues.

From the data it shows that the overall mean on equity has changed slightly for organic producers between 2006 and 2009 from 2,36 to 2,53. The level of equity for conventional producers has almost remained the same. In 2006 they value was 2,14 and in 2009 this was 2,15.

Looking at the graph in which the scores on all the items of the Likert scale are presented clarifies the situation and the change occurred in the past years more.

Fig. 27 Mean equity



N= 40 Organic N=18 Conventional N=22

It shows that while for the organic producers a shift is visible from a household in which the man primarily is making the decisions to a household in which both man and women have a say. Contrary, in the case of the conventional producers the influence of the women seems to decline over the years.

The analysis of the different aspects of equity provides more or less the same picture although a little less straightforward. The influence of women in household decisions over the years has increased a little bit within organic families. In conventional producing families there are fewer instances in which the woman has the most influence but the number of families who are taking decisions together has increased from 50% to 73%. Thus the results on this specific topic are a bit ambiguous.

For the change in influence in farming decisions the results are less ambiguous. The percentage of families in which the decision making power concerning farming solely was with the man decreased for

¹⁴ The topic equity is a bit different from all the other topics. The Likert scale does not range from none to very much but differentiates between man, mostly man, together, mostly woman, woman. Analyzing the means of the questions on this topic thus has to be done in a somewhat different manner. The higher the score, the more influence the woman has in the family. Since women often still have a secondary role in the family, the more influence they have, the 'better'.

organic producers from 56 to 50 per cent while it increased for conventional farmers from 86 to 91 per cent. Besides, both in 2006 and 2009 there were no conventional producer families in which the decisions concerning the farm were primarily or solely taken by the woman.

The interviews with the key-persons support the above presented results. The influence of entering organic certification on the equity between wife and husband is large. One cause is the larger extend of openness towards change within organic producing families. However the main cause for the change seems to be the trainings provided to organic producers by Alter Vida. The amount of women participating in these trainings has increased over the years. They become more knowledgeable and as a result are starting to form more part of the production process. In most cases the decisions in organic producing families are taken together. However, this change is not totally caused by the organic certification.

Alter Vida perceives it very important to stimulate women in participating in production trainings provided by Alter Vida. Women are actively stimulated to participate. Especially in the organic markets that are organised on a weekly basis in Santa Rosa and San Ignacio women participate actively. According to Wilfrido Ortiz the change in the position of women in the family is both being caused through the trainings which are part of the organic certification and by the work of Alter Vida.

Equipment

The amount of equipment that the producers own can be impacted by the certification scheme via increasing earnings or investments by other actors in the supply chain. With increased access to production equipment the production process can become less labour intensive which will positively influence the productivity of the producers. Increased productivity can in its turn lead to increased yields which will positively influence the income of the producer.

The change in the amount of equipment that the producers own is measured through asking them about the amount of equipment they owned in 2006 and 2009, about whether they owned sufficient equipment to cultivate their crops and finally whether their amount of tools changed in the past three years and what the cause was for this change.

In 2009 organic producers in general own a little bit more production equipment as in 2006. In 2009 they believed to own a reasonable amount of equipment. Also conventional producers in general have undergone a positive change in possession of equipment, while the increase was smaller for them. However, still 21% of the conventional producers believe that they have undergone a negative change in the ownership of equipment between 2006 and 2009. This negative change is mainly said to be caused by low product prices, low yields and defective equipment. Oppositely, organic producers only believe they have undergone a positive change.

The cause for the positive change in ownership of equipment of organic producers mainly lies in the training they have received. This has positively influenced the amounts of products produced, their income and consequently the amount of money available to buy equipment. In the case of conventional producers the cause for improvement is a bit more dispersed. The mention causes like more work, better money management, trainings and the committee they are part of.

The key-persons support the results and believe that there is almost no influence from organic certification on the ownership of equipment. The increase in incomes for organic producers is not sufficient to enable producers to expand their ownership of equipment. Although there is some increase in income visible for organic producers their earnings remain relatively low and probably insufficient to make large expenses. The higher income does however seem to be sufficient to replace defective equipment since none of the organic producers has experienced a negative change contrary to conventional producers.

Housing

Housing is also one of the aspects that can be influenced by the certification scheme via increased earnings. Depending on the quality of the housing of the producers this aspect is of great or less importance. In Paraguay, most houses have a reasonable quality although they are humble and in some cases exist of less strong materials such as wood.

This aspect is measured via asking the respondents about the number of household members, the number of rooms and the amount of problems with the house in 2006 and 2009. They were also questioned about the change their house had undergone between 2006 and 2009 and the cause for this change.

The household size has changed for both organic and conventional producers in a similar manner. The change for organic producers is from 5,15 to 4,5 and for conventional producers from 5,27 to 4,46. The

amount of rooms within a house has increased for both producer groups: for organic producers from 3,6 to 4,25 and for conventional producers from 4,19 to 4,42.

In general organic and conventional producers do not have many infrastructural problems with their house. There is almost no change visible between 2006 and 2009. Only one organic and conventional producer believe they have undergone a negative change in housing the past years, while in total 7 organic and 11 conventional producers believe they have undergone a positive change.

The causes brought up for the positive change in their housing are very diverse. They make it hard to make any claims about the influence of organic certification on housing quality. None of the results seems to point to any causal relation in this aspect between organic certification and the housing quality.

Based on the key-person interviews there is not much to add to the results as presented. According to them the influence of organic certification on the housing of the producers is small. The change in financial income is too small to be able to positively contribute to the quality of housing.

Roads

Roads is one of the aspects that can be impacted by the certification scheme via investments of producers or other stakeholders.

This aspect is measured by asking the producers about the availability of the five types of roads in their community that are most common in that area. These types range from very bad vegetal roads to asphalt roads. Besides they were questioned about the amount of maintenance work on the roads, the level of change in the road quality in the past three years and the responsible agent for this change.

The total mean score on the subject of roads has slightly improved for both producers groups. The mean for organic producers increased from 2,00 to 2,22 and for conventional producers from 1,92 to 2,19.

In general all the scores within the subject of roads are almost the same and display almost no change. Both organic and conventional producers believe that the roads have improved a little bit over the last three years. Of the total respondents 90,3% held the local, regional or national government responsible for this change. None of the respondents mentioned anything whatsoever related to increasing incomes or investments of the cotton industry as a cause for the small improvement in roads. Based on this and the very similar results for both producer groups it is fair to state that organic cotton certification does not influence the road availability, quality or maintenance.

Finally it is interesting to note that both producers groups consider the availability, quality and maintenance of roads reasonably important and that their score is almost exactly the same; 3,19 for organic producers and 3,17 for conventional producers.

All the key-persons confirm that organic cotton certification does not influence the availability, quality and maintenance of roads. Road construction and maintenance is part of the public works of the government. So far there is no influence in this from people or institutes outside of the government.

Profit distribution

Equal profit distribution is one of the claimed aims of SSCG-systems. Producers do not only have to comply to several production demands but they also receive a better and fairer price in return.

Measuring the impact of certification on profit distribution is a somewhat special case. The aspect is very difficult to study through questions to the producers. The producers have very little economic awareness and are very likely to have no insight in price setting further up in the supply chain. Moreover, it is very difficult to study this aspect since it is influenced throughout the whole supply chain which necessitates a gathering of data throughout this whole supply chain. Therefore, the goal in this research was to gain some insight in the fairness of profit distribution at the beginning of the supply chain; between the producer, cotton buyer and the international market. The information obtained on this subject provides some insight in the situation but more specific data is needed to be capable of drawing strong conclusions.

Equal profit distribution is represented in the prices paid to the different links in the value chain. In the case of certification schemes the question is whether also the lower links in the value chain, the farmers especially, receive a share of the higher price paid for the certified product. Based on the gathered information only some insight in the profit distribution in the southern side of the value chain can be provided.

Aratex Organica is the company that is buying the organic cotton from the producers in Misiones and is also providing the certification to them. Aratex pays for this certification and also owns the certificate. For the organic cotton that they buy from organic producers they pay a higher price as the market price for conventional cotton. They are unable however to put a predetermined percentage of extra they pay for the

organic cotton because there are too many market influences which makes this impossible. Moreover, the plus is very small. They would want to pay higher prices but there are several reasons which make it difficult to do so.

In the last years the plus paid for organic cotton became less and less. According to Aratex this is the consequence of changing international market prices of organic cotton. Whereas before the plus paid for organic cotton on the international market was 30%, currently this difference is almost none. This hinders the possibility of paying a significant plus to organic producers.

Paying a better price is furthermore hindered by the large costs attached to organic certification. Membership of a certification body brings along high annual costs. In the case of Aratex this is around US\$ 12.000 a year. Besides, they have very high costs in the trainings they need to provide to the organic producers. In Misiones these trainings are arranged by Alter Vida, but in other regions Aratex has to arrange and pay for all the trainings themselves. Especially in the first few years these costs are tremendous. After a few years the producer group becomes consolidated and the knowledge internalized. From this point onwards the costs become less.

Aratex is only possible to realize this certification and pay for all the costs entangled with it because they are producing end-products. Besides producing and selling fibres they also produce cloths which they sell on the European market. Because of doing so they are able to compete on different levels, such as product quality, service, design and logistics. In these levels they can obtain a better margin which enables them to pay for all the certification costs. This story is supported by Manufactura Pilar who only exports fibres. Manufactura Pilar wishes to enter in organic cotton certification but is unable to pay for all the costs.

Finally, Aratex assures that they directly pay the producers, instead of paying the middlemen. According to Aratex most companies pay the middle men, subsequently they pay the producers. How much they actually pay to the producers is not controlled. For Aratex this is different. They pay the producers and middle men separately, to assure that the producers receive a fair price.

Summarizing, it seems that on this side of the value chain there is no inequity in profit distribution. The little that Aratex receives extra for their organic cotton they use for covering several expenditures that come along with certification. Based on their and Manufactura Pilars information it would even be impossible for them to proceed the organic certification without converting the raw cotton into end-products.

Ownership: product processing and certification

In a development world context the shift from producers from production to processing is considered very important for their development. When a producer not only produces but also processes its product it can add more value to it which can consequently influence its economic situation. When producers develop more economic awareness through being part of certification scheme they might also start more value adding processes.

This subject in the questionnaire proved to be very difficult for the farmers to understand. It clearly shows that the farmers lack education and that they find it hard to relate to concepts such as product processing and certification. After a lot of explaining most farmers understood more or less what product processing means but I nonetheless have the feeling that the answers on these questions are not reliable. The same is true for the questions concerning certification. Of the organic producers 75% answered that they are the owners of their certificate. However, my interviews with Alter Vida and Aratex showed that Aratex owns the certificate. The farms are approved organic farms by Alter Vida and Aratex but they do not receive a certificate which they can use to show to other cotton buyers. Owning such a certificate themselves is much too expensive for the small scale farmers and is thus impossible.

What the result does tell is something about the sense of ownership of the organic producers. They apparently do feel very responsible for their organic certification and have the feeling it is something they have (achieved).

In annex 11 some more graphs are presented concerning the effects on the social and economic situation of organic producers to provide more detailed inside.

8.4.2. Environment

The expected aspects of impact of cotton certification on the environment are aquatic eco-toxicity, climate change, loss of land support functions, salinisation, biodiversity, depletion of a-biotic resources (seeds), human toxicity, terrestrial eco-toxicity and the depletion of water and land resources. Most of these environmental aspects are affected by the use of pesticides and synthetic fertilizers, the frequency of use and the toxicity of the pesticides and synthetic fertilizer used and the spraying method. Concerning

biodiversity some measures can be taken to increase the biodiversity level on the farm, such as planting of trees and fallow ground. The operationalization of the concept is further explained in annex 6.

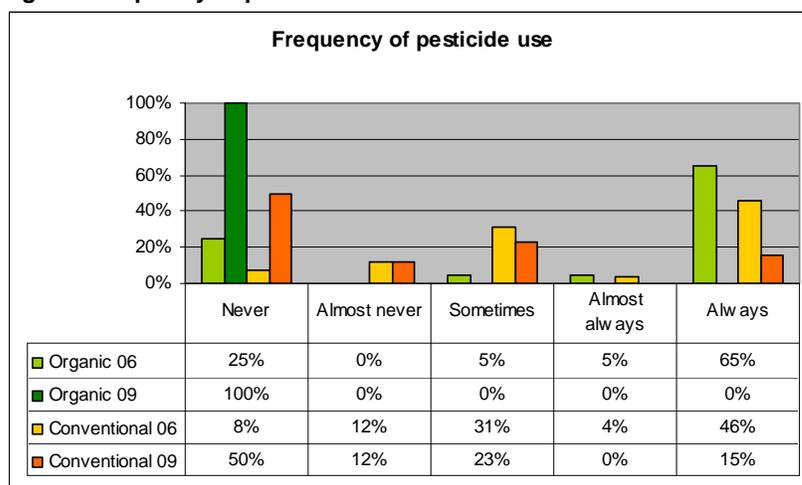
Pesticides¹⁵

The impact on pesticide use should be present through the certification scheme since it is forbidden for organic producers to use pesticides when entering the certification scheme.

To measure the impact of certification on the use of pesticides producers were asked about the whether they used pesticides, the type of pesticide they use, the amount they applied and the frequency of applying the pesticides before entering the certification process and in 2009.

In 2006 pesticides were used on a frequent basis by both producer groups. Pesticides were used by 92% of the conventional producers and 70% of the organic producers. In 2009 this declined to none of the organic producers and 50% of the conventional producers. The frequency of use also changed significantly. For organic producers the mean score on frequency changed from 2,15 to 5 and for conventional producers from 2,31 to 3,81 between 2006 and 2009. This is in further detail displayed in the figure 30.

Fig. 28 Frequency of pesticide use



N= 46 Organic N=20 Conventional N=26

For both producer groups there is a large decrease visible in the group that always used pesticides (65% and 31% decrease). The amount of pesticides that producers use has also declined over the years. Organic producers stopped using pesticides totally, but conventional producers also largely decreased the amount of pesticides they use. None of the producers uses very much pesticide any more and only 12% of the conventional producers use much pesticide. The largest number of pesticide users (62%) claim to only use a little amount of pesticides on their cotton crop.

The key-persons totally agree with the results as presented. According to them the biggest impact of organic certification is on the use of pesticides, together with the impact of knowledge and capacities. Furthermore, they see the trainings provided to organic producers in their committees as the cause for the decline in use of pesticides of conventional producers. Conventional producers learn a lot from these trainings as presented in the 'production knowledge and capacities' aspect. As a consequence they start to take over the production methods taught to organic producers. Consequently they use less pesticides and also less synthetic fertilizer. Consequently their health improves and the amount of green manure used increases as well.

¹⁵ Because farmers are forced to stop using pesticides and synthetic fertilizer already three years before entering the certification with these specific aspects the questions did not ask about the situation in 2006 and 2009 but about the situation before entering the certification process and in 2009.

Synthetic fertilizer

The use of synthetic fertilizer is also prohibited in the organic certification scheme. Therefore, the impact on this aspect is expected to be large.

The change in use of synthetic fertilizer is measured through asking the respondents about the use, the type, the amount and frequency of use of synthetic fertilizer before entering the certification process and in 2009.

In 2006 80% of the organic producers used synthetic fertilizer and 58% of the conventional producers. In 2009 none of the organic producers used synthetic fertilizer and 31% of the conventional producers still used synthetic fertilizer.

Although there is also a decline visible in the amount of producers that apply fertilizer, those that continue using it seem to apply more as before. The increase in producers that apply much is very high. It could however be that the producers that decided to quite using fertilizer were the producers that used a small amount of fertilizer already. However, not only the percentage but also the number of people using much synthetic fertilizer has increased. Before, 2 conventional producers used much synthetic fertilizer and one used very much. Currently, 5 conventional producers use much synthetic fertilizer. The type of synthetic fertilizers used in 2006 and 2009 are foliar, chemical manure and a fertilizer with the technical description 12,1272. Foliar is the most used synthetic fertilizer.

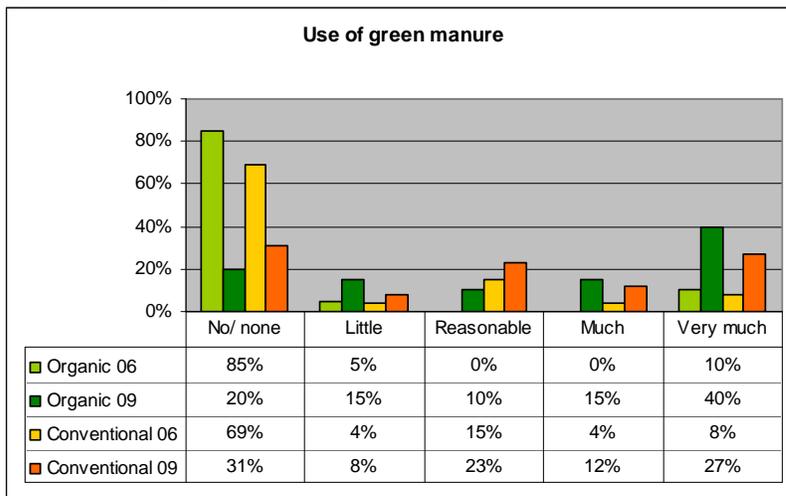
Green manure

In the certification scheme the use of green manure is promoted. Therefore the impact of the certification scheme on the use of green manure is expected to be relatively high.

The impact on the use of green manure is measured through asking them whether they use green manure, which type and with which frequency in 2006 and 2009.

The mean score in green manure has changed over the years consequently. Organic farmers score 1,45 in 2006 and 3,4 in 2009. Conventional farmers scored 1,77 in 2006 and 2,96 in 2009. The increasing score is again higher for the organic producers, although the difference is a bit smaller in this particular subject.

Fig. 29 Use of green manure



N= 46 Organic N=20 Conventional N=26

It is interesting to see that the producers that start to use green manure immediately seem to be using it very much. This is probably possible for them because the expenses are low. The type of green manure that is often used are Lupine, Moscuna, Canavalia, Havana negra and cow-poop. Also other types of plants which are found on the farm are used.

Key-persons agree that organic certification positively influences the amount of green manure used, for both organic and conventional producers. This is not only a positive change as such but also contributes to biodiversity and soil nutrition. In stead of killing all life around the crops pests are fought with natural measures and soils are fed in a natural manner.

Crop rotation & seed variety¹⁶

One of the demands of the organic certification scheme is that the organic producers cannot use treated seeds. Besides, crop rotation is promoted. Therefore, a change in both aspects is expected.

The impact on crop rotation is measured by asking the farmers whether they conducted crop rotation with their cotton crop in 2006 and 2009. The impact on seed variety is measured through asking the producers which type of seeds they used in 2006 and 2009.

In 2006 more conventional producers (85%) as organic producers (60%) did crop rotation with their cotton. In the group of organic producers we see however a rapid increase towards the year 2009, when all organic producers make use of crop rotation. The increase for conventional producers has stagnated; it remains on a high stable level of 88%.

Organic certification has clearly influenced the seed variety being used. While before 2006 the seed variety used was quite diverse, in 2009 out of 20 organic producers 19 used IAN425, an organic seed. These seeds are provided via Alter Vida to the farmers. The seed varieties used by the conventional producers were diverse in 2006 and remained so in 2009. The largest share of conventional producers however used treated seeds in both 2006 and 2009.

Health

The impact on health of the certification scheme is expected to occur because of not using pesticides and synthetic fertilizer anymore. Both severely influence the health of producers and their family in a negative manner.

The impact of the certification scheme on the health of the producers is measured by asking the producers how they applied their pesticides, how they store their pesticides and whether their health has changed in the past three years. If a change in their health situation occurred they were asked to mention the cause for this change.

All but one producer apply pesticides with a backpack with spray. This means the producers are in direct contact with the pesticides. Only 7% of the organic producers and 50% of the conventional producers made use of protection in 2006 while spraying the pesticides. In 2009 the number of conventional producers which made use of protection even declined a bit to 43%. The farmers that do make use of protection use a combination of gloves, boots, protective clothes, a cap or a mask. Most of the producers have two or three of these items. The producers that do use the protection do not always use it. Not only the amount of conventional producers, but also the frequency with which they use the protective wear has slightly declined over the years.

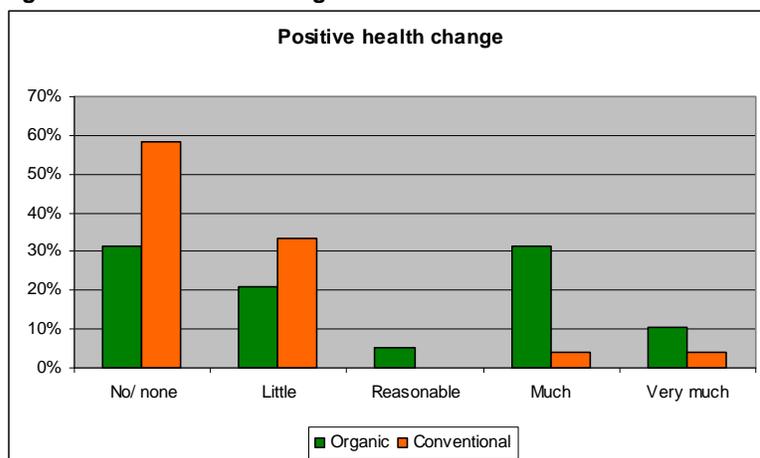
Pesticides are most of the time stored in a deposit. Where this deposit is depends per family. Most of the producers have a separate deposit outside of the house. Some of the producers store their pesticides in trees or between plants around the house. Very few producers store their pesticides within the house. The storage place for pesticides has not changed over the last years.

The frequency with which farmers are exposed to health risks when using pesticides is perceived much higher by the organic producers in 2006 as it is by the conventional producers. This can be explained by the fact that the questions are posed after the fact. Organic producers have by now learned about the danger of pesticide application, in contrary to most conventional farmers. Therefore, they perceive their risk much higher as conventional farmers do. Conventional producers currently perceive their risks a little bit less as three years ago.

Producers were also questioned about the change in their health situation over the last three years. From the 20 organic producers 12 said they had undergone a positive change. From the 26 conventional producers this was the case for 10 of them.

¹⁶ During the interviews it showed that many producers do not know which seed variety they are exactly using. They often only knew part of the name which forced me to interpret the results a bit. Since Alter Vida provides the seeds to all organic producers it might seem possible that the one organic producer that mentioned to use a different seed variety might have made a mistake. Nonetheless, the results are clear and speak for themselves.

Fig. 30 Positive health change



N= 43 Organic N=19 Conventional N=24

Most conventional producers believe that their health has improved a little bit. Most organic producers on the other hand believe that their health has improved much or even very much. Almost all organic producers mention as the cause for this positive change the fact that they stopped using pesticides. Of the conventional producers 50% also mentioned stopping with using pesticides as the cause for the positive change. The other 50% named different causes.

Native forest, reforestation and fallow land

As part of the demands of the certification scheme producers have to take some biodiversity improving measures in their farm. That can either be preserving their native forest, plant trees or assure that a certain amount of hectares is in rest.

For measuring this aspect producers were asked about the amount of hectares they used for the different biodiversity measures in 2006 and 2009.

Table 12 Biodiversity measures – Change in mean hectares used

	Organic Yes/No	N	Mean
Native forest 06	Yes	20	1,7500
	No	26	1,6923
Native forest 09	Yes	20	1,8500
	No	26	1,6538
Reforestation 06	Yes	20	1,6500
	No	26	1,7308
Reforestation 09	Yes	20	1,9500
	No	26	1,7308
Fallow land 06	Yes	20	2,2500
	No	26	1,7692
Fallow land 09	Yes	20	1,7500
	No	26	1,6923

The mean score of conventional producers has remained almost the same on all subjects. The amount of land covered with native forest and reforested land has slightly increased for organic producers. The hectares of ground in rest however have declined.

Cultivated land

In general there is a downturn visible in the amount of cotton hectares cultivated in Paraguay. Therefore, it is interesting to see whether the downturn amongst organic producers is less because of the benefits they receive from producing in an organic manner.

Producers were asked about the amount of hectares they cultivated with cotton in 2006 and 2009 and their forecast for the coming year.

The mean hectare of land owned by organic producers is larger as the mean hectare owned by conventional producers. For organic producers this is 12,25 while for conventional producers this is 9,86. The hectare of cotton cultivated by conventional producers in 2006 was higher as those cultivated by organic producers. For both types of producers the amount cotton cultivation decreased; from 1,9 to 1,18 for organic producers and from 2,85 to 1,88 for conventional producers. Although conventional producers originally cultivated more hectares the decline is more or less the same. The data about the planned cotton cultivation for the coming season prelude a bigger decline in cotton cultivation for conventional producers (from 1,88 to 0,82) as for organic producers (from 1,18 to 0,96).

Out of the interviews with the key persons the same tendency came to the front, producers are producing each year less cotton. Low cotton prices, degraded soils and failed harvest due to climatic issues have caused farmers to change to other crops.

The decreasing cotton production could be a future problem to cotton buyers. Manufactura Pilar, a nearby cotton buyer is already taking measures in order to stop the decrease in cotton production. Actions taken are price incentives and currently they are running a test with organic cotton. They are unable to become part of organic cotton certification, because of the high costs, but nonetheless producing organic cotton has many positive impacts on farmers. This could motivate them more to continue producing cotton. In the same line of reasoning some of the key persons claim that organic farmers are not decreasing, but increasing the amount of cotton produced.

In annex 12 some more graphs on the effects of organic cotton certification on the environment are presented to provide more detailed insight into the results.

9. Test Results: Methodological considerations towards the balanced measurement system

The results from the test concerning the effects of organic cotton certification on the economic, social and environmental situation of the producers are presented. However, the main purpose of the test was to gain more insight in the functionality of the measurement system I have designed. The main goal was to study whether the characteristics attached to validity, reliability and manageability are appropriate and sufficient to address these issues. This will be discussed subsequently.

9.1 Validity

External validity is about whether the results can be externalized to other populations. In the measurement system this was hoped to be accounted for by drawing random samples, thick description and critical and reflexive practices of the researcher in order to discern possible influences on the outcome.

Drawing a random sample proved difficult. The population was determined with help of Alter Vida. However, it showed that they do not have lists with participating organic farmers, they know them by heart. Moreover, they mostly work with organic producers and not with conventional producers. Consequently, drawing up a list from which the control group sample could be drawn proved even more difficult. Finally, we were able to create a list for the organic producers and draw a sample from there. But practice also showed that due to practicalities, such as logistics and people that were not at home we sometimes differentiated from this list. Moreover, for the conventional producers we were unable to draw up a list and were forced to make use of a transactional ride/walk. Concluding, drawing a random sample proves difficult in a developing world context.

However, by doing thick description and taking a critical and reflective attitude it is still possible to obtain insight in the local situational factors that might have influenced the results. With this insight it is also possible to get grip on whether the results can be generalized. Context specific factors can be appointed and consequently compared to other contexts. Based on this comparison the results can or cannot be generalized to other populations.

In this research the fact that the production trainings are provided in committees which often consist of organic and conventional producers is a very important contextual factor. If this was not the case the effects of the certification on conventional producers would have been far less. The importance of describing this local contextual factor is herewith stressed.

Construct validity determines whether the dependent and independent variable accurately measure the construct of interest and not the construct of disinterest. In the measurement system this is accounted for in the funded theoretical basis of the sustainability aspects.

Making statements about whether construct validity is attained, is difficult because one can never be 100% certain that the total construct of interest is measured. However, a good theoretical foundation for the different aspects is a good starting point. Moreover, it showed during the research that it is importing to crosscheck the local meaning of the different concepts used to measure the aspects. When the local perception of the concepts or terms used to measure the sustainability aspect is different to how it is defined in the research, constructs of disinterest are measured which will negatively influence the construct validity. In order to overcome this problem it is of utmost importance to test the questionnaire and discuss the local perception of concepts and terms used with local stakeholders to assure clear understanding. Finally, cross-checking during the interview can also prove valuable.

Internal validity is realized when we are measuring what we claim to measure and there are no influences from external variables. In the measurement system this is accounted for via triangulation and reflexive practices of the researcher as well as respondent validation. Triangulation was meant to be realized through contextualizing the results through documents, key-person interviews and studying previous research. All these methods seem to serve their purpose very well. Respondent validation was not tested. However, it still is very important to conduct this step in future research because it can provide more insight in the accuracy of the conclusions drawn.

Besides, the influence of external factors on the results is controlled for by addressing the attribution problem. This is brought into practice by making use of a target group and control group, two separate communities, through asking about the current situation as well as the past and through asking questions about causal explanations.

Although it proved impossible to find separate communities that only had target or control group respondents the use of these two groups was very valuable. The possibility to compare the results of both groups enables the researcher to clearly discern the influence of certification. The time dimension clarified the situation even more. Both obvious as well as less obvious impacts could be discerned. The results

obtained coincided in large with the opinion of key-persons, which creates more certainty and trust in the results and consequently also in the method used.

However, the test of the system also showed the huge importance of the questions about causal explanations and triangulation. The questions about causal explanations largely influence the internal validity. It offers the opportunity to state with a high level of certainty that the observed change is being caused by organic certification or not. Triangulation is very valuable in gaining more insight in the local context. With this insight the measured effects can be better understood and attributed. Having more insight in and understanding of the context in which the changes take place, increases the value of these results.

Table 13 Final validity characteristics

Validity	Characteristics	Test Lessons/ Conclusions
Construct	Theoretical Foundation Multi-criteria analysis	Test & adjust questionnaire to local perception Cross-checking during interview
Internal	Reflexive and rigour practice Respondent Validation Triangulation: - key-person interviews - documented data Attribution problem: - Target & Control Group - Two communities - Before and after - causal explanation questions	Measures taken serve their purpose well - No separate communities but still very valuable Importance of: - Causal explanation questions - Key-person interviews (triangulation)
External	Random sample Thick description Critically reflect upon results	Drawing random sample is difficult – local organisation Thick description very important – e.g. mixed committees

9.2 Reliability

Reliability determines the extent to which the results are free from random error. In the measurement system this is hoped to be accounted for through systematic, logical and well documented research. It is however very difficult to determine up to which level this succeeded. Crosschecking the answers during the interview can overcome some of the random errors. While conducting the questionnaires it proved very important to crosscheck all the time whether the questions were correctly understood. Sometimes answers on two similar types of questions collide. In this case it is important that the researcher takes account of it and checks whether the answers given are correct. Besides, analyzing the gathered data before filing it in SPSS can also filter out some random errors.

Table 14 Final reliability characteristics

	Characteristics	Test lessons/ Conclusions
Reliability	Research design: - Consistent - Systematic - Logical - Well documented - Cross- check	Difficult to determine Cross-checking important Analyze data before filing into SPSS

9.3 Manageability

Manageability in this research is determined by the applicability of the measurement system in a development world context, the costs and the inputs (time, technical staff etc.) needed to be able to conduct the measurement. The method of result representation and the comparability of the results are also an important aspect of manageability.

In the measurement system manageability is realized through the use of a Likert scale system. The inputs for the use of this system are low: The time to conduct a questionnaire is short, the questionnaire format is easy and can be conducted by inexperienced researchers, the representation is clear and the results do not depend on the availability of external documented data.

From the test of the system in Paraguay all the values attributed to the use of Likert scale type questions as explained in the latter are proved in reality. The questions are easy to use, the information gathered can be practically and easily filed and used and transformed in clear graphs and figures, the length of interviews can be kept short and the results do not depend on external documented data. Moreover, because of the short answers many questions can be asked which allows a broad research and crosschecking via multiple questions on one topic. As such it does not only contribute to the manageability but also to the validity of the results.

The different options for processing the results make them not only very manageable but very useful as well. Depending on the wishes of the researcher the results can be studied in depth or a general overview can be obtained. This multi-functionality is a very valuable aspect of the system. Finally it also makes the results easy to compare between regions. Means and graphs can be easily put together and be compared.

It is however also important to stress that the results from the Likert scale questions are not so valuable without the causal explanation questions and the key-person interviews. The Likert scale-questions provide you with focused data, which makes it more difficult to understand the background and causes of changes measured. Therefore, inserting questions about causal explanations and conducting key-person interviews is very important.

Accessibility and sample location are also part of the manageability. For addressing these aspects the six step research schedule proved very valuable as is discussed subsequently.

Table 15 Final manageability characteristics

	Characteristics	Test lessons/ Conclusions
Manageability	Likert scale questions Research schedule: Six steps	Likert scale questions is a positive choice → Low inputs (time, costs, staff) No dependence on documented data Clear representation without losing its explaining capacity Six steps function in practice

The six steps research schedule

The research design as presented for conducting the research itself turned out to function well in the test-case. The different steps assisted in conducting the research in a structured manner and delivered sufficient and useful data on which different conclusions can be based. Consequently, based on the test no adjustments are needed in the six steps as presented.

However, the test did shed light on some of the most important and central issues which were very important for conducting the test in a successful manner. One of these focal points is the importance of establishing useful contacts in the research nation. Especially contact with one of the main actors in the local certification process is recommended. Their assistance proved very valuable in different areas: finalizing the questionnaire, contacting key-persons, determining the research area, drawing random samples, accessibility in terms of infrastructure and in social terms to the research population and other stakeholders, assistance with all sort of infrastructural issues, with language issues and all others sorts of

issues that come along while conducting a research. I do not claim that without their help it would have been impossible to conduct the research. With some personal effort many things can be achieved also without the help of a local organisation but it did make all the steps that had to be taken easier.

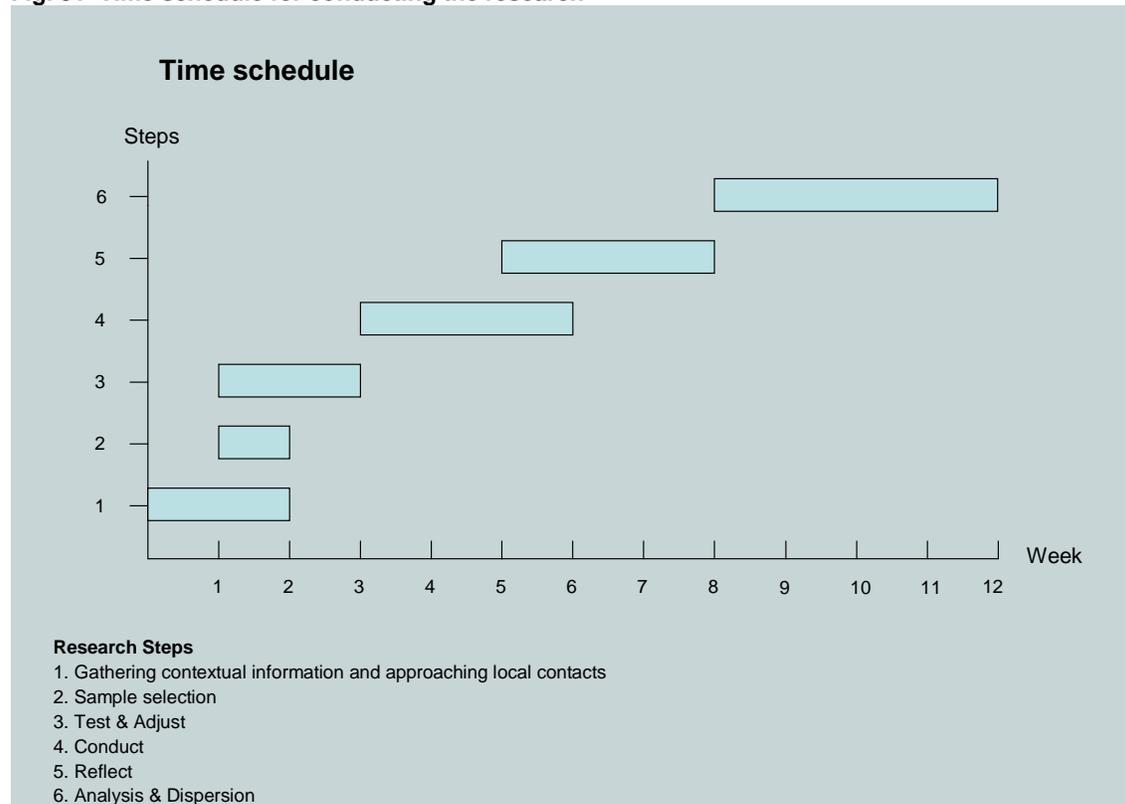
A second focal point in conducting the research that has also been put forward while discussing the validity and reliability of the results is the importance of discussing the questionnaire with local actors and testing the questionnaire with great precision. By doing so many errors and misunderstanding can be filtered out which gives more certainty that you really measure what you claim to measure.

A third focal point is that during all stages of the research good planning is of major importance. As a researcher you should always aim to think a few steps ahead to assure that the process will develop smoothly. There are always things that can suddenly occur on which you did not account on beforehand. For example, the fact that, aside from the language problem, the producers did not want to meet for a group meeting in the week that I had planned because they were in harvest season of some other crops they were growing. Because the producers were very willing to participate during the whole research I did not think there would be any problems. But I did not account on them to be in harvest season. With more planning I might have foreseen this problem and would have been able to adjust my program.

A fourth and final focal point are that of time and statistical significance. Both focal points are closely related to one another and to the issue of manageability.

The time that is needed to conduct an interview is relatively little; between 45 minutes and 1 hour and 15 minutes. More or less 25 interviews can be conducted a week, based on full access to transport and a five day work week. Some preparation time is needed and off course time to file all the data and present the conclusions. The timeline for conducting the whole research will look as follows¹⁷:

Fig. 31 Time schedule for conducting the research



Statistical significance

Almost none of the results proved to be statistical significant. The cause for this is most likely the relatively small n of the two samples. The reasonable solution for this problem is drawing a larger sample, which means more respondents. However, more respondents will have its impact on the time and money inputs

¹⁷ The preparation time at home for the research process is not part of the time schedule.

and thus on the manageability of the research. The question is thus how manageable the research is when we increase the n up to such a level that we are likely to obtain statistically significant results.

As presented in the table the process of conducting the research for 50 respondents is eleven weeks. The two processes that are influenced when increasing the n are step 4, conducting the research, and step 6, analyzing the results. The preparatory work will remain the same. The number of key-persons interviewed does not have to increase since this is not influenced by the size of the n , unless part of the questionnaires are conducted in a different region. The number of group-meetings can also stay the same because only a sample of the respondent groups will have to participate.

Consequently, if the n is increased to a total population of 100 either the amount of weeks of step 4 has to be doubled, or more researchers need to be hired. The same is true for step 6. It is also important to note that only 1,5 week of the 4 weeks displayed in step 6 are for filing the data of the questionnaires, the rest is for analyzing the results and writing the report. The amount of extra weeks necessary for doubling the research population is 4,5.

The question then remains at which size of n the results will be statistically significant. It is difficult to make any statements about this because it depends on the certainty one wants to realize and on the size of the total research population. It can consequently differ per product group and research area. The time schedule however does present a time indication for the different sizes of n .

In respect to both the level of statistical significance and the costs spend on the measurement system it is difficult to come up with clear cut data based on my research results. However, the time indication can form a basis for an indication of costs. Besides via assuring manageability the inputs and thus the costs are kept as low as possible while still realizing validity and reliability. Moreover, different approaches can be sought to keep costs low, such as the use of students for conducting the research.

9.4 System specific considerations

Different lessons about the system and its usefulness in practice were learned during the test in Paraguay. One of these issues is how to make best use of the key-person information.

As mentioned before it is important to interview key-persons throughout the whole research project. They can provide very valuable information throughout the whole process. However, it is also very important to remain an open focus towards the information they provide. Key-persons can just as well overlook certain issues, or misinterpret reality. Besides they can be biased depending on their stakeholder-ship. It is important for the researcher to pay attention to this when conducting the research and analyzing the results. The researcher should make notes of the position of the key-person towards the subject and record situations in which a certain perspective comes specifically to the front.

The researcher should also pay careful attention to the answers provided by the respondents. Because of the use of answer categories errors are less on the surface. To overcome this problem and note errors in answers multiple questions about one subject proved very valuable. It provides a method of control.

This cross-checking is especially important since some of the concepts were not clearly understood by the respondents. The difference between producing capacities and knowledge was hardly understood. Often the producers understood production capacities to mean to be the possibilities they have to produce well. With this they mainly aimed at sources such as land, the climate, production equipment available to them, etc. Besides the word investments was understood in a different manner. From their perspective investments do not necessarily comprise monetary investments but can be in kind as well. Furthermore, some respondents perceived expenses like school fees as an investment. This was not the definition of investments I had in mind. Not noticing this could have lead to wrong conclusions. A different concept that was hardly understood proved to be crop-processing. It was difficult to explain to the producers what was meant with this concept. All these examples stress the importance of discussing the concepts used in length with a local stakeholder or key-person before using the questionnaire in order to overcome as many misunderstandings as possible.

In this specific test-case language turned out to be a problem. To assure that no misunderstandings would take place the people from Alter Vida helped me with translating whenever necessary. A positive lesson learned from this language problem is that no high level of knowledge on the local language is needed to be capable to conduct the questionnaires. For interviewing the key-persons however a reasonable knowledge is required.

The order in which the questions were asked is also a subject of discussion. In the questionnaire I started with questions concerning equity between husband and wife. Before asking these questions it is firstly important to ask whether the respondent has a husband or wife or partner. If not, the questions do not make sense. Secondly, one could argue that it would be better to ask these types of questions later on in the interview. The subject can be a sensitive issue which necessitates trust and an open attitude. At a later point in the interview the chance is higher that this is the case. Nonetheless, in Paraguay there did not

seem to be any issues with answering the questions in such an early stage in the interview. Moreover, the answer- scales are different for this aspect as for the others. It might be better not to mingle the aspects with different answer categories to prevent confusion.

9.5 Question specific lessons

With the aspect of labour demand it would be better to firstly ask which amount of family members are working on the farm before asking whether they are hiring any labourers. This order is preferred because of the more sensitive approach. Besides, it would be better to not only differentiate between family and external labourers but also between the working hours that different family members spend on the farm. In this manner more certainty can be obtained about what has really changed on the farm with respect to labour input.

With respect to food security it was difficult to find difference in the frequency with which certain types of food were consumed. However, the level of food shortages did change significantly. Therefore, it would be interesting to differentiate between food eaten from their land and expenses on food. Moreover, the question concerning the expenses on food should be more precise. Instead of asking about the frequency with which money is spent it will make more sense to ask about the amount of money spent. One farmer could spend a large amount of money but just once a week, while another farmer might spend very little every day. When asking about the frequency the second farmer gives the impression to spend more while this is not the case. Therefore, in future research the questions should be about the amount spend, rather as the frequency of spending.

With respect to change in income a question should be added to ask about the cause of change. In other aspects this proved very valuable for obtaining insight in the actual cause of change. Besides cotton income should not be determined in general but rather per kilo. In this manner it is easier to establish what the price difference is for both producer groups. Moreover, it is also very important to add a question about yields per hectare because this can also influence the income. While the price of cotton might not go up, the yields can be improved through new production methods which will lead to increased incomes.

The questions concerning the road quality should be put in a more relative manner. In their current form it was sometimes difficult for the respondent to choose one of the options available. Therefore the questions should be put in such a manner that it ask about the total amount of a certain type of road in the community in relation to the total amount. In this manner it becomes easier for the respondents to given the correct answer.

With respect to cotton production it is very important to ask the hectares of cotton cultivation throughout the years. In Paraguay it showed that cotton production is declining. It is very interesting to gain insight in the differences of the development of this trend for the two producer groups. Moreover, it could be a causal explanation for declining incomes.

Finally, in future use it would be advisable to discuss the environmental aspects with an expert to assure that the right and locally important aspects are measured. Issues such as the types of pesticides, chemical fertilizer etcetera can be discussed. But also the possible impact on the quality of cotton and other issues that could be affected by the certification.

10. Conclusion & Discussion

10.1 Conclusion

The research question as presented in the first chapter of this thesis is:

What characteristics should a monitoring system have in order to measure 2nd and 3rd order effects of Sustainable Supply Chain Governance-systems, functioning in a developing world context, in a valid, reliable and manageable way?

Besides answering this question, I hoped to overcome several often prevailing issues in designing a measurement system for SSCG-systems. For attaining the three aspects; validity, reliability and manageability, several characteristics were proposed throughout the research and subsequently tested on organic cotton certification in Paraguay. The conclusion presented here will therefore be focused on whether these characteristics are useful and sufficient to attain the three aspects. Which aspects should be adjusted or focused on, and can the measurement system be perceived successful for future research?

The characteristics proposed to realize construct validity were a strong theoretical foundation of the sustainability aspects that are being used and multi-criteria analysis. The strong theoretical foundation of the sustainability aspects is a good manner of realizing construct validity. However, the test showed that this is not sufficient if the concepts are not correctly translated to the local language and meaning. The diverging interpretation of concepts and words used in the questionnaire between the researcher and its respondents reaffirm the importance of local context. This problem can be overcome, and consequently construct validity realized, through crosschecking the local interpretation of the concepts and wording used in the questionnaire both before and meanwhile conducting the research. Examining the questionnaire with local stakeholders, precisely testing the questionnaire before taking it in use and cross-checking during the interview are of utmost importance for realizing construct validity.

The characteristics proposed to address internal validity were; the use of a target and control group, questions about the past and current situation, causal explanation question, triangulation through key-person interviews and documented data, respondent validation and reflexive and rigour practice of the researcher. These measures prove very valuable in gaining more confidence in the level of causality between the dependent and independent variables. It is a very manageable and widely applicable strategy to obtain a large amount of certainty about the effects a certification system can have on its local environment. It is especially valuable in those contexts where there is a lack of statistical data or baseline information, such as often the case in development world contexts.

It is however important to note that in case of using this strategy in combination with Likert-scale questions the importance of the questions concerning causal explanations and triangulation via key-person interviews are tremendous. Without the causal explanation questions many effects might have been interpreted wrongly or remain unclear. The key-person interviews contribute in large to the understanding of the system and its functioning as a whole. Together these aspects add in a large to the insights obtained.

One of the proposed measures to realize internal validity was not applicable in the test case. The organic and conventional producers are not living in a separate community which has negatively influenced the capacity to make statements on community level. When the two producers groups cannot be divided on community level it becomes harder to attribute the change measured on community level to the certification scheme. It becomes almost impossible to control for all the possible external influences. Therefore, less structural side effects could be measured and consequently less insight was obtained in 3rd order effects. In cases where this division is possible it is advised to gather information on community level to enable more statements about 3rd order effects. When this is not possible the 3rd order effects should be based on the aspects that touch upon this topic, such as labour demand and investments. When the local situation is known on beforehand more community focused aspects can be included to obtain more insight in 3rd order effects. Besides, information about the effects on the community can be filtered from the key-person interviews.

A final measure that was proposed was to validate the research results with the research population via focus group meetings. Due to language problems and planning problems this measure was not applied in the test, which stresses the importance of a good detailed planning and careful examination of the research area. Nonetheless, I do recommend making use of this validation method in future research. It can improve the confidence about the internal validity of the research results and moreover provide useful insights in possible areas of improvement for the certification system. This is especially considered important when one wants to make the step from a proving to an improvement approach.

To realize external validity the following characteristics were proposed; drawing a random sample, thick description of the local context and critically reflecting on the results. Drawing a random sample proved somewhat difficult, even when help is available from local organisations. It depends to a large extent on the availability of data about the research populations. Therefore, thick description of the local context proved to be very important to be able to generalize the results. Thick description can appoint context specific factors that might have influenced the results. In the test-case the mixed producer committees proved to be one of these important context specific factors. Addressing the context specific factors offers the opportunity to generalize the results to other contexts after analyzing how these context specific factors apply in the target area.

Assuring full reliability was hoped to be realized via a consistent, systematic, well documented research and via cross-checking the results. It is difficult to make any hard statements about whether this has succeeded. The researcher has a large responsibility in doing what lies in his capabilities to prevent random error. Through paying close attention to the answers during the interviews and by analyzing the results before filing them in SPSS some of the errors can be filtered out.

Manageability was hoped to be realized through the use of Likert-scale questions and through the creation of a six steps research schedule for conducting the research. The use of this method has shown to be a very positive choice in the eye of realizing manageability. It offers the opportunity to obtain a large share of information in a short amount of time. Consequently costs are kept low. There is no dependency on the availability of documented data and the results can be presented in a clear manner without losing its explaining capacity. This makes the results useful for future policy-making and thus coincides with the improvement approach.

The six steps research schedule proved useful in practice. The test clearly appointed some of the focus points while conducting the research. The assistance of a local organisation proved to be very helpful throughout the whole process and is strongly recommended. The testing of the method with local stakeholders is also very important to filter out possible biases in the results. Besides it also brought up one of the important issues in creating a sound measurement system that was not addressed as an aspect of importance initially; statistical significance.

Almost none of the results were statistically significant. To solve this problem the n should be increased, which will negatively influence the time and costs inputs. The question consequently is whether the measurement system will still keep its manageability? Answering this question depends in large on the demands of the researcher or research initiating institute. It depends on the value they attach to conducting a valid, reliable and statistically significant research. If they do perceive that to be important in my opinion this is a viable measurement system to make use of.

Finally, the test in Paraguay addressed the importance of the characteristics of the questionnaire. The clarity of questions, the order of questions, the concepts used and how the aspects are operationalized all are important aspects in the quality of the questionnaire.

All the characteristics that should be attributed to a valid, reliable and manageable measurement system are presented. The system showed to function well and fairly valid and reliable results are obtained. The mix of different research methods to create a measurement system for the specific SSCG-system context showed to function in practice. However, future use can only prove the usefulness of the system for the different actors within the SSCG – systems. The level to which it will be perceived applicable and useful will most likely depend on the goals of the researcher or researching-institute. When they attach value to a valid and reliable measurement the measurement system as presented is a viable option. However, if they do not attach value to this they can organize a few focus meetings to obtain a general impression in a much smaller amount of time and thus with less costs. The same is true for the level of statistical significance the researcher wants to achieve. Much, and also the future use of this measurement system, depends on the demands and choices made by the researcher or the research initiating organisations.

10.2 Discussion

In the construction of a valid, reliable and manageable measurement system many different aspects play an important role. The newness of the subject makes it that many of those aspects can be put up for discussion and there is a large arena of future research that could be conducted:

1. Concerning the statistical significance of the results it would be interesting to gain more insight in the sample size necessary to realize statistical significance. Further tests with the system as well as more literature study on statistical significance can shed more light on this aspect.
2. Using a larger sample size in future test can also help explore the boundaries of the possibilities that the measurement system offers. With an increased sample size the statistical program SPSS

can be used to its full potential. More statistical tests can be conducted which can provide even better insight in the reliability and validity of the measurement system.

3. Closely related to this aspect of sample size is the viability of the measurement system in a market or civil society context. The goals of achieving validity, reliability and statistical significance are science based demands. Although in this system a synthesis is sought between these goals and the manageability of the system it remains questionable how viable the system is for market actors within the supply chain.
4. The measurement system has so far been only tested on one product in one area. Many interesting insight can be obtained from applying the system on different products groups in different areas. The functionality of the system can be further addressed as well as the comparability of the results amongst product groups and regions.
5. One of the areas of improvement in the measurement system is that the measurement of effects on the environmental aspects. If time allows it, the best approach would be to operationalize these aspects in cooperation with an expert in this field to realize more accuracy in the measurements. Besides it is very important to realize that these aspects might have to be adjusted to each specific product group and context.
6. An issue that was not addressed in this measurement system is that of self-selection bias. This is about the difference that might naturally exist between those producers that decide to enter in a certification scheme and those that do not. One could claim that producers that enter certification schemes take more initiative and are better capable of improving their own economic position. According to this line of reasoning these producers would have also had a better economic/social position if they would not have entered the certification scheme. This issue is very difficult to overcome in those cases where individuals choose to enter certification themselves. Only when a measure, in this case certification, is implied on the total population this bias will not play a role. In future research more attention could be paid to this aspect and measures can be sought to overcome the problem.
7. During the test of the measurement system in Paraguay it showed that there were several concepts and questions used in the questionnaire that were not fully understood by the respondents. Besides, the order of questioning the different aspects is also an issue up for discussion. It is recommended to study some of the existing literature on these aspects when further developing the measurement system.¹⁸
8. Because of the lack of measurement systems to measure 2nd and 3rd order effects in SSCG-systems this research was about developing such a measurement system. An interesting future approach could be to take two or three other measurement systems that are in development and compare their effectiveness on realizing validity, reliability and manageability. New insights from such a test can be obtained and it offers the opportunity to state with more certainty why one measure is considered 'better' as the other.

Finally there is a general issue that is not focused on the methodological aspects of the measurement system but places the subject in a larger perspective. This issue concerns the question which agent is responsible for conducting these types of measurements within SSCG-systems. Should it be the market, as the main actor within SSCG-systems, the government, from its caretaking role, or civil society from its watch-dog position? Different arguments could be proposed for keeping either of these actors responsible.

¹⁸ Sheskin I.M., 1985. Questionnaire development. Chp 4 in: Geographers, survey research.
Willis, K., 2006. Interviewing. Chp 15 in Doing development research. SAGE Publications: London
Simon, D., 2006. Your questions answered? Conducting questionnaire surveys. Chp 17 in: Doing Development Research. SAGE Publications: London
Laws et al., 2007. Research for Development. A practical guide. SAGE Publications: London
Lindsay, M.J., 1997. Questionnaire, interviews, and allied techniques: getting information from people. Chp 4 in: Techniques in Human Geography. Routledge: New York.

REFERENCES

- Albersmeier et al., 2009. Evaluation and reliability of the organic certification system: perceptions by farmers in Latin America. *Sustainable Development*. Vol. 17, p. 311-324.
- Banuri, T., 1998. Pakistan: Environmental of Cotton Production and Trade. [online] 06-06-09. Retrieved from: http://www.tradeknowledgenetwork.net/pdf/pk_banuri.pdf
- Barrientos, S., 2003. Labour Impact Assessment: Challenges and opportunities of a learning approach. EDIAIS conference, University of Manchester, November 2003. [online] 04-05-09. Retrieved from: <http://www.sed.manchester.ac.uk/research/iarc/ediais/pdf/EINJuly04.pdf>
- Bauman H., Tillman A., 2004. *The Hitch Hiker's Guide to LCA*. Studentlitteratur: Lund
- Beazley H., Ennew J., 2006. *Participatory Methods and Approaches: Tackling the Two Tyrannies* Chp 20 in: *Doing Development Research*. SAGE Publications Ltd.: London
- Bitzer et al., 2008. Intersectoral partnerships for a sustainable coffee chain: Really addressing sustainability or just picking (coffee) cherries? *Global Environmental Change*. Vol. 18, p. 271-284
- Briassoulis, H., 2001. Sustainable Development and its Indicators: Through a (Planner's) Glass Darkly. *Journal of Environmental Planning and Management*. Vol. 44 (3), p. 409-427
- Bryman, A., 2004. The nature of qualitative research. Chp. 13 in: *Social Science Research Methods*, p. 266-290. Oxford University Press Inc.: New York.
- Bryman, A., 2001. *Social Research Methods*. Oxford University Press.
- BSCI, 2010. The BSCI Code. [online] 02-02-10. Retrieved from: <http://www.bsci-eu.org/index.php?id=2020>
- Carney, D., 1998. Sustainable Rural Livelihoods: What Contribution Can We Make? Paper Presented at the DFID Natural Resources Advisers Conference, July 1998 London.
- Chambers, R., 2007. From PRA to PLA and Pluralism: Practice and Theory. IDS Institute of Development Studies. Working Paper 286.
- Chambers, R., 1994. Participatory Rural Appraisal (PRA): Challenges, Potentials and Paradigm. *World Development*. Vol. 22, No.10, p. 1437-1454.
- Cramer, J.M., 2008. Organizing corporate social responsibility in international product chains. *Journal of Cleaner Production*. Vol. 16, p.395-400.
- Creative Commons, 2010. Value chain map organic cotton. [online] 12-04-10. Retrieved from: http://images.google.nl/imgres?imgurl=http://upload.wikimedia.org/wikipedia/commons/7/71/Value-chain-map-organic-cotton.png&imgrefurl=http://commons.wikimedia.org/wiki/File:Value-chain-map-organic-cotton.png&usq=__xiuOr9QqLR5M8Vw-u9yfcf2Ytl=&h=1700&w=2329&sz=168&hl=nl&start=27&sig2=1gTbuRgnOAtbwB449IcjtQ&um=1&itbs=1&tbnid=QGpGhM4oQKCJZM:&tbnh=109&tbnw=150&prev=/images%3Fq%3Dcotton%2Bchain%26start%3D18%26um%3D1%26hl%3Dnl%26sa%3DN%26ndsp%3D18%26tbs%3Disch:1&ei=3FTZS6ztEYSIOlio8boP
- Custance, J., Hillier, H., 1998. Statistical Issues in Developing Indicators of Sustainable Development. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. Vol. 161, No. 3, p. 281-290
- Daly, H. E. (1990) *Toward Some Operational Principles of Sustainable Development*, *Ecological Economics*, Vol. 2, p. 1-6.
- Dietze, R., 2009. Programa Nacional de Algodón. Periodo Agricolo 2009/10 y 2019/20.
- Eyhorn et al., 2005. The Impact of Organic Cotton Farming on the Livelihoods of Smallholders. Evidence from the Makaail bioRE project in central India. Research Institute of Organic Agriculture. [online] 06-06-09. Retrieved from: <http://www.fibl.org/fileadmin/documents/en/development-cooperation/production-systems/maikaal-research-report.pdf>

- Feller et al., 2006. Value Chains versus Supply Chains. BPTrends. [online] 06-07-09. Retrieved from: <http://www.ceibs.edu/knowledge/papers/images/20060317/2847.pdf>
- Ferrigno et al., 2010 forthcoming. Spinning the real farmers Yarn. Farmer self-assessment methodology to measure progress in organic cotton production. A Training Manual for facilitators' and farmers (personal communication).
- Ferrigno S., Lizzagara A., 2009. Components of a Sustainable Cotton Production System: Perspectives From the Organic Cotton Experience. ICAC Recorder March 2009, p. 13-23
- Ferrigno, S., 2009. New Self-Assessment Scorecard (personal communication).
- Ferrigno, S., 2008. KPI Scorecard (personal communication).
- Ferrigno et al., 2006. Sowing the Seeds of Change: Weaving Innovation and Integrity into Organic Agriculture. Organic Exchange: USA.
- Ferrigno et al., 2005. Organic Cotton: A new development path for African smallholders? IIED Gatekeepers series 120. [online] 06-06-09. Retrieved from: http://www.povertymonitoring.go.tz/documents/IIED_on_organic_cottons_potential_in_Tz.pdf
- Giovannucci D., Potts J., 2008. Seeking Sustainability. COSA Preliminary Analysis of Sustainability Initiatives in the Coffee Sector. IISD. [online] 08-07-09. Retrieved from: http://www.iisd.org/pdf/2008/seeking_sustainability.pdf
- Glin et al., 2006, Living with Poison: Problems of Endosulfan in West-African growing systems. [online] 06-06-09. Retrieved from: <http://www.pan-uk.org/Projects/Cotton/Resources/downloads/livingwithpoison.htm>
- Guinée et al., 2002. Handbook on Life Cycle Assessment, Operational Guide to the ISO Standards. Kluwer Academic Publishers; New York.
- Hilderink, H.B.M., 2004. Towards human and social sustainability indicators, A multidimensional approach. RIVM report 550012002: Bilthoven.
- Huitt, W., 2004. Maslow's hierarchy of needs. Educational Psychology Interactive. Valdosta, GA: Valdosta State University. [online] 08-06-09. Retrieved from: <http://chiron.valdosta.edu/whuitt/col/regsyst/maslow.html>
- Hutchins M.J., Sutherland J.W., 2008. An exploration of measures of social sustainability and their application to supply chain decisions. Journal of Cleaner Production. Vol. 16, p. 1688-1698
- ICAC - International Cotton Advisory Committee 2009. Structure of World Cotton Trade. [online] 10-04-10. Retrieved from: http://www.icac.org/cotton_info/publications/reviews/articles/structure_wct09.pdf
- ICAC - International Cotton Advisory Committee, 2008. Literature Review and Research Evaluation relating to Social Impacts of Global Cotton Production for ICAC Expert Panel on Social, Environmental and Economic Performance of Cotton (SEEP). [online] 23-05-09. Retrieved from: http://icac.org/seep/documents/reports/literature_review_july_2008.pdf
- ICAC - International Cotton Advisory Committee, 2008. Literature Review and Research Evaluation relating to Social Impacts of Global Cotton Production for ICAC Expert Panel on Social, Environmental and Economic Performance of Cotton (SEEP). [online] 06-06-09. Retrieved from: http://www.icac.org/seep/documents/reports/literature_review_july_2008.pdf
- International Crisis Group, 2005. The Curse of Cotton: Central Asia's Destructive Monoculture. [online] 23-05-09. Retrieved from: <http://www.crisisgroup.org/en/regions/asia/central-asia/093-the-curse-of-cotton-central-asias-destructive-monoculture.aspx>
- IICA - Instituto Interamericano de Cooperación para la Agricultura, 2010. Evolución y situación: Algodón (Paraguay) [online]. 03-01-10. Retrieved from: <http://www.iica.org.py/observatorio/producto-paraguay-algodon.htm>

IICA - Instituto Interamericano de Cooperación para la Agricultura, 2009. Situación y perspectivas del sector agro-rural Paraguayo 2008. Documento de trabajo. [online]. 12-12-09. Retrieved from: http://www.iica.org.py/Sit_y_Pers_2008.pdf

ISEAL Alliance, 2010. Code of Good Practice for Assessing the Impacts of Social and Environmental Standards Systems. [online] 05-02-10. Retrieved from: <http://community.isealalliance.org/sites/default/files/Draft%20Impacts%20Code%2020100201.pdf>

ISEAL Alliance, 2009. Evaluation Approaches, HOW It Is Being Measured. Part of the ISEAL series State of the Art in Measuring the Impacts of Social and Environmental Standards. [online] 06-07-09. Retrieved from: http://www.isealalliance.org/sites/default/files/R081_ISEAL_Measuring_Impacts_of_Standard_Systems_Public_Summary_Jan09.pdf

ISEAL Alliance, 2008. Key Impact Issues, WHAT Is Being Measured. Part of the ISEAL series State of the Art in Measuring the Impacts of Social and Environmental Standards. [online] 06-07-09. Retrieved from: <http://www.isealalliance.org/sites/default/files/R082%20ISEAL%20Key%20Impacts%20Issues%20WHAT.pdf>

Kates et al., 2005. What is Sustainable Development? Goals, Indicators, Values, and Practice. *Environment*. Vol. 47 (3), p. 8-21

Kates et al., 2003. Long-term trend and a sustainability transition. *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 100 (14), p. 8062-8067.

Kilian et al., 2006. Is sustainable agriculture a viable strategy to improve farm income in Central America? A case study on coffee. *Journal of Business Research*. Vol. 59, p. 322-330.

Kilian et al., 2004. Can the Private Sector be Competitive and Contribute to Development through Sustainable Agricultural Business? A Case Study of Coffee in Latin America. *International Food and Agribusiness Management Review*. Vol. 7 (3), p. 21- 45

Kooistra K.J., Termorshuizen A.J., 2006. The sustainability of cotton. Consequences for man and environment, Wetenschapswinkel Wageningen Universiteit en Research Centrum. Rapport No. 223. ISBN: 90-6754-90-8585-000-2.

Labuschagne et al., 2005. Environmental and social Impact Considerations for sustainable Project Life cycle management in the Process Industry. *Corporate Social Responsibility and Environmental Management*. Vol. 12, p. 38-54.

Labuschagne et al., 2003. Assessing the sustainability performances of industries. *Geoforum*. Vol. 37, p. 1021-1034.

Levett, R., 1998. Sustainability indicators – integrating quality of life and environmental protection. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. Vol. 161 (3), p. 291-303.

Mahdi et al., 2009. Livelihood Change and Livelihood Sustainability in the Uplands of Lembang Subwatershed, West Sumatra, Indonesia, in a Changing Natural Resource Management Context. *Environmental Management*. Vol. 43, p. 84-99.

Marine Stewardship Council, 2010. MSC Environmental Standard for Sustainable Fishing. [online] 02-02-10. Retrieved from: <http://www.msc.org/about-us/standards/standards/msc-environmental-standard>

Mayoux, L., 2006. Quantitative, Qualitative or Participatory? Which Method, for What and When? Chp. 13 in: *Doing Development Research*. SAGE Ltd.; London.

Mentzer, J.T., 2002. What is supply chain management? Chp. 1 in: *Supply Chain Management*, p. 1-27. SAGE Publications Ltd: London.

Meyer et al., 2008. Cotton and Wool Situation and Outlook Yearbook. A report from the USDA Economic Research Service. [online]. 10-04-10. Retrieved from: <http://usda.mannlib.cornell.edu/usda/ers/CWS-yearbook//2000s/2008/CWS-yearbook-12-15-2008.pdf>

- Mullan et al., 2009. Evaluation of the Impact of the Natural Forest Protection Program on Rural Household Livelihoods. *Environmental Management*. [online] 07-06-09. Retrieved from: <http://www.springerlink.com.proxy.library.uu.nl/content/9161702702377x4r/fulltext.pdf>
- Muradian et al., 2005. Governing the Coffee Chain: The Role of Voluntary Regulatory Systems. *World Development*. Vol. 12, p. 2029- 2044
- Narayan, D., 1997, *Voices of the Poor, Poverty and social capital in Tanzania*. The World Bank: Washington D.C.
- Nelissen et al., (1997) *Our Common Future*, Chp 22 in *Classics in Environmental Studies*: 275-284. International books: Utrecht.
- Nelson et al., 2002. Methodological challenges to impact assessment of codes of practice. Paper presented at the 5th annual Corporate Citizenship Conference on International Perspectives on Corporate Citizenship, Warwick Business School, Warwick, 7 October.
- NSSD - National Strategies for Sustainable Development, 2003. Indicators and Information Systems for Sustainable Development. [online] 02-06-2010. Retrieved from: <http://www.nssd.net/references/SDInd/Donella.html>
- O'Leary, Z., 2004. Indicators of Good Research, Chp 5 in: *The essential guide to doing research*. SAGE: London
- PAN Germany., 2007. No food, no home, no partner? What is left when biodiversity disappears? [online] 04-06-09. Retrieved from: <http://www.pan-germany.net/baumwolle/Version2/en/index.html>
- Paredes, M., Rojas, M., 2007. *Costos de Producción y Rentabilidad de Productos Agropecuarios Orgánicos y Agroecológicos y de Microempredimientos Comunitarios*. Alter Vida, European Union, ICCO.
- Pierr, H.P., 2003. Environmental policy, agri-environmental indicators and landscape indicators. *Agriculture, Ecosystems and Environment*. Vol. 98, p. 17-33.
- Putnam, R., 1993. *Making democracy work: Civic traditions in modern Italy*. Princeton: Princeton University Press.
- Rakodi, C., 1999. A Capital Assets Framework for Analyzing Household Livelihood Strategies: Implications for Policy. *Development Policy Review*. Vol. 17, p. 315-342.
- Raynolds et al., 2007. Regulating sustainability in the coffee sector: A comparative analysis of third- party environmental and social certification initiatives. *Agriculture and human Values*. Vol. 24, p. 147-163.
- Robinson, G.M., 1998. *Methods and techniques in human geography*. John Wiley & Sons: New York.
- Schmidt et al., 2004. SEEBalance: Managing sustainable Products and Processes with the Socio-Eco-Efficiency analysis of BASF. *Greener Management International*. Vol. 45, p. 79-94.
- Soth, J., 2006. Mit Kompost gegen den Klimawandel in Partnerschaft. Vol. 185, p. 20-23. [online] 06-06-09. Retrieved from: http://www.helvetas.ch/global/pdf/media/partnerschaft/pa_185_d.pdf
- Studer, C., 2009. RISE – Response Inducing Sustainability Evaluation. Model Synopsis. [online] 07-06-09. Retrieved from: http://www.shl.bfh.ch/fileadmin/docs/Forschung/KompetenzenTeams/Nachhaltigkeitsevaluation/RISE/Model_en.pdf
- Sustainable Coffee Partnership, 2007. The COSA Project. A Multi-criteria Cost- Benefit Analysis of Sustainable Practices in Coffee. [online] 08-07-09. Retrieved from: <http://www.iisd.org/pdf/2007/cosa.pdf>
- UN, 2002. *Johannesburg Declaration on Sustainable Development* part of: UN Documents - Gathering a Body of Global Agreements. [online] 12-05-2010. Retrieved from: <http://www.un-documents.net/jburgdec.htm>
- UN, 2009. *Our Common Future, From One Earth to One World* part of: UN Documents – Gathering a Body of Global Agreements. [online] 03-06-09. Retrieved from: <http://www.un-documents.net/ocf-ov.htm>

UNCSD, 2007. Indicators of Sustainable Development: Guidelines and Methodologies. Third Edition.

UNCTAD, 2010. INFOCOMM: INFORMATION ON COTTON [online] 10-04-10. Retrieved from: <http://www.unctad.org/infocomm/anglais/cotton/sitemap.htm>

Utting, K., 2009. Assessing the Impact of Fair Trade Coffee: Towards an Integrative Framework. *Journal of Business Ethics*. Vol. 86, p. 127-149.

UTZ CERTIFIED, 2010. Impact in Coffee growing, Professionalism, local communities and the environment. [online] 02-02-10. Retrieved from: <http://consumer.utzcertified.org/index.php?pageID=218>

Veleva V., Ellenbecker M., 2000. A Proposal for Measuring Business Sustainability, Addressing Shortcomings in Existing Frameworks. *GMI*. Vol. 31, p. 101-120.

Vermeulen, W.J.V., 2008. Creating sustainable Global product chains: a three levels game. *ICOVACS*

WRI, 2008. Pathways. Creating sustainable business, one leader at a time. Drivers.[online presentation] 08-03-08 Retrieved from: <http://pathways.wri.org/index.asp?Topic=2>

Online presentation:

Gonzales, A.T., 2008. WWF. Choosing what to measure. Presentation at ISEAL Impacts Workshop, Bonn, February 8, 2009. [online] 14-12-09. Retrieved from: www.iseal.org.

Further reading:

Laws et al., 2007. *Research for Development. A practical guide*. SAGE Publications: London

Lindsay, M.J., 1997. Questionnaire, interviews, and allied techniques: getting information from people. Chp 4 in: *Techniques in Human Geography*. Routledge: New York

Sheskin I.M., 1985. Questionnaire development. Chp 4 in: *Geographers, survey research*.

Simon, D., 2006. Your questions answered? Conducting questionnaire surveys. Chp 17 in: *Doing Development Research*. SAGE Publications: London

Willis, K., 2006. Interviewing. Chp 15 in *Doing development research*. SAGE Publications: London

ANNEXES

Annex 1

	<u>External validity</u>		<u>Internal validity</u>	<u>Reliability</u>
	Sample size	Sample drawing	Attribution/ causality	
RISE		Voluntary; over-representation	Analyze status quo, no contextualization	
The Cosa Project	51 farms --> advise more	List of selection criteria to ensure balanced representation/ random selection	Control group of conventional farmers/ statistical tests	
Mullan et al.	Large to be able to use PSM. 285 household interviews and 32 village leader interviews from 33 villages	Select individuals based on their similarity. Village chosen with village leaders, participants approached individually	"Natural experiment"; test and control group. Control observable and unobservable factors through the use of covariates or by estimating the propensity score.	Recall bias not likely to give bias because it is the same for both groups
Mahdi et al.	160 households, key informants and focus groups in three villages.	Households were selected randomly from three income groups; low, middle and high. This was based on the latest monthly household income per capita records, kept in the villages.	Cannot determine the specific agent of change	Clear described method.

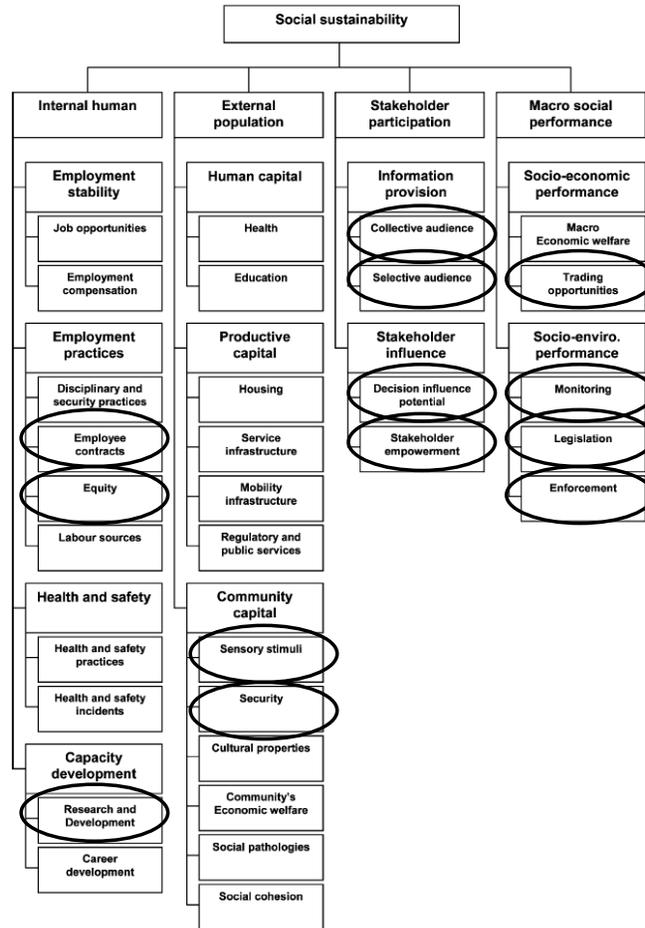
Utting			Impacts asked before and after entering sustainable trade initiative. Contextualization. Triangulation.	
Organic Exchange	Determined by size of Internal Control Group, less important because of participatory method	Try to assure good representation but have no clear tactics to assure this.		
ISEAL			Baseline data, control group, theory of change , explain external influences , opinions and attitudes of people.	Transparency

Manageability						
Accessibility	Locating sample	Costs	Time	Resources	Results representation	Comparability
		relatively high	3-6 houses per interview/ over three years	Interviews & databases	Sustainability polygon	Globally applicable, polygon makes it easy.
	hard to find farms only enrolled in one type of certification system		Annual visit over three years/ farmer economic log to minimize time	Need help from experts, not all answers can be provided by farmers	Spider graph	
	Assistance from local forest bureau		PSM is time expensive	Face to face, no problems with databases	SPSS statistics	Easy to compare because of statistical approach
	Had some help of local counties. All people in the sub-watershed were affected by the policy thus finding the target group was not difficult.			Some basic income data were available. More was not needed.	Data of two points in time represented in asset pentagon.	It is possible to compare the data between different regions and to use the system in different parts of the world. Most indicators are relative and can therefore be compared between different regions. Asset pentagon.

				Some local docs needed and interviews	Qualitative data; no easy analysis.	Accommodates for differences in the social, political, environmental and institutional context. Different objectives of systems, levels of intervention and the stakeholder involved. Qualitative data make it more difficult.
	Own staff working together with local stakeholders	Are kept low by only having one meeting with few experts.	No problem	No records needed, ask farmers directly.	Bar charts	Use the same tool in different areas. Realize the importance of translating terms to local contexts.
	Encouraged by stakeholder engagement	Kept low through cooperation.	Cooperate, use existing reports.			Only when other standard system uses same method. Simplicity for analyzing is promoted.

Annex 2

Labuschagne et al., Social indicator framework



)

Annex 3

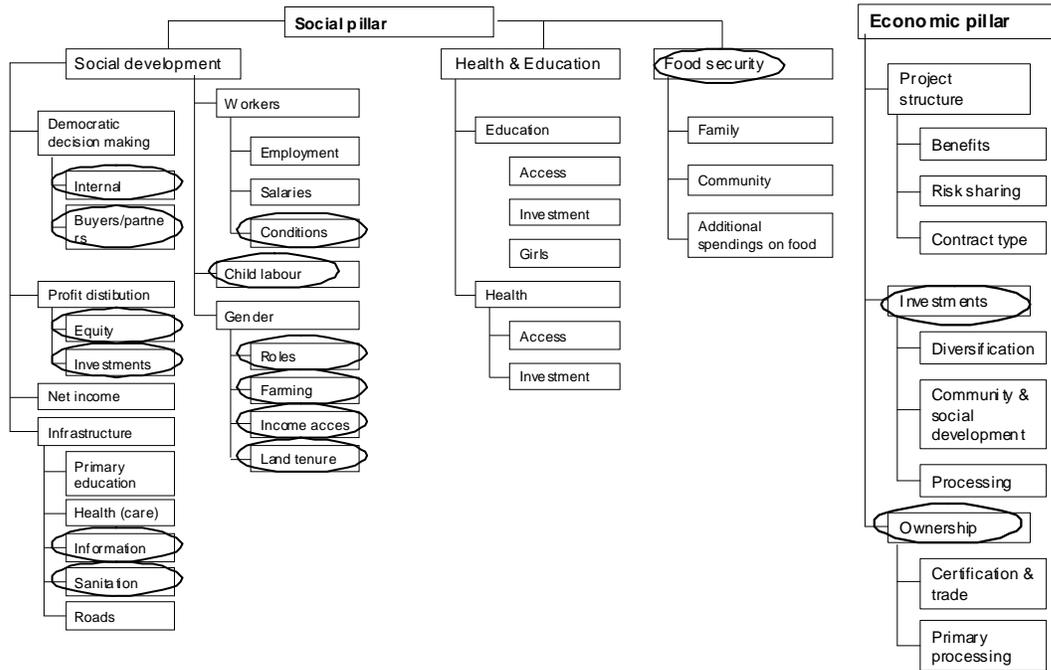
UNCSO Social and Economic Indicator framework

<i>Theme</i>	<i>Sub-theme</i>	<i>Core indicator</i>	<i>Other indicator</i>
Poverty	Income poverty	Proportion of population living below national poverty line	Proportion of population below \$1 a day
	Income inequality	Ratio of share in national income of highest to lowest quintile	
	Sanitation	Proportion of population using an improved sanitation facility	
	Drinking water	Proportion of population using an improved water source	
Poverty (continued)	Access to energy	Share of households without electricity or other modern energy services	Percentage of population using solid fuels for cooking
	Living conditions	Proportion of urban population living in slums	
Governance	Corruption	Percentage of population having paid bribes	
	Crime	Number of intentional homicides per 100,000 population	
Health	Mortality	Under-five mortality rate	
		Life expectancy at birth	Healthy life expectancy at birth
	Health care delivery	Percent of population with access to primary health care facilities	Contraceptive prevalence rate
		Immunization against infectious childhood diseases	
	Nutritional status	Nutritional status of children	
Health status and risks	Morbidity of major diseases such as HIV/AIDS, malaria, tuberculosis	Prevalence of tobacco use	
		Suicide rate	
Education	Education level	Gross intake ratio to last grade of primary education	Life long learning
		Net enrolment rate in primary education	
		Adult secondary (tertiary) schooling attainment level	
	Literacy	Adult literacy rate	
Demographics	Population	Population growth rate	Total fertility rate
		Dependency ratio	
	Tourism		Ratio of local residents to tourists in major tourist regions and destinations

Economic development	Macroeconomic performance	Gross domestic product (GDP) per capita	Gross saving
		Investment share in GDP	Adjusted net savings as percentage of gross national income (GNI)
			Inflation rate
	Sustainable public finance	Debt to GNI ratio	
	Employment	Employment-population ratio	Vulnerable employment
		Labor productivity and unit labor costs	
Share of women in wage employment in the non-agricultural sector			
Information and communication technologies	Internet users per 100 population	Fixed telephone lines per 100 population	
		Mobile cellular telephone subscribers per 100 population	
Economic development (continued)	Research and development		Gross domestic expenditure on R&D as a percent of GDP
	Tourism	Tourism contribution to GDP	
Global economic partnership	Trade	Current account deficit as percentage of GDP	Share of imports from developing countries and from LDCs
			Average tariff barriers imposed on exports from developing countries and LDCs
	External financing	Net Official Development Assistance (ODA) given or received as a percentage of GNI	Foreign direct investment (FDI) net inflows and net outflows as percentage of GDP
Consumption and production patterns	Material consumption	Material intensity of the economy	Domestic material consumption
	Energy use	Annual energy consumption, total and by main user category	Share of renewable energy sources in total energy use
		Intensity of energy use, total and by economic activity	
	Waste generation and management	Generation of hazardous waste	Generation of waste
		Waste treatment and disposal	Management of radioactive waste
	Transportation	Modal split of passenger transportation	Modal split of freight transport
			Energy intensity of transport

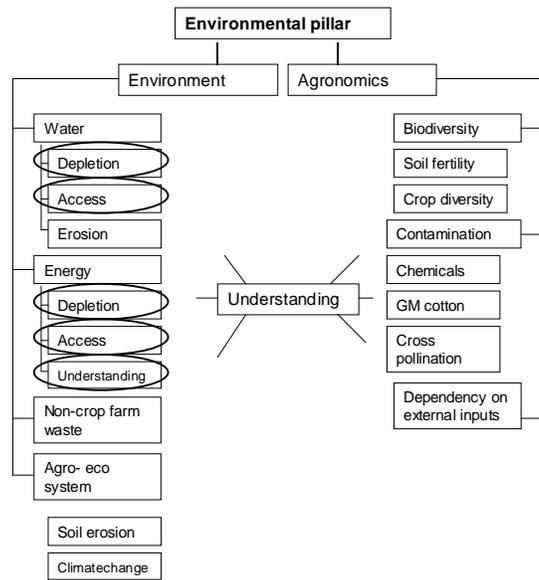
Annex 4

Organic Exchange Social framework & Economic framework



Annex 5

Organic Exchange Environmental framework



Aspect operationalization

Equity

Gender → Roles, farming, income access, land tenure, decision making power
Organic Exchange provides a broad description of what the concept of equity contains: “This topic is about equity between the sexes in terms of decision making and equal share in proceeds of work, distribution of roles, access to income, land tenure and ownership. It addresses who cultivates, what they can cultivate, who gets the proceeds and who decides on spending. A further question relates to who is the nominal holder or owner of the land, and if women can farm and/or enjoy the returns from land they hold or own.

Community investments (for example if farmer associations invest in their communities).

Community investments in this research concerns whether the farmers, the farmer associations, or the labelling organisation make any monetary investments in their community they are part of. These investments concern projects which benefit the community. The investments of the farmers and association do not necessarily have to be made within their own community but can also be done in other communities. Investments of the labelling organisation do only count if they are made in the area in which their providers (the framers they buy the products from) live.

Production capacities and knowledge

Capacities in this research concern an ability or talent developed through experience or training.

Food security;

Family, additional spending on food

Food security concerns “when all people at all times have access to sufficient, safe, nutritious food to maintain a healthy and active life” (WHO, 2009).The main attention in this research is at access to sufficient food and the diversity in their food provision. Furthermore, attention is paid to additional spending on food.

Housing

A definition of housing can be as follows: Housing concerns the physical state of the house the people are living in. A house is considered good when it is appropriate for the circumstances the house is functioning in. This considers size, construction type, appropriateness to the local weather conditions etc.

However, in this research the attention is on the changing housing state and whether the owners consider their house is having any discrepancies.

Production equipment

Production equipment concerns all the instrumentalities which can be used in the production process.

Labour demand

Labour demand concerns the level of demand for labour on the farm as well as outside the farm.

Ownership; product processing and certification

Ownership of the certificate.

Processing of raw products.

Earnings (net)

Net earnings consider the incomes minus the costs for production. In this research this is measured for the farm as a whole as well as for cotton production specifically.

Credits

Credits concern the access of the producers to external financial sources.

Savings

Savings are a fund of money put aside.

- Climate change

Global warming is influenced by energy use and the burning of organic materials (some extent CH₄ produced through animal labour). Loss of trees.

- Loss of life support functions

Destruction or alteration of land, or the results from harvesting biotic resources. For example with forests. If forests are harvested it is for example desertification which can occur.

Also caused by salinisation.

Salinisation of soil

The accumulation of soluble [mineral](#) salts near the surface of [soil](#), usually caused by the capillary [flow](#) of [water](#) from saline ground [water](#). Where the [rate](#) of surface [evaporation](#) is high, [irrigation](#) can exacerbate the problem by moistening the [soil](#) and causing [water](#) to be drawn from deeper levels as [water](#) evaporates from the surface.

The [evaporation](#) of pure [water](#) leaves the salts behind, allowing them to accumulate, and they can [reach](#) concentrations that are toxic to plants, thus sterilising the [land](#).

[definition source: [EEA](#) multilingual environmental [glossary](#)]

Terminology source: <http://glossary.eea.europa.eu>

- Salinisation

Influenced by the manner of watering of the soils. In Paraguay the producers do not make use of irrigation methods, therefore this aspect is not influenced and not taken up in the research.

- Aquatic eco-toxicity

Influenced by the use of pesticides, spraying methods, frequency of use, quantity per application and toxicity of the pesticide.

- Depletion of abiotic resources

The focus on this aspect is solely on seeds. The depletion of natural (non-chemical) cotton seeds. The focus is solely on the types of seeds that are being used and not on cross-pollination.

- Terrestrial eco-toxicity

Influenced by the use of pesticides. spraying methods, frequency of use, quantity per application and toxicity of the pesticide.

- Loss of biodiversity (trees, plants, animals)

Biodiversity loss is determined by the use of pesticides, natural habitat conversion and loss of land due to degradation and shifting farming policies.

Influenced by use of pesticides. spraying methods, frequency of use, quantity per application and toxicity of the pesticide. Measured through analyzing these aspects as well as the biodiversity measures taken by the farmers; planting of trees and the amount of land that lays fallow.

- Water resources

The depletion of water resources depends on the type of irrigation systems used and the efficiency of the system. Drip irrigation is for example far more efficient as flood-or-furrow irrigation.

- Land resources

The depletion of land resources can be measured through combining the impacts on above mentioned aspects such as biodiversity, terrestrial eco-toxicity, salinisation, acidification, eutrophication, impacts of land use and the loss of life support functions (= salinisation amongst others). Therefore, the impact on land resources will be determined based on the above mentioned aspects.

- Human toxicity

impacts on health, caused by use of pesticides. spraying methods, frequency of use, quantity per application and toxicity of the pesticide

Annex 7

Compliance rules; quick overview

A producer may not:

- use pesticides
- use synthetic fertilizers
- use chemically treated seeds
- burn any materials on their land
- plough up fallow land
- have any agrochemical in his farm whatsoever.
- cannot use any equipment contaminated with agrochemicals without cleaning them beforehand.

A producer does have to:

- use agro-ecological measures to maintain soil fertility and avoid erosion.
- control plagues in a natural manner
- use a backpack that is only used for natural remedies
- try to avoid contamination of agrochemicals on the farm by creating a protection strip
- cooperate with the internal control and the consultation and advices provided by Alter Vida.
- cultivate organic fertilizers
- pay-off all the open bills with Aratex at the end of each harvest season
- deliver its entire yield to Aratex
- 10% of the total farm area should be reserved for ecological purpose

Annex 8

Interviewed Key- persons

José L. Rolón – President of Cadelpa – The Paraguayan cotton chamber

Carlos Céspedes Zarza - Aratex

Wilfrido Ortíz – Responsible for Alter Vida Misiones.

Rafael Dominguez Martinez – CEPAG Local NGO

Estela Chena - Manufactura Pilar

Ubaldo Britos - Professor University of Asunción

Felix Ortíz – Agroeconomico

Head of MAG Santa Maria

Mayor of Santa Maria

Local bank employee

Annex 9

Fig. 1 N=46

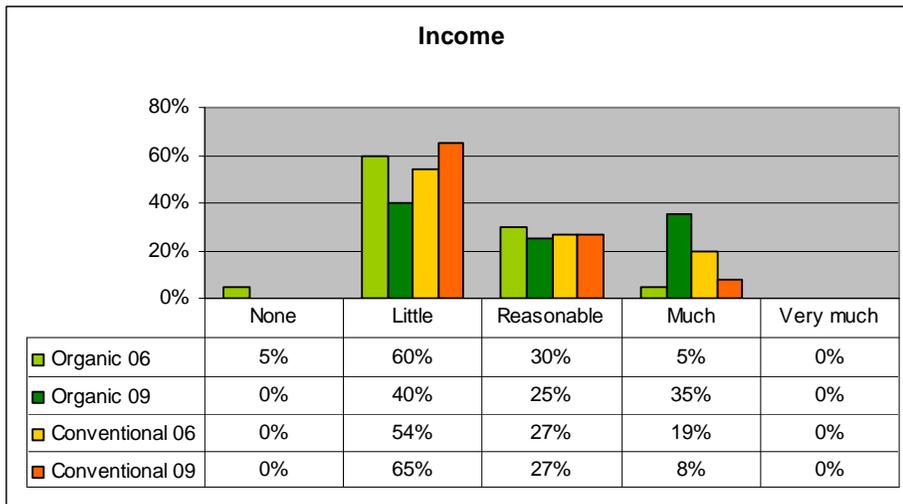


Fig.2 N=46

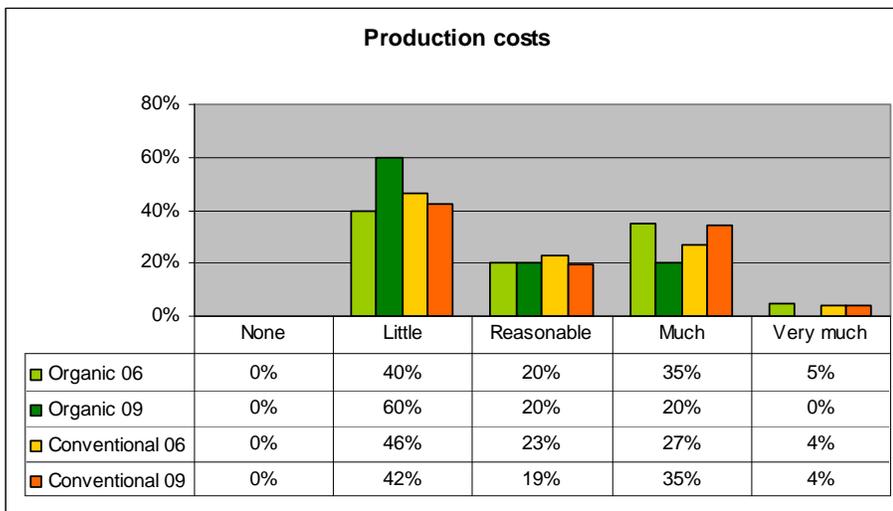


Fig. 3 N=41

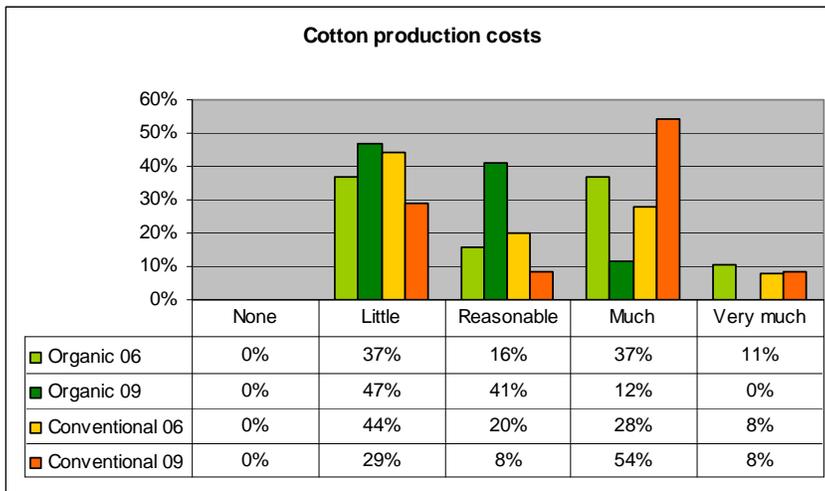


Fig. 4 N=46

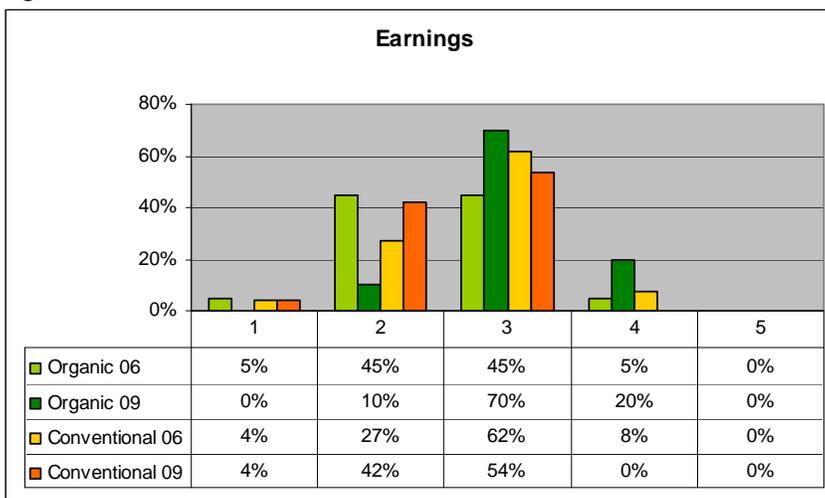
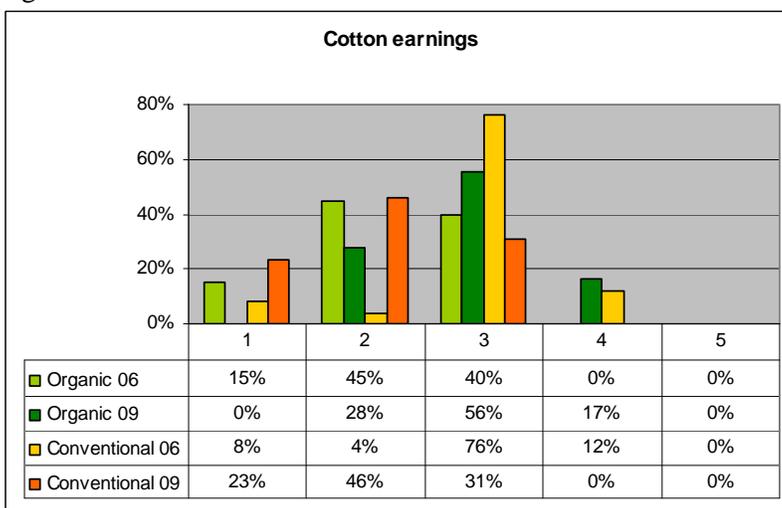


Fig. 5 N=44



Banco Nacional de Fomento – Credit provision

For farmers there are two options for receiving credit at the BNF. Farmers can obtain credit to make investments in their business and they can obtain credit for operational capital. The latter concerns investments in the production cycle such as labour, soil preparation, sowing, harvesting etc. The former type of investment concerns long term investments. The farmers need to have some sort of guarantee in order to obtain such a credit. This should be some form of proof of ownership. This can be from a house, land, or an animal. The farmer can receive up to 60% of the value of the item he can prove ownership of. These loans are mostly over a period of 5 years with an interest of about 13%. With the proof of ownership the farmer can also obtain another 30% of the value in credit that can be used for investments in the production cycle. This is a short term loan which should be paid back after harvest. Finally, also farmers that do not have a proof of ownership can obtain short-term credit. This credit is meant for investments in the production cycle and should be paid a month after harvest. The amount can increase up to 70% of the estimated costs of production.

There is no differentiation in the type of producers that do or do not receive credit. It all depends on whether they possess a proof of ownership or not and if they can show a good estimation of the future production costs. The type of crop does not influence the opportunities a farmer has at receiving credit either. The amount a farmer receives can differ per crop simply because the production costs differ per crop as well.

In general the amount of loans that the BNF has provided over the last years has declined enormously. Three years ago the bank financed around 300 small scale farmers, now it is only 2. The bank is not so eager at providing loans to small scale farmers because the prices of most crops have been declining, especially that of cotton. However, there is a different bank which still provides these small loans to small scale farmers.

Annex 11

Additional graphs – effects on the economic and social situation

Fig. 1 N=40

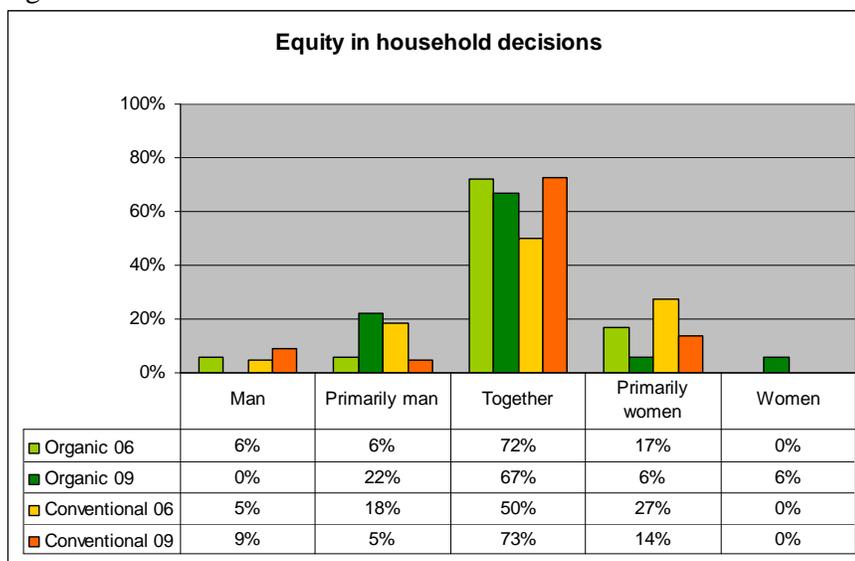


Fig. 2 N=40

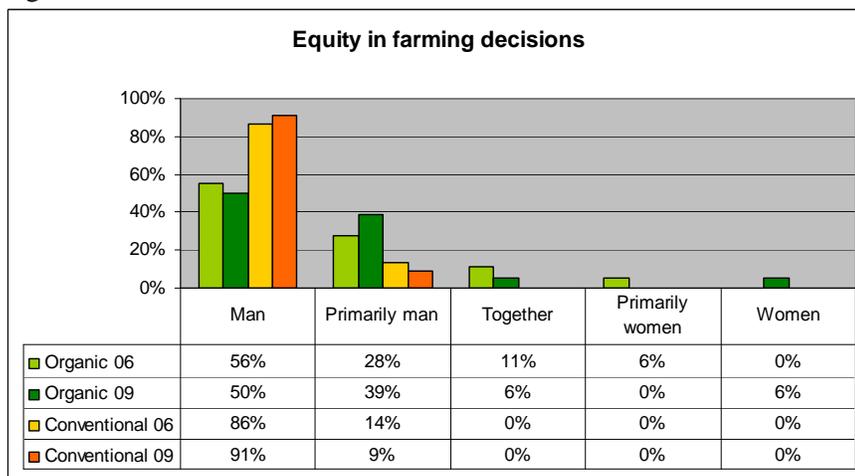


Fig. 3 N=46

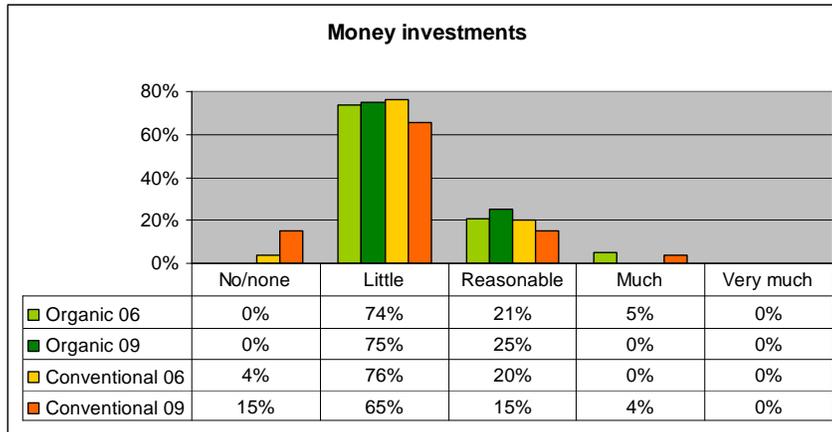


Fig. 4 N=44

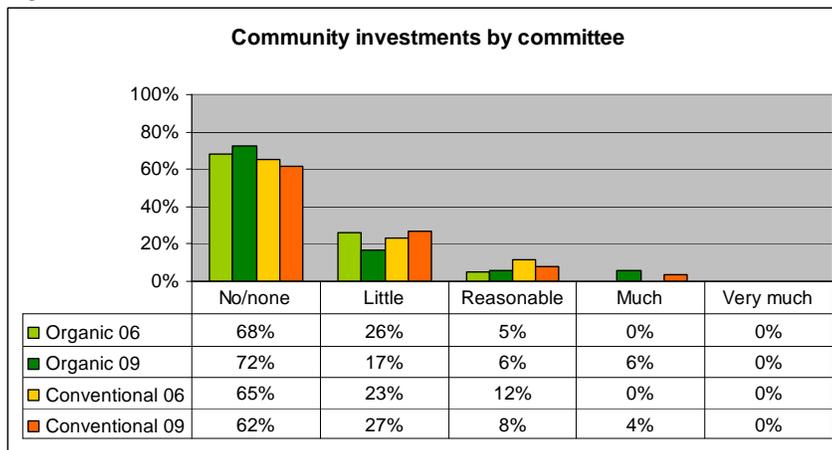


Fig.5 N=10

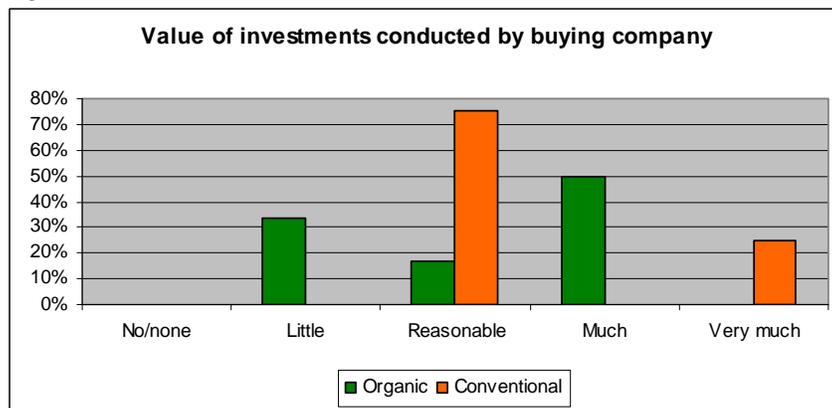


Fig. 6 N=41

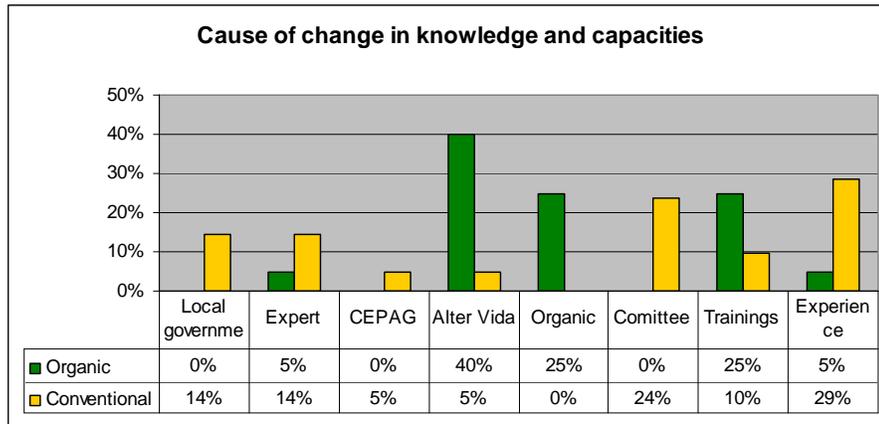


Fig. 7 N=19

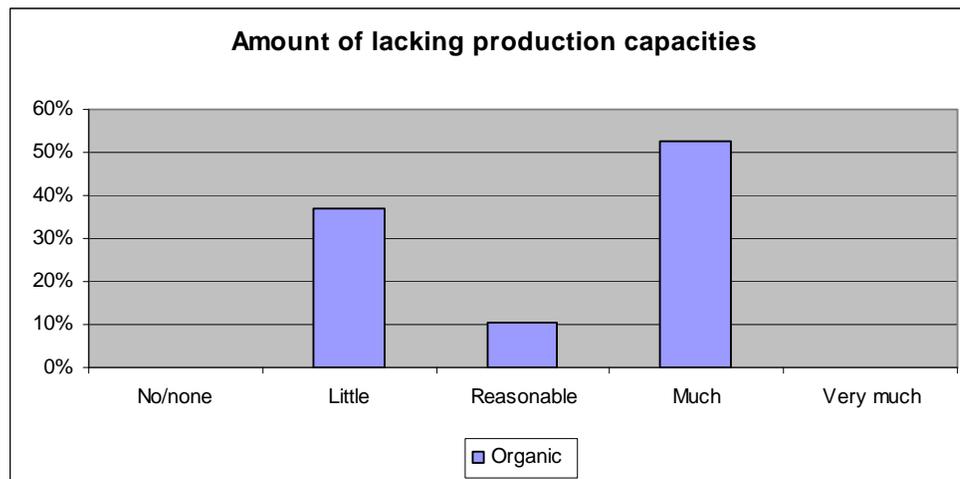


Fig. 8 N=17

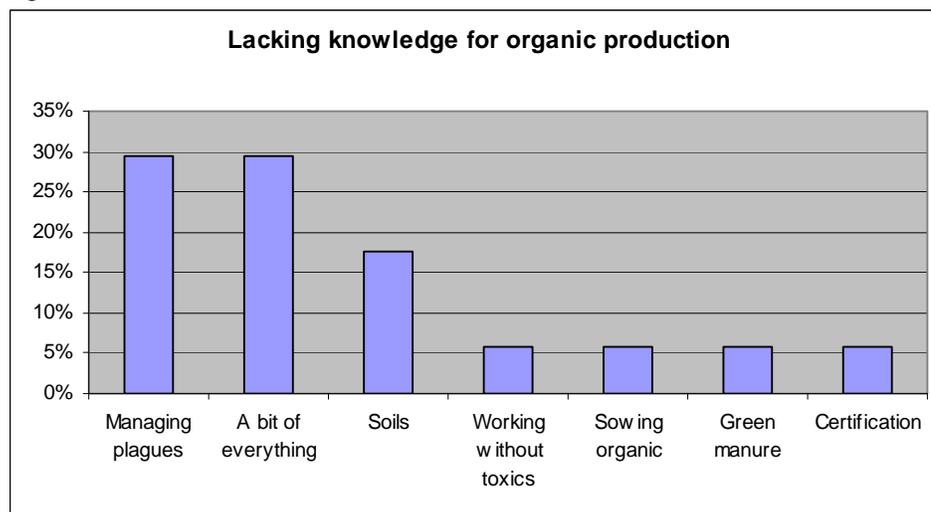


Fig. 9 N=46

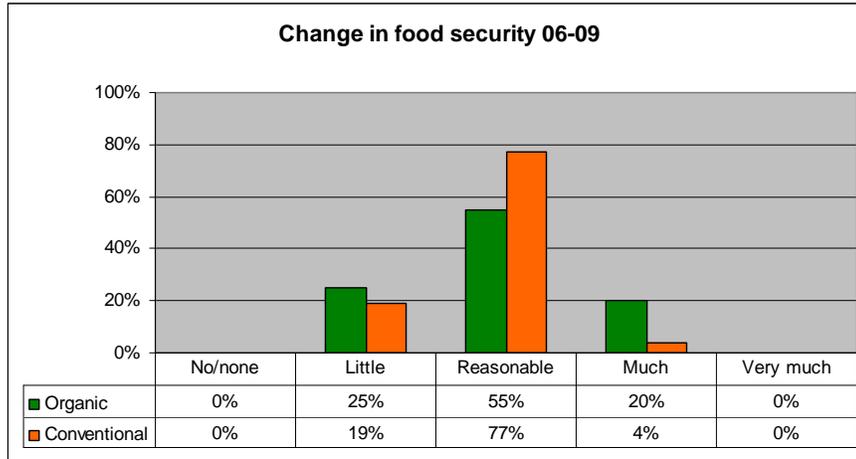


Fig. 10 N=15

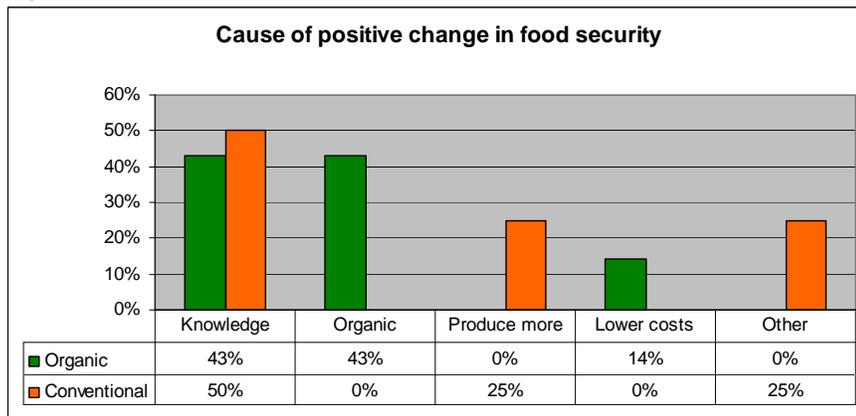


Fig 11 N=46

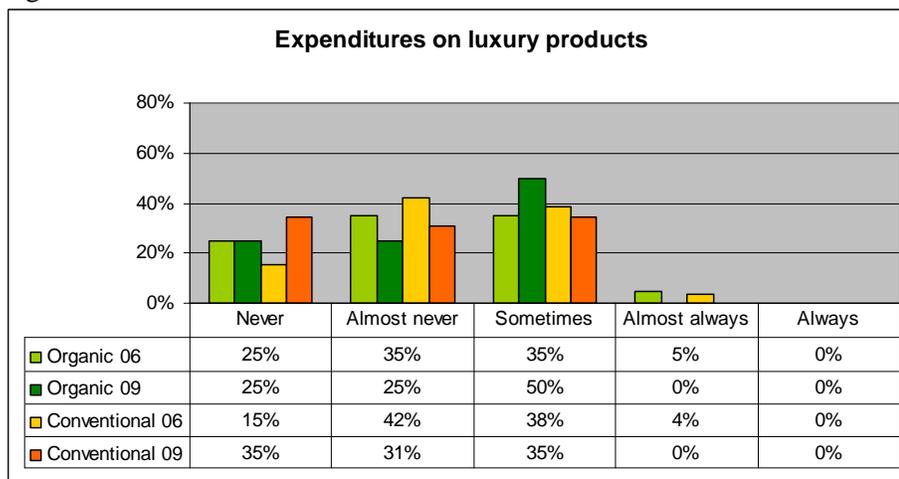


Fig. 12 N=46

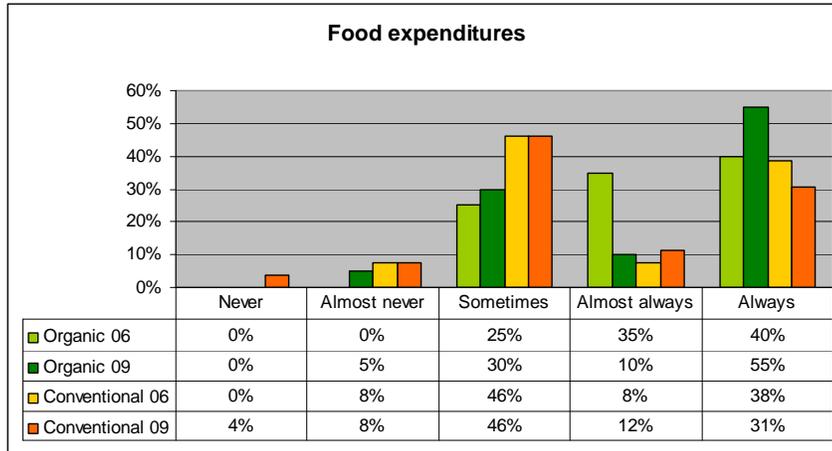


Fig. 13 N=46

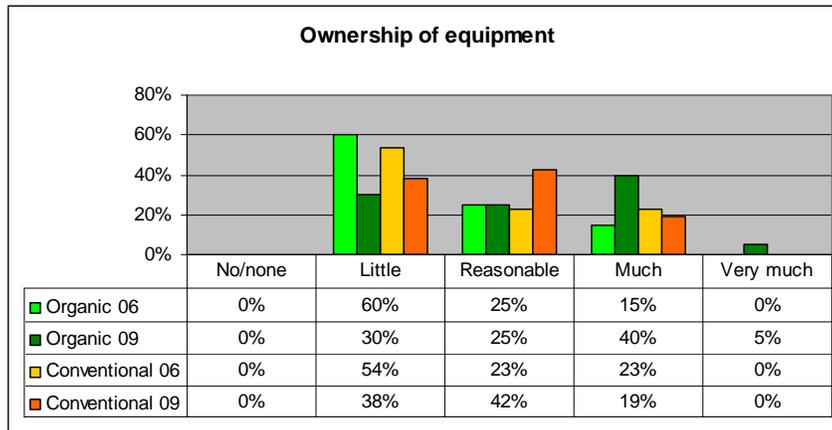


Fig. 14 N=21

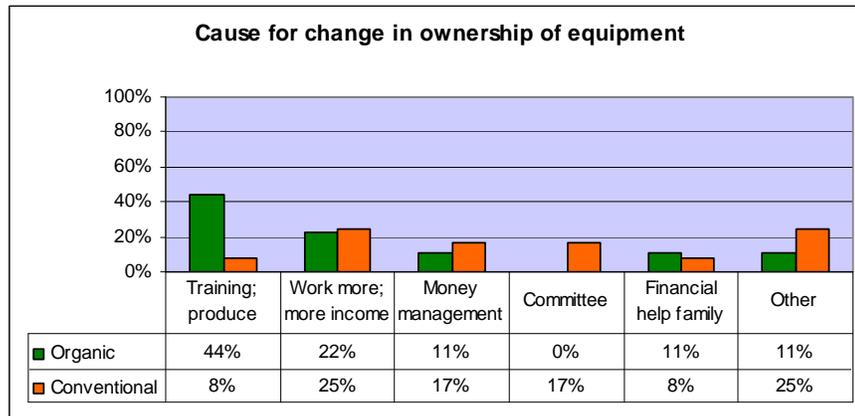


Fig. 15 N=46

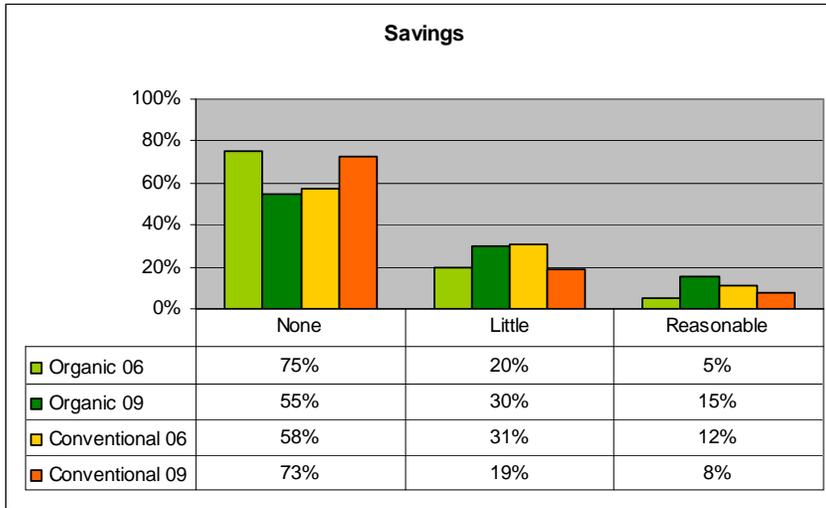


Fig. 16 N=46

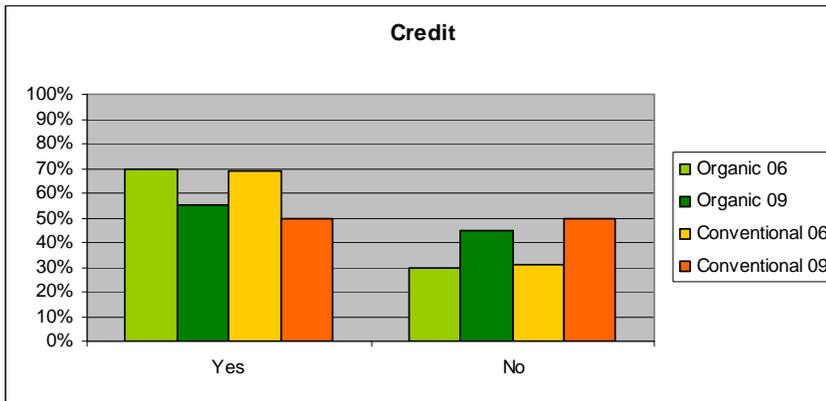


Fig. 17 N=45

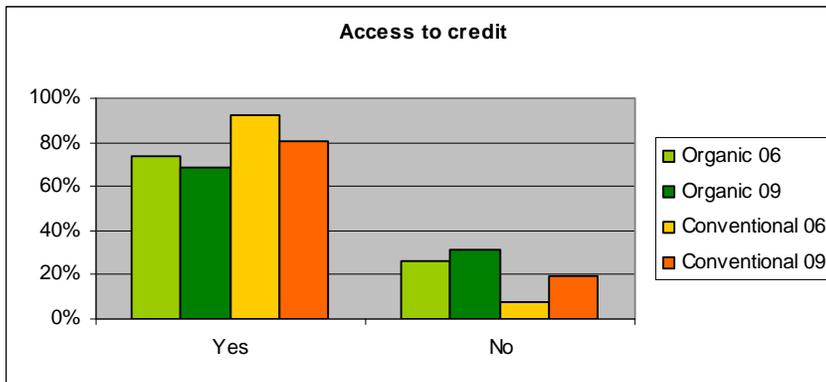


Fig. 18 N=45

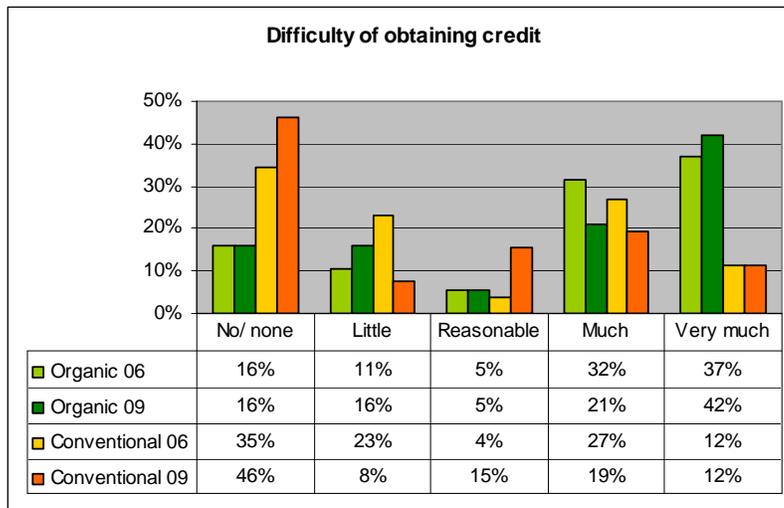


Fig. 19 N=20

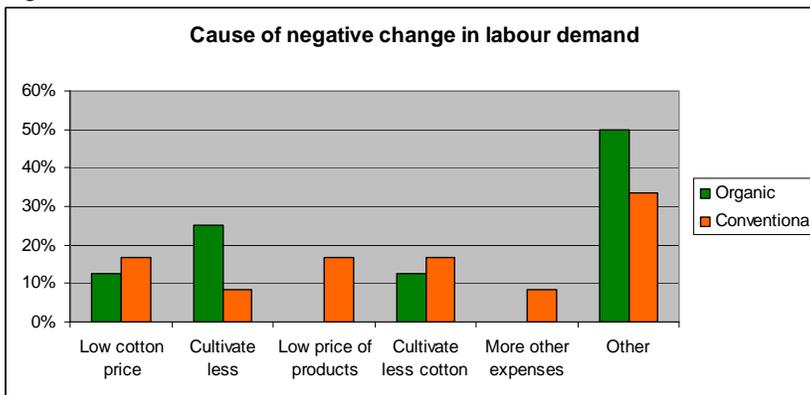
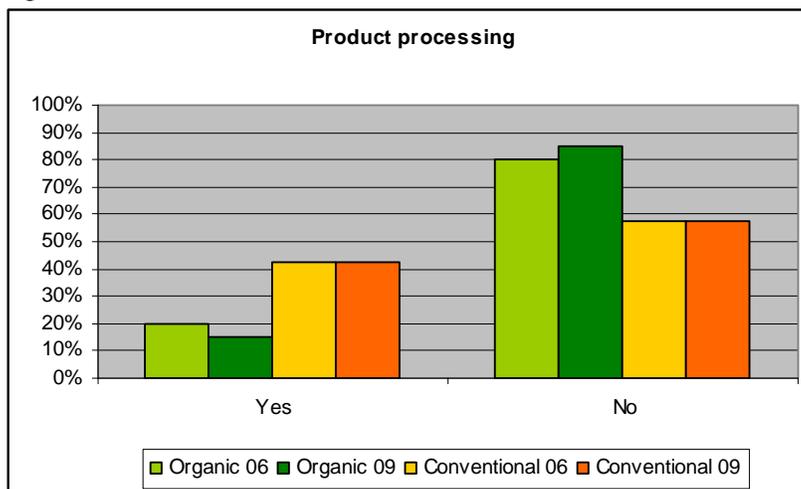


Fig. 20 N=46



Annex 12

Additional tables and graphs- effects on the environment

Table 1 Change in hectare of cotton cultivated

		Report			
Orga yes/no		Hectares	Hectares cotton 06	Hectares cotton 09	Hectares cotton planned
yes	Mean	12,2500	1,9038	1,1885	,9562
	N	20	13	13	8
	Std. Deviation	10,22445	,96576	,76052	,52197
no	Mean	9,8600	2,8500	1,8750	,8182
	N	25	20	20	11
	Std. Deviation	9,18050	2,49789	,97164	,71668
Total	Mean	10,9222	2,4773	1,6045	,8763
	N	45	33	33	19
	Std. Deviation	9,62062	2,06757	,94525	,62945

Fig. 16 N=46

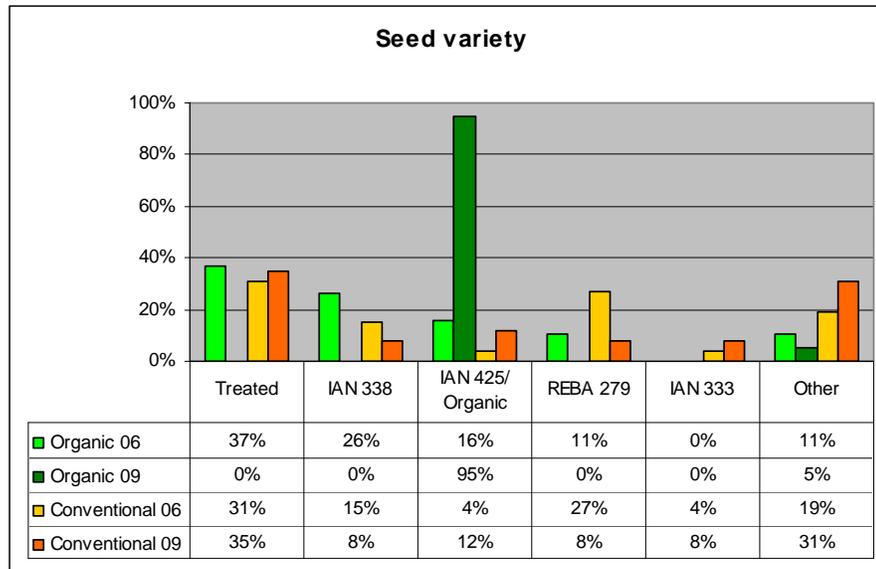


Fig. 17 N=46

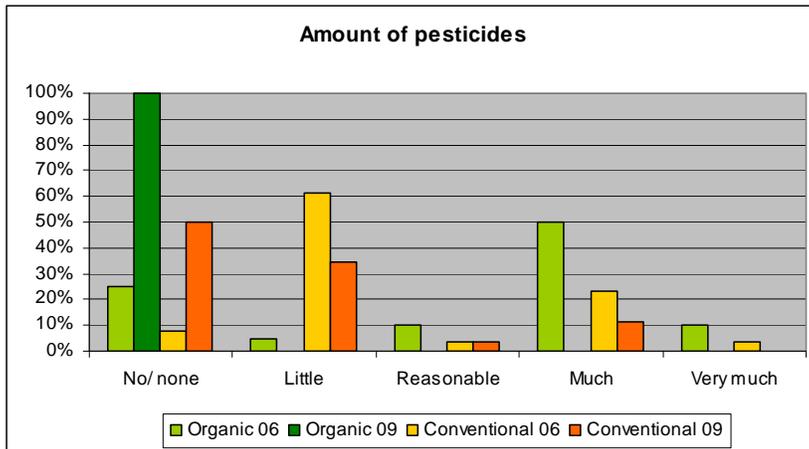


Fig. 18 N=20

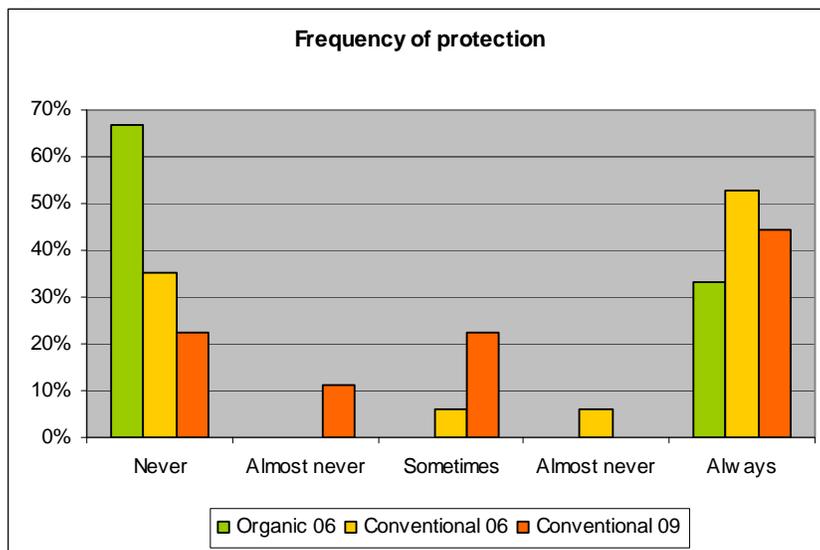


Fig. 19 N=25

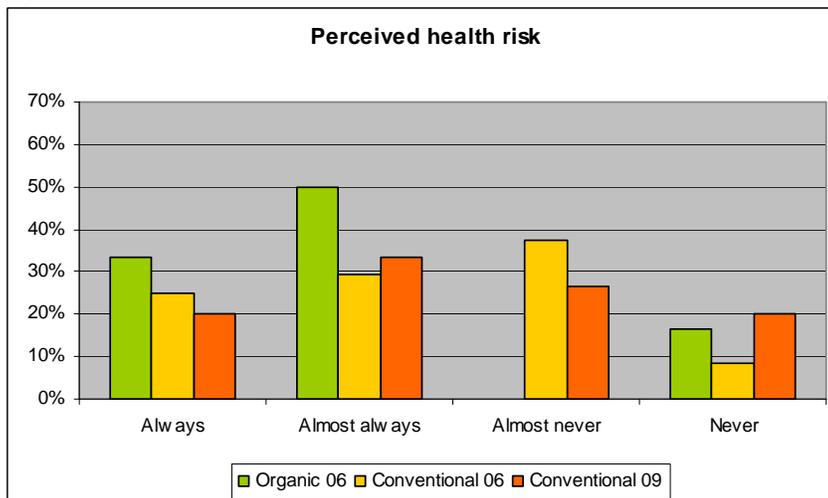


Fig. 20 N=24

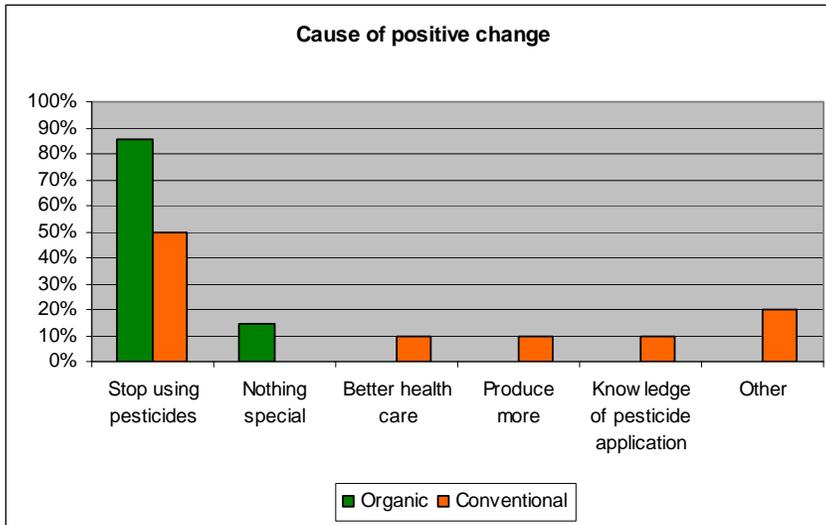


Fig. 21 N=44

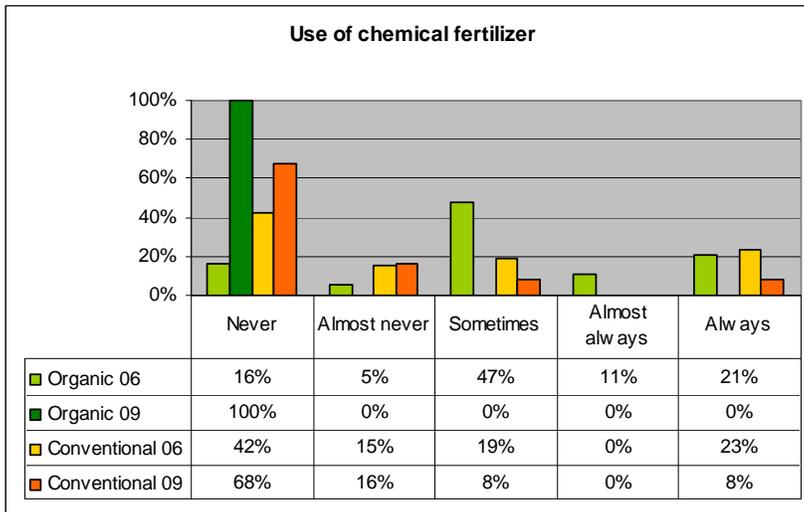


Fig. 22

